



Can photovoltaics be used to estimate cloud cover?

Stavros Stylianou , Rogiros Tapakis & Alexandros G. Charalambides

To cite this article: Stavros Stylianou , Rogiros Tapakis & Alexandros G. Charalambides (2020) Can photovoltaics be used to estimate cloud cover?, International Journal of Sustainable Energy, 39:9, 880-895, DOI: [10.1080/14786451.2020.1777129](https://doi.org/10.1080/14786451.2020.1777129)

To link to this article: <https://doi.org/10.1080/14786451.2020.1777129>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 10 Jun 2020.



Submit your article to this journal [↗](#)



Article views: 307




View related articles [↗](#)



View Crossmark data [↗](#)

Can photovoltaics be used to estimate cloud cover?

Stavros Stylianou , Rogiros Tapakis and Alexandros G. Charalambides

Chemical Engineering Department, Cyprus University of Technology, Lemesos, Cyprus

ABSTRACT

The current study evaluates a new method for the estimation of cloud cover using only photovoltaics (PV) as ground-based irradiance sensors. Seven different scenarios were examined using 30, 50, 75, 100, 500 and 1000 randomly and 1000 non-randomly distributed residential PV stations in a study area of 10 km × 10 km. For each scenario, 3500 different fractal-based cloud shadows were produced and the accuracy of the model in estimating the cloud cover, in percentage cover and in oktas, based on the number of stations being shaded, was calculated. With 30 PV stations, the model produced an average accuracy of only 57% for the predicted okta cover, while its accuracy increased to 90% and 94% with 500 and 1000 PV stations.

ARTICLE HISTORY

Received 7 January 2020
Accepted 25 May 2020

KEYWORDS

Cloud cover; meteorology;
weather prediction;
photovoltaics

Introduction

Clouds can affect climate, not only because they may produce rain and snow but also because they strongly modulate the flows of both incoming irradiance (warming the earth's surface) and infrared light (cooling the earth as it is radiated to space) through the atmosphere. Many climate scientists study the behaviour (formation and movement) of clouds based on the daily weather using both observations and models with the seasonal cycle, and with year-to-year changes.

Estimating, and in extent predicting cloud cover has numerous practical applications, not only for climate-related research but also for energy production from solar energy systems (solar energy forecasting) (Madhavan et al. 2017). Unlike conventional power generation or non-variable renewable energy sources, solar energy is considered a variable energy source (Ela et al. 2013; Perez et al. 2016). This is because electricity production from solar technologies cannot be controlled based on the demand but is highly depended on the intensity of the incident solar irradiance which varies, amongst other factors, on cloud cover (Adaramola 2016; Gul and Stenzel 2005; Martins, Souza, and Pereira 2003). Therefore, to estimate accurately the global radiation in the short-term interval, it is necessary to understand the presence and effect of clouds on solar energy production (Falayi and Rabiou 2011; Li, Moreau, and Arking 1997).

Cloud cover is defined as the ratio of cloud area to whole sky area at the observed point (Yamashita, Yoshimura, and Nakashizuka 2004) and is usually given in units called oktas, where each okta represents one eighth of the sky covered by clouds; ranging from 0 oktas (clear sky) to 8 oktas (completely overcast). Several methods and equipment such as ground-based systems and satellites (Martinis et al. 2013), as well as human observations made at weather stations, have been used for determining cloudiness conditions or the cloud fraction above the earth's surface. Human daily cloud observations at

CONTACT Alexandros G. Charalambides  a.charalambides@cut.ac.cy

This article has been republished with minor changes. These changes do not impact the academic content of the article.

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

meteorological stations are considered to be the earliest method used for cloud cover estimation and cloud type determination and although it is considered a useful method, Sakellariou and Kambezidis (2004) have used data in oktas to predict the cloud cover, obtaining satisfactory results, the introduction of the human factor weakens its credibility (Kazantzidis et al. 2012; Silva and Souza-Echer 2016). Ground-based equipment such as irradiance metres/radiometers, sky cameras and radars although often cover longer time scales, are mainly used to measure data for specific locations, whereas satellites are used for measuring data over continents and provide global coverage (Ackerman and Cox 1981; Werkmeister et al. 2015). Ground-based instrumentation has often a smaller field of view and needs additional equipment to be installed at the locations of interest, limiting the ability of monitoring the formation of clouds and their movement over larger sky areas such as over cities, regions and countries. Nevertheless, since the state of the sky is monitored on-site, they provide a sufficient accuracy for the local variations of solar irradiance due to clouds (Tapakis and Charalambides 2013). On the other hand, satellites provide large-scale cloud information and multispectral measurements from different sensors, but the provided spatiotemporal data is often in low resolution and may contain errors; small clouds are often overlooked due to the limited analysis and low or thin clouds are not always distinguishable (Dybbroe, Karlsson, and Thoss 2005; Heinle, Macke, and Srivastav 2010; Ricciardelli, Romano, and Cuomo 2008). The cost of equipment is also an important factor for cloud measurements, since using satellites is much more expensive than most ground instruments. Yet, some specialised ground meteorological observing systems such as RADARs, may be equally expensive (Clothiaux et al. 1998; Seiz, Baltasvias, and Gruen 2002; Wang and Sassen 2004).

In this paper, a simple method for the automatic cloud cover estimation in oktas using existing PV power stations as ground-based irradiance sensors is being investigated (Tapakis and Charalambides 2016). To understand and identify the optimal number of PV stations needed for an accurate estimation of cloud cover in an area of 10 km × 10 km, six different scenarios using 30, 50, 75, 100, 500 and 1000 randomly distributed residential PV stations were examined. The PV stations are used as binary data points and thus the production of these in the model is either 0 (not shaded) or 1 (shaded). However, it is expected that in a real-life scenario PVs will be of the size of 1 to 5 kWp, as they will be installed on rooftops, and the appropriate distances were kept in our model. One more scenario was carried out using 1000 distributed residential PV stations over the same area but excluding parks, roads and waterways where no PV stations were placed. For each scenario, 3500 different fractal-based cloud shadows, (500 per okta from 1 to 7 oktas) were produced, and the accuracy of our method in estimating the cloud cover based on the number of stations being shaded was calculated. In the context of this work, the methodology does not consider the conversion efficiency of the PV stations or the position of the sun. The calculations are made when the sun is at its highest point and the solar irradiance is perpendicular to the clouds surface, thus the cloud shadow assumed to be equal to the size of the cloud.

Given the fact that nowadays photovoltaics are used widely, especially in populated areas, we believe that without the need for any additional sophisticated and expensive equipment or the use of any meteorological data, PV data can provide interesting meteorological information such as cloud cover. This innovative work acts as the foundation for the development of an algorithm to predict cloud motion in real time using only data from photovoltaics, that can then be correlated back to predict energy production from photovoltaics. To the best of our knowledge, this is the first time such a methodology is applied using only data from PV stations to calculate the cloud cover over an area.

The format of this paper is organised as follows: Section 2 explains the methodology used for the creation of the input data and the calculation of the cloud cover over the study area, Section 3 presents the results from the six scenarios and finally Section 4 concludes this paper.

Methodology

The methodology presented in this paper for the estimation of cloud cover over an area using only the output from PV stations was performed in three steps. The first step involved the production of

3500 random fractal-based cloud shadows over an area of 10 km × 10 km using the midpoint displacement algorithm (Fang 2009; Saupe 1988). These simulated cloud shadows were used in the model to represent the real cloud cover in order to identify which PV stations were under shade and thus estimate the predicted cloud cover (Section 2.1). In the second step, randomly distributed PV stations were generated in the study area and the number of PV stations under cloud shade were tabulated. The shade coverage in square kilometres per okta was then predicted as is analysed in depth in Section 2.2. The final step comprised of the calculation of the simulated oktas based on the generated shadow cover and the estimation of the predicted cloud cover from the PV stations under shade as will be shown in Section 2.3. The simulated cloud cover was then compared with the predicted cloud cover to calculate the accuracy and the associated errors of the method in each scenario using the Root Mean Square Error (RMSE). Throughout this paper, ‘simulated’ cloud cover refers to the cloud cover calculated by the shaded pixels in the image, while ‘predicted’ to the cloud cover estimated by the number of the shaded PV stations.

Fractal cloud shadow generation

For the identification of the PV stations being shaded and as a mean of estimating the predicted cloud cover from the PV stations, 3500 different fractal-based cloud shadows were produced using the midpoint displacement algorithm (Saupe 1988). For all scenarios, 500 fractal shadows were created per okta, ranging from 1 to 7 oktas, while shadows for 0 and 8 oktas, which represent zero and complete cloud cover, were not computed. Real cloud cover data from ground-based systems and/or satellite imagery could have been used to create the required 3500 images with random cloud cover, but since this required either the installation of additional equipment or the purchasing of expensive high-resolution satellite imagery, it was decided to use numerical models, that generate random cloud shadow patterns on the ground, instead.

Fractal-based modelling has been used by many researchers to reproduce the naturally irregular shapes found in nature, such as cloud shadow patterns, and therefore generate more realistic results than models based on simple geometrical shapes (Cai and Aliprantis 2013). The algorithm used to generate the fractal shadows in this work (Appendix A) is based on a modified version of the Mid-PointFM2D () midpoint algorithm from the ‘Algorithms for random fractals’ in ‘The Science of Fractal Images’ (Saupe 1988). The required parameters for the algorithm are: (a) *maxLevel*, which sets the number of iterations that would be used as $N = 2maxLevel$, (b) *sigma*, the initial standard deviation and (c) parameter *H*, which determines the fractal dimension ($D = 3 - H$). These parameters can be modified to generate different types of cloud shadows.

The midpoint displacement method is iterative and takes $N = 2maxLevel$ stages to complete. In each stage, the centre midpoint values are calculated based on the four adjacent corner points of each square and then the values on the midpoints of the edges are calculated (Figure 1). The final output of the algorithm is a two-dimensional fractal surface. Some examples of generated cloud shadows for oktas 1, 3, 5 and 7, that have been used in this paper are illustrated in Figure 2, where $H = 0.6$, $sigma = 1.9$ and $maxLevel = 9$ were used.

The algorithm presented here describes the main part of the process where the simulated shadows are being generated (Figure 2). The okta score of these shadows is calculated from the area covered by the clouds sorted into eight equally space bins (see Table 1). The midpoint displacement algorithm generates an array of dimension N with zeros (0) and ones (1), where zero values represent the shadow and one clear sky. The percentage of the cloud cover over the area is calculated following:

$$sp = \frac{\text{shadow}}{S} \times 100, \quad (1)$$

where sp is the percentage cloud cover over the area, shadow is the number of the zeros in the array of the fractal shadow and S represents the dimension squared (N^2). Finally, the okta category is assigned to the fractal shadow image based on the values in Table 1 in Section 3.1.

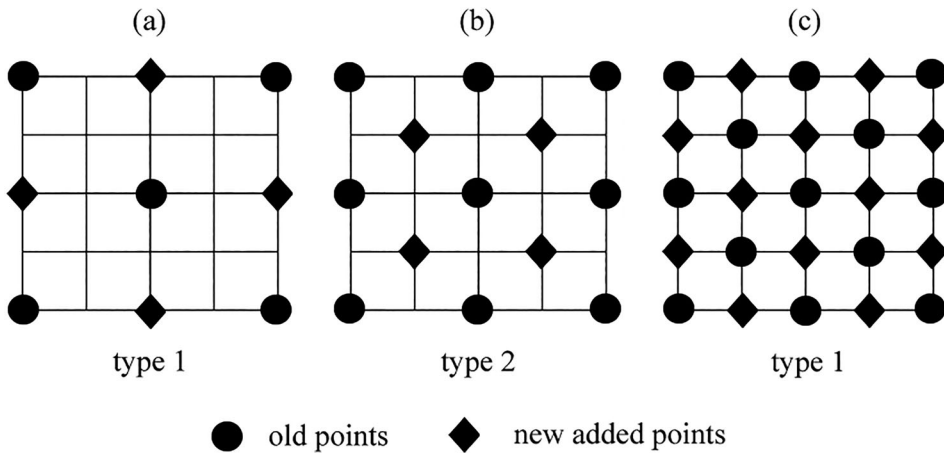


Figure 1. Midpoint displacement method. A grid of type 1 (a) is generated at the beginning from which a type 2 grid is given (b), with mesh size is $1/\sqrt{2}$ times the old mesh size. Following in a similar step a type 1 grid is again generated (c) and so forth.

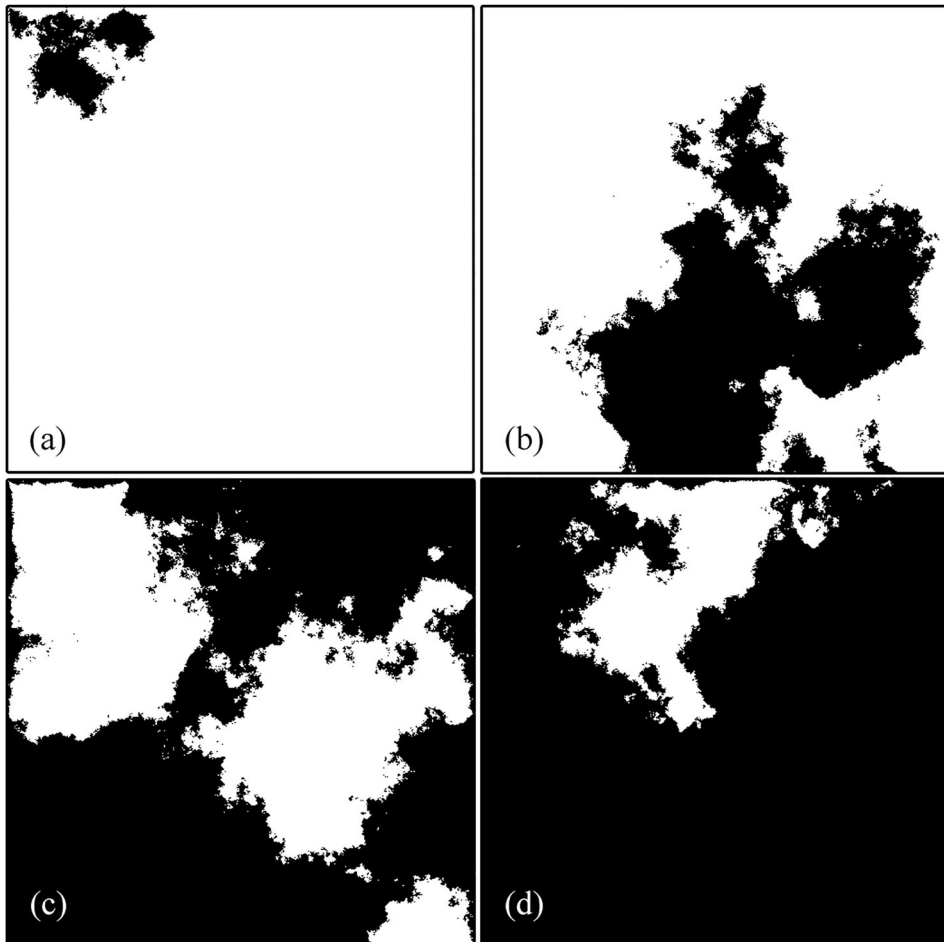


Figure 2. Cloud shadow patterns representing: (a) 1 oktas, (b) 3 oktas, (c) 5 oktas and (d) 7 oktas cloud cover in the study area. The cloud shadow is presented with black colour while a clear sky with white colour.

Table 1. Scale of cloud cover measured in oktas (eights) and the calculated percent scale with the meteorological symbol for each okta.

Scale in oktas	Percent scale (%)
0 – Sky completely clear	0
1	0.1–14.2
2	14.2–28.5
3	28.5–42.8
4 – Sky half cloudy	42.8–57.1
5	57.1–71.4
6	71.4–85.7
7	85.7–99.9
8 – Sky completely cloudy	100

Identifying shaded PV stations

The fractal-based cloud shadows obtained using the midpoint displacement method in the previous step were used as the basis to calculate which PV stations were under shade. For this task, the method presented here utilises geographic information systems (GIS) geoprocessing to (a) georeference the cloud images according to the 10 km × 10 km study area, (b) generate the randomly distributed PV stations for each scenario and finally (c) calculate the number of PV stations that fall under shade. As the first step in our methodology, we have only calculated cloud shade during the time of the day when the sun is at its highest point and the solar irradiance is perpendicular to the clouds surface (Figure 3), thus the cloud shadow was equal to the size of the cloud (considering and the distance of the sun). The PV stations are used as binary data points and thus the production of these in the model is either 0 (not shaded) or 1 (shaded). As part of our future work, cloud shadow for all sun angles will be calculated and validated and real electricity production datasets will be used in the methodology.

PV performance issues such as the natural degradation of the solar cells over time, physical damage or panel discoloration, soiling due to dusty environments or inverter issues were not modelled at the present stage and the functionality of the PVs under study assumed to be uniform, thus not affecting the accuracy and reliability of cloud cover estimation.

The algorithm used in our method is iterative and takes 3500 stages to complete for each of the seven scenarios using 30, 50, 75, 100, 500 and 1000 PV randomly distributed PV stations throughout the study area and the 1000 PV stations where roads, waterways and parks are excluded (Figure 4). In each stage of the process the PV stations for each scenario are distributed randomly in the study area with a 20 metres minimum distance between them (Figure 5) and then are intersected with the georeferenced cloud shadow layers. Thus, the PV stations under shade are calculated. The final results of the number of PV stations under shade for each okta condition and the shadow coverage in square kilometres are presented in tabular form.

The oktas scale used in this study, for practical reasons, ranges from 1 to 7 oktas instead of the one meteorologists use from 0 to 9 oktas. This is because 0 oktas represent the completely clear sky, while 8 oktas the completely overcast scenarios that can be easily predicted by both a human observer and any predictive software. The 9th okta category is also excluded from our methodology as it describes the sky being obstructed from view. Overall, 24,500 simulations for all seven scenarios were run.

Cloud cover estimation

For estimating the ‘predicted’ cloud cover, a percent scale was formed based on the oktas cloud scale that is being widely used (Werkmeister et al. 2015; WMO 2015) (Table 1). The percent scale was created in order to be easily implemented within the algorithm for the calculation of the simulated cloud cover (fractal-based cloud shadows) and the estimation of the predicted cloud cover from the PV stations under shade in oktas, and the comparison of the two.

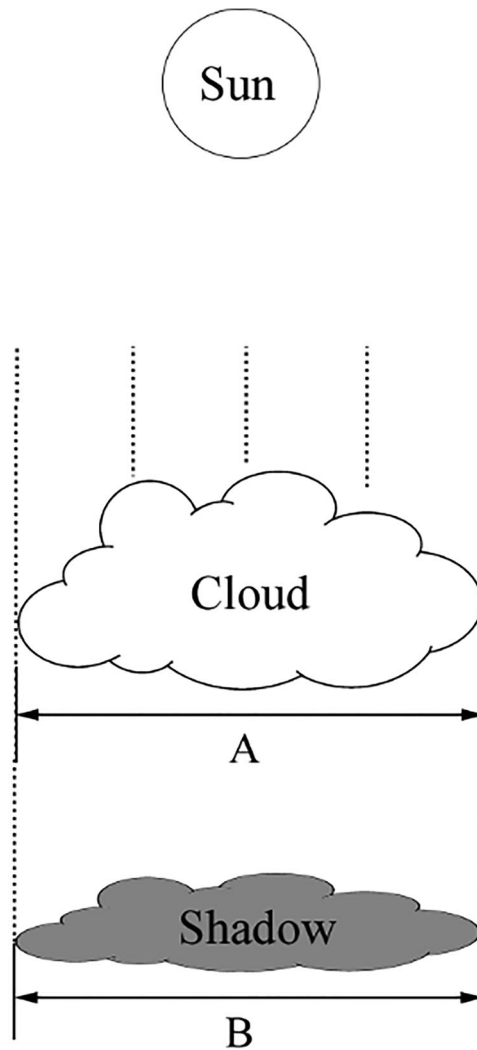


Figure 3. Cloud shadow formation when the sun is at its highest point. The rays of light coming from the sun are considered parallel to each other and the shadow will have about the same size of the cloud itself ($A \approx B$).

Based on the number of the PV stations under the shade and the calculated coverage of each of the simulated cloud shadows in square kilometres (see section 2.2), the simulated and the predicted cloud cover are calculated and presented in a new table. Any wrong estimations of the cloud oktas (i.e. simulated one was 2 oktas, while predicted one was 3 oktas) are counted as errors after the comparison of the simulated and predicted cloud cover in oktas. The simulated cloud cover refers to the calculated cloud cover in oktas scale from the fractal-based cloud shadows layers, while the predicted one refers to the estimated one from the number of the PV stations that were under shade using the following equation:

$$C = \frac{P_{\text{shade}}}{S_{\text{total}}} \times 100, \quad (2)$$

where C is the percentage cloud cover, P_{shade} is the number of the PV stations under the shade, and S_{total} is the total number of PV stations used in each scenario. The calculated cloud cover C is then

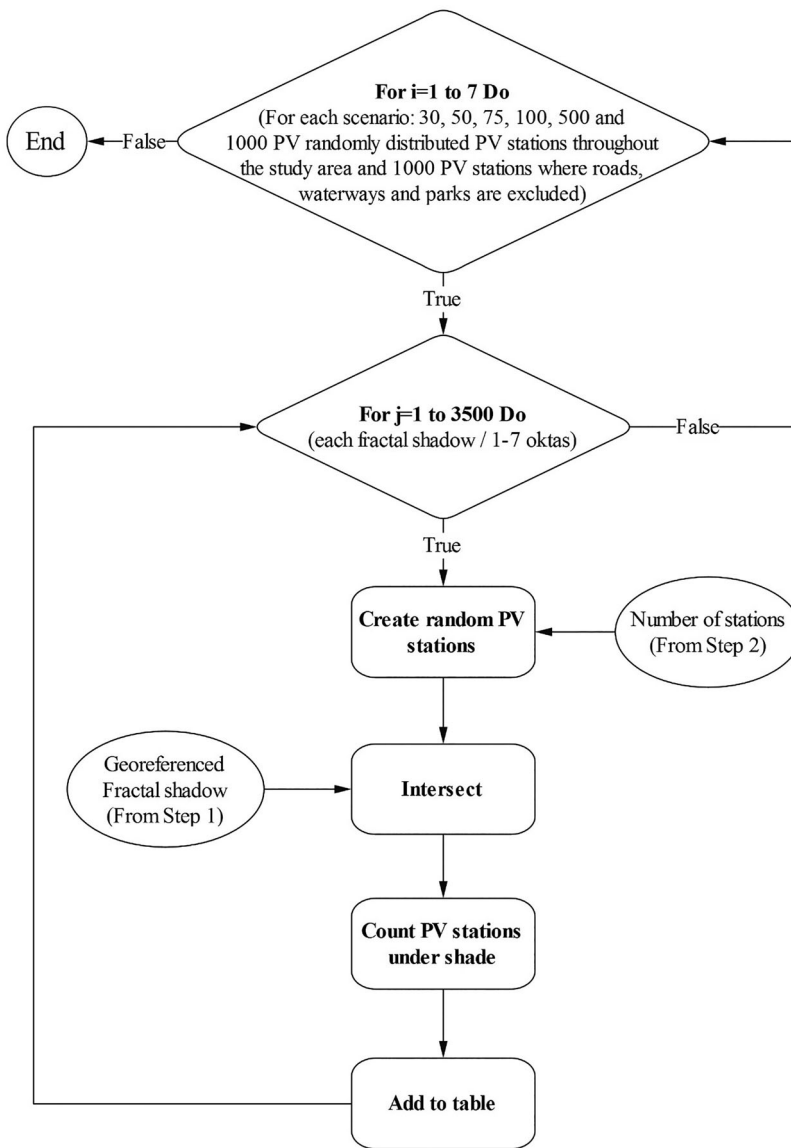


Figure 4. Flowchart of the methodology for identifying shaded PV stations (Step 3 of process).

compared with the percentage scale (Table 1) and the relative okta was being assigned. For example, in the 500 PV stations scenario, if 52 PVs were shaded, C was equal to 10.4% $((52/500) \times 100)$ and the predicted oktas was equal to 1. Predicted oktas were then compared to the simulated one and the results are presented in confusion matrices, as will be shown in the following section.

Results and discussion

Six different scenarios using 30, 50, 75, 100, 500 and 1000 randomly distributed PV stations in a study area of $10 \text{ km} \times 10 \text{ km}$ and one with 1000 distributed PV stations excluding parks, roads and waterways, were used as input to measure the accuracy of our methodology in calculating the cloud cover in oktas using only the power output from these PV stations. For each scenario, 3500

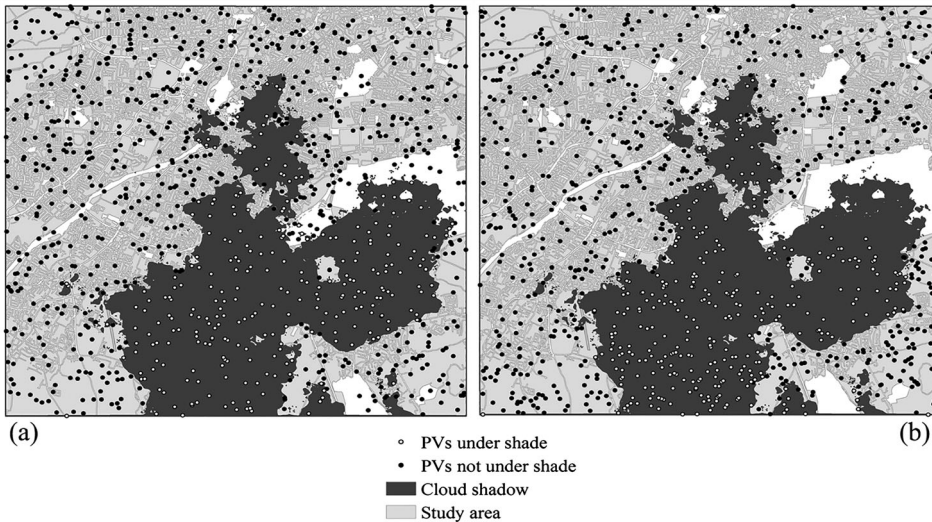


Figure 5. (a) 1000 PV stations randomly distributed throughout the study area of 100 km² with a shadow from a 3 okta cloud cover. (b) 1000 PV stations randomly distributed in the study area of 100 km² with a shadow from a 3 okta cloud cover introducing blind areas (parks, roads and waterways) indicated by the white area. Black dots represent the randomly generated PV stations not under shade while the white dots the randomly generated PV stations under shade. The shadow is represented with black colour.

different fractal-based cloud shadows were produced using the midpoint displacement algorithm, and the accuracy of the estimated cloud cover based on the number of stations being shaded was predicted. The results from each scenario were compared with the simulated results (fractal shadows) and the RMSE was calculated using the following equation (Equation (3)) to identify the optimal scenario and the least number of PV stations needed for an accurate estimation of the cloud cover expressed both as a percentage and in oktas.

$$\text{RMSE} = \frac{\sqrt{\sum_{i=1}^n (s_i - e_i)^2}}{n}, \quad (3)$$

where n equals to 3500, which is the total number of the simulated cloud shadows examined in the methodology for each scenario, s is the simulated values (observations) and e the predicted values from the methodology. Throughout this paper 'simulated' cloud cover refers to the cloud cover calculated by the shaded pixels in the image and used as the observations in the methodology, while 'predicted' to the cloud cover estimated by the number of the shaded PV stations.

Results from each scenario are presented in Tables A1–A7 in Appendix B. In the tables, the column entries denote the fraction of inferred oktas (bins) for a fractional cloud cover again in oktas, as given by the first column. As it can be seen from the results in the confusion matrices, the accuracy of the estimation in each okta category is improved as the number of stations increases. The scenario with 30 PV stations shows a higher degree of error, ± 2 oktas while the rest scenarios ± 1 oktas error. For example, with 30 PV stations, 1% of the simulated cloud cover of 5 oktas was wrongly classified by our methodology as 3 oktas (Table A1). Scenarios with 30, 50, 75 and 100 PV stations tend to overestimate and/or underestimate more frequently and at higher degree the simulated cloud cover, while the scenarios with 500 and 1000 PV stations have produced the best estimates.

Using the first scenario with only 30 PV stations in the methodology the highest accuracy achieved was 64%, for at least one okta range, while for the scenarios with 500 and 1000 PV stations was 97% and 99%, respectively. Similarly, comparing the lowest accuracies produced by the methodology between the lowest and the maximum number of stations used, the 30 PV stations achieved a 49% accuracy, while using 500 and 1000 PV stations, 85% and 91%, respectively. In most of the

scenarios, the methodology achieved lower accuracies between 3 and 5 oktas which could be attributed to the more uneven distribution of ‘cloud’ and ‘sky’ pixels.

The scenario with 1000 PV stations, where parks, roads and waterways were excluded has produced a slightly less accurate cloud cover estimations than the one where the same number of PV stations were distributed throughout the study area, with an average accuracy of 82%. The highest accuracy achieved, for at least one okta range, was 96%, and the lowest 73% (Table A7). It seems that when distributed PV stations are not present throughout the whole area, lower accuracies as regard the cloud cover in oktas are achieved.

Figure 6 presents the mean percentage accuracy of the estimated cloud cover in oktas for all seven scenarios. As it can be seen from the results of scenarios one to six, there is a positive correlation between the number of stations used in each scenario and the accuracy of the estimation. The average accuracy of the model increased from 57%, using 30 PV stations as ground-based irradiance sensors, to 90% in scenarios 5 and 94% in scenario 6 with 500 and 1000 PV stations respectively. Scenario 6 with 1000 PV stations, produced better results than the 500 PV stations scenario, with the individual cloud cover estimations per okta to be more accurate. This is demonstrated in Figure 7, where the RMSE with the lowest score of 0.25 indicates that scenario 6 performs better with regard to the estimation of the cloud cover in oktas. The error bars in Figure 6 representing standard deviation indicate that scenarios 4 (100 PV stations) and 6 (1000 PV stations) with the lowest values produce more coherent predictions as the results for each okta range are closer to the mean. When considering the percentage cloud cover estimation, the RMSE scores presented in Figure 8 indicate that scenario 6 with 1000 PV stations, as the best one having a RMSE score of 1.32. Scenario 7, where no PV stations distributed in parks, roads and waterways, produced an average accuracy of 82% having a RMSE score of 0.42 as regards to the per okta cloud cover estimation and 3.92 for the percentage one.

As the results indicate, there is no need to evaluate the methodology with more than 1000 PV stations due to the fact that RMSE scores from 500 and 1000 PV stations are close to each other and because we need to keep a feasible number of stations within an area. The difference in the resulting estimations for the okta and percentage cloud cover between the scenarios used in this study lies to the fact that the oktas and percentage (%) cloud cover estimations are calculated based on the number of stations covered by clouds in the whole area, where more PV stations produce a more accurate estimation.

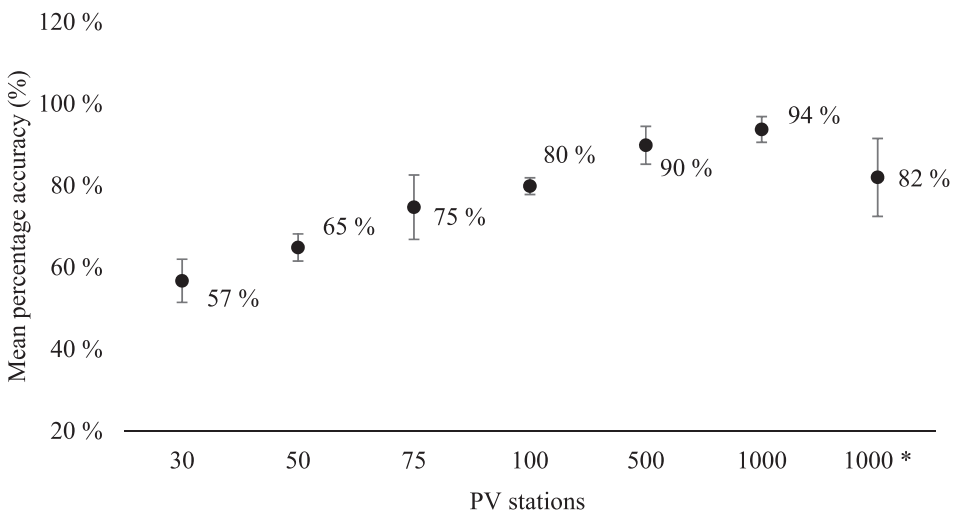


Figure 6. Mean percentage accuracy of the estimated cloud cover in oktas per stations used in each scenario. * Parks, roads and waterways excluded.

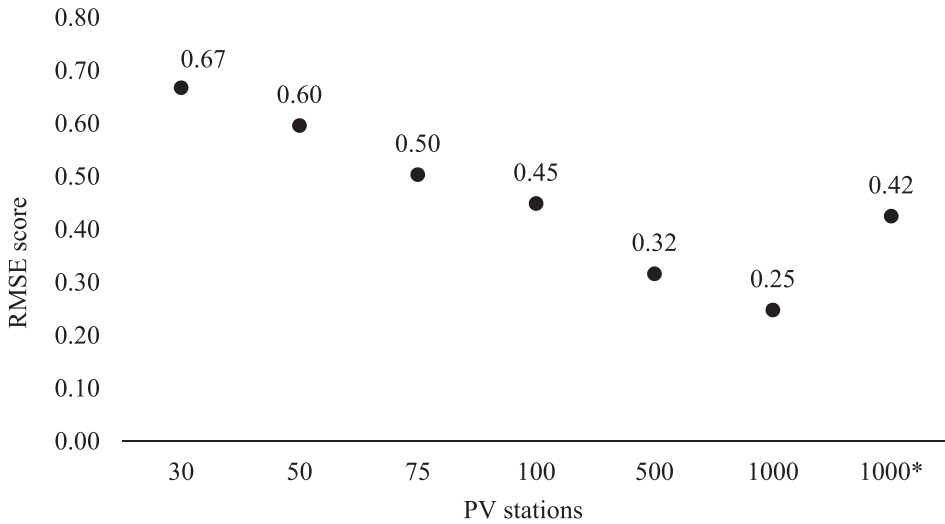


Figure 7. Root Mean Square Error of the estimated cloud cover in oktas per stations used in each scenario. * Parks, roads and waterways excluded.

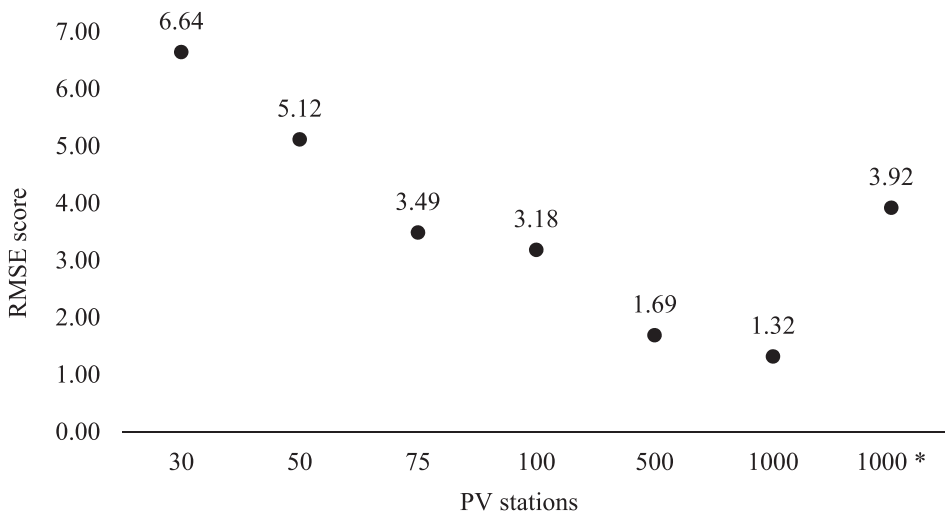


Figure 8. Root Mean Square Error of the estimated cloud cover expressed as a percentage per stations used in each scenario. * Parks, roads and waterways excluded.

Figure 9 presents the percentage accuracy of the estimated cloud cover for each scenario per okta. Scenario 6 with the 1000 PV stations has produce the best estimations for all oktas with an accuracy of 90% and higher. It is also observed that our model performs more accurately at low and high oktas, as we either have almost clear skies or overcast, for all scenarios.

RMSE per okta per scenario was also calculated to study the okta category, where the model produces the more accurate results as regards the percentage cloud cover estimation (Figure 10). Similarly, with the conclusions from the confusion matrices, the RMSE scores indicate that when the cloud cover is between 3 and 5 oktas, correct cloud cover estimation is lower and the RMSE is higher, due to the sparse distribution of the clouds in the region. Nevertheless, the percentage cloud cover estimation of the model (Figure 6) has proved to be more than sufficient in scenarios with 500 and

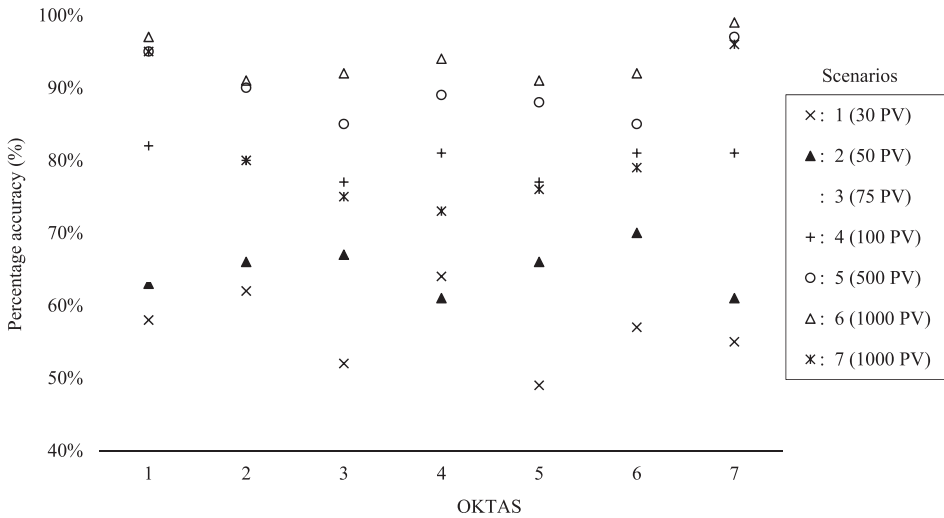


Figure 9. Percentage accuracy (%) of the estimated cloud cover per okta per scenario.

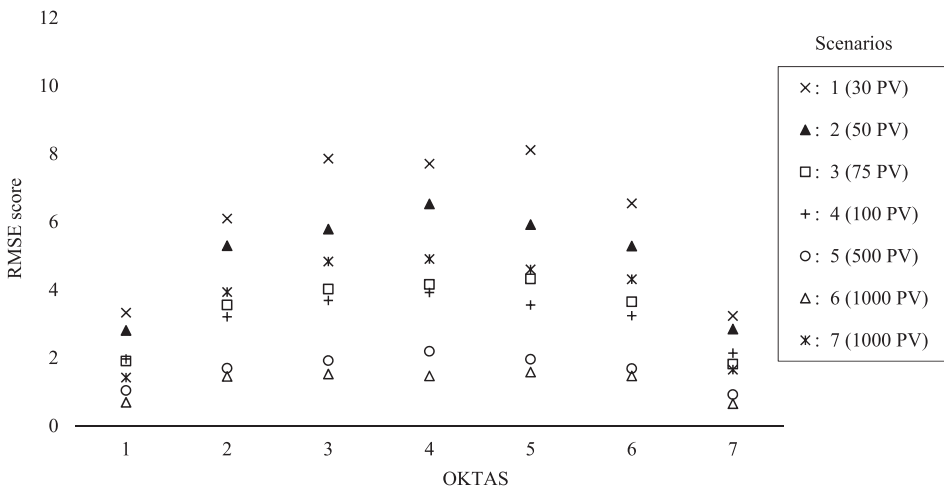


Figure 10. Root Mean Square Error on the percentage cloud cover estimation per okta per scenario.

1000 PV stations, where the error score comparing the simulated cloud cover with the estimated one, is very low, regardless of the cloud cover.

Our findings suggest that a number between 500 and 1000 photovoltaics distributed in an area of 10 km×10 km can be used to provide interesting meteorological information such as cloud cover without the need of any additional equipment or meteorological data. The results obtained from the methodology used in this study have shown that no more than 1000 PV stations are needed to produce the most accurate cloud cover estimation.

The primary objective of this study was to answer the question if photovoltaics can be used to estimate cloud cover. We believe that this question has been answered through the results presented here and this study will become the basis for further development of an algorithm that will be used to predict cloud movement in real time from PV data. Nevertheless, there are some limitations that need to be overcome in order to have a more complete algorithm that will include missing important variables and provide more accurate results. First of all, we assume PVs to be binary (shaded or not)

and as part of our research we currently working on predicting the real time decrease on PV energy production without exogenous information. Furthermore, in our methodology we have only calculated the cloud shade during the time of the day when the sun is at its highest point and the solar irradiance is perpendicular to the clouds surface, thus the model can provide accurate cloud cover estimations only the time of the day when the sun is at its highest point (noon).

If we use the present methodology at a different time of the day the resulting estimation will have a large deviation from the reality because the shade from the clouds will be deformed by the angle of the sun. The solar angle based on the geographic location, season and time of the day must be calculated within the model to produce accurate estimation of the cloud movement based on the cloud shadows during the daytime. PV performance issues such as the natural degradation of the solar cells over time, physical damage or panel discoloration, soiling due to dusty environments or inverter issues were not modelled at the present stage and the functionality of the PVs under study assumed to be uniform, thus not affecting the accuracy and reliability of cloud cover estimation. All these variables must be taken into account and determine the way in which they may affect the resulting cloud cover estimations. Furthermore, in combination with real electricity production data, the aforementioned variables can be correlated with the different types of clouds.

Conclusions

This paper proposes a new methodology for cloud cover estimation using nothing more than measurements taken from photovoltaics that can be used continuously in real time. The methodology proposed in this study utilises geoprocessing to calculate the cloud cover over an area based on whether a PV is covered by shade or not. For practical reasons the method considers the time of the day when the sun is at its highest point and the solar irradiance is perpendicular to the clouds surface, thus the cloud shadow is assumed to be equal to the size of the cloud.

To identify the optimum number of PV stations needed for an accurate cloud cover estimation, the model was tested in an area of 10 km×10 km under six different scenarios using 30, 50, 75, 100, 500 and 1000 randomly distributed residential PV stations throughout the study area and one using 1000 PV stations over the same area but excluding parks, roads and waterways where no PV stations were placed.

Fractal-based cloud shadows were generated using a modified version of the midpoint displacement algorithm and used as the observation data in the model. The estimated cloud cover from each scenario was then compared with the simulated cloud cover (fractal shadows) and the optimum scenario was selected based on the results.

As was shown in the results, the model exhibits great dependence on the number of PV stations used each time. As the number of PV stations increases, the accuracy of the cloud cover estimation also increases, while the error decreases. Starting with an average accuracy at 57% using 30 PV stations in the first scenario, the model produced a 90% accuracy in scenario 5 and 94% in scenario 6 with 500 and 1000 PV stations respectively. Scenario 7, where no PV stations distributed in parks, roads and waterways produced a less accurate cloud cover estimation in relation to scenario 6, with the same number of PV stations used, with an average accuracy of 82%. It seems that when PV stations are not distributed throughout the whole area we get lower accuracies, but good enough to have a solid cloud cover estimation.

Given the fact that a sporadic populated area of 10 km × 10 km consists of more than 30,000 houses such as, Limassol district in Cyprus (Statistical Service of the Republic of Cyprus 2011) and that it is not unreasonable to assume that 3% of the houses will have PVs on their rooftops, we believe that PV data without the need of any additional sophisticated and expensive equipment can provide interesting meteorological information such as cloud cover.

Based on our findings, either 1-second instantaneous or 1-minute average cloud cover estimations would produce the same results. A low number of PV stations (such as 30 PV stations) would

however not produce accurate results. For example, if for 1 oktas the cloud cover is ‘stationary’ over an area with no shaded PVs, the predicted cloud cover will be zero in both cases.

Future work includes the validation of the results from the methodology using real PV electricity production and meteorological data, the calculation of the cloud cover due to the solar angle based on the geographic location, season and time of day, and finally the development of an algorithm to predict cloud movement in real time from PV data is in progress.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the SOLAR-ERA.NET Cofund Joint Call/Cyprus Research Promotion Foundation (CRPF) [grant number KOINA/SOLAR-ERA.NET/1216/0014].

ORCID

Stavros Stylianou  <http://orcid.org/0000-0003-3457-1154>

References

- Ackerman, S. A., and S. K. Cox. 1981. “Comparison of Satellite and All-Sky Camera Estimates of Cloud Covering During Gate.” *Journal of Applied Meteorology* 20: 581–587. doi:10.1175/1520-0450(1981)020<0581:COASAAS>2.0.CO;2.
- Adaramola, M. S. 2016. “Distribution and Temporal Variability of the Solar Resource at a Site in South-East Norway.” *Frontiers in Energy* 10: 375–381. doi:10.1007/s11708-016-0426-6.
- Cai, C., and D. C. Aliprantis. 2013. “Cumulus Cloud Shadow Model for Analysis of Power Systems With Photovoltaics.” *IEEE Transactions on Power Systems* 28: 4496–4506. doi:10.1109/TPWRS.2013.2278685.
- Clothiaux, E. E., G. G. Mace, T. P. Ackerman, T. J. Kane, J. D. Spinhirne, V. S. Scott, E. E. Clothiaux, et al. 1998. “An Automated Algorithm for Detection of Hydrometeor Returns in Micropulse Lidar Data.” *Journal of Atmospheric and Oceanic Technology* 15: 1035–1042. doi:10.1175/1520-0426(1998)015<1035:AAAFDO>2.0.CO;2.
- Dybbroe, A., K.-G. Karlsson, and A. Thoss. 2005. “NWCSAF AVHRR Cloud Detection and Analysis Using Dynamic Thresholds and Radiative Transfer Modeling. Part I: Algorithm Description.” *Journal of Applied Meteorology* 44: 39–54. doi:10.1175/JAM-2188.1.
- Ela, E., V. Diakov, E. Ibanez, and M. Heaney. 2013. “Impacts of Variability and Uncertainty in Solar Photovoltaic Generation at Multiple Timescales.” *National Renewable Energy Laboratory*. Accessed December 19, 2018. <https://www.nrel.gov/docs/fy13osti/58274.pdf>.
- Falayi, E. O., and A. B. Rabiu. 2011. “Estimation of Global Solar Radiation Using Cloud Cover and Surface Temperature in Some Selected Cities in Nigeria.” *Archives of Physics Research* 2: 99–109.
- Fang, D. 2009. “The Study of Terrain Simulation Based on Fractal.” *WSEAS Transactions on Computers* 8: 133–142. Accessed December 19, 2018. <http://www.wseas.us/e-library/transactions/computers/2009/31-860.pdf>.
- Gul, T., and Stenzel, T. 2005. “Variability of Wind Power and Other Renewables, Management Options and Strategies.” Paris: IEA Publications.
- Heinle, A., A. Macke, and A. Srivastav. 2010. “Automatic Cloud Classification of Whole Sky Images.” *Atmospheric Measurement Techniques* 3: 557–567. doi:10.5194/amt-3-557-2010.
- Kazantzidis, A., P. Tzoumanikas, A. F. Bais, S. Fotopoulos, and G. Economou. 2012. “Cloud Detection and Classification with the Use of Whole-sky Ground-based Images.” *Atmospheric Research* 113: 80–88. doi:10.1016/j.atmosres.2012.05.005.
- Li, Z., L. Moreau, and A. Arking. 1997. “On Solar Energy Disposition: A Perspective From Observation and Modeling.” *Bulletin of the American Meteorological Society* 78: 53–70. doi:10.1175/1520-0477(1997)078<0053:OSEDAP>2.0.CO;2.
- Madhavan, B. L., H. Deneke, J. Witthuhn, and A. Macke. 2017. “Multiresolution Analysis of the Spatiotemporal Variability in Global Radiation Observed by a Dense Network of 99 Pyranometers.” *Atmospheric Chemistry and Physics* 17: 3317–3338. doi:10.5194/acp-17-3317-2017.

- Martins, C., J. Wilson, P. Zablowski, J. Baumgardner, J. L. Aballay, B. Garcia, P. Rastori, and L. Otero. 2013. "A New Method to Estimate Cloud Cover Fraction Over El Leoncito Observatory From an All-Sky Imager Designed for Upper Atmosphere Studies." *Publications of the Astronomical Society of the Pacific* 125: 56–67. doi:10.1086/669255.
- Martins, F. R., M. P. Souza, and E. B. Pereira. 2003. "Comparative Study of Satellite and Ground Techniques for Cloud Cover Determination." *Advances in Space Research* 32: 2275–2280. doi:10.1016/S0273-1177(03)90554-0.
- Perez, R., M. David, T. E. Hoff, M. Jamaly, S. Kivalov, J. Kleissl, P. Lauret, and M. Perez. 2016. "Spatial and Temporal Variability of Solar Energy." *Foundations and Trends in Renewable Energy* 1: 1–44. doi:10.1561/27000000006.
- Ricciardelli, E., F. Romano, and V. Cuomo. 2008. "Physical and Statistical Approaches for Cloud Identification Using Meteosat Second Generation-Spinning Enhanced Visible and Infrared Imager Data." *Remote Sensing of Environment* 112: 2741–2760. doi:10.1016/j.rse.2008.01.015.
- Sakellariou, N. K., and H. D. Kambezidis. 2004. "Cloud Cover in the Athens Area." *Fresenius Environmental Bulletin* 13: 66–68.
- Saupe, D. 1988. "Algorithms for Random Fractals." In *The Science of Fractal Images*, edited by Heinz-Otto Peitgen and Dietmar Saupe, 71–136. New York: Springer.
- Seiz, G., E. P. Baltasvias, and A. Gruen. 2002. "Cloud Mapping From the Ground: Use of Photogrammetric Methods." *Photogrammetric Engineering & Remote Sensing* 68(9):941–951.
- Silva, A. A., and M. P. Souza-Echer. 2016. "Ground-based Observations of Clouds Through Both an Automatic Imager and Human Observation." *Meteorological Applications* 23: 150–157. doi:10.1002/met.1542.
- Statistical Service of the Republic of Cyprus. "Census of Population 2011 – Volume II Households and Housing Units," Republic of Cyprus, Nicosia (2011).
- Tapakis, R., and A. G. Charalambides. "System and Method for Predicting Solar Power Generation," International (PCT) Patent Application No. PCT/EP2016/055889 (Patent pending), (2016).
- Tapakis, R., and A. G. Charalambides. 2013. "Equipment and Methodologies for Cloud Detection and Classification: A Review." *Solar Energy* 95: 392–430. doi:10.1016/j.solener.2012.11.015.
- Wang, Z., and K. Sassen. "An Improved Cloud Classification Algorithm Based on the SGP CART Site Observations," In: Proceedings of the 14th ARM Science Team Meeting, Albuquerque, New Mexico, 1–11 (2004). Accessed 19 December 2018. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.648.9416&rep=rep1&type=pdf>.
- Werkmeister, A., M. Lockhoff, M. Schrempf, K. Tohsing, B. Liley, and G. Seckmeyer. 2015. "Comparing Satellite- to Ground-Based Automated and Manual Cloud Coverage Observations – A Case Study." *Atmospheric Measurement Techniques* 8: 2001–2015. doi:10.5194/amt-8-2001-2015.
- World Meteorological Organization (WMO), "Manual on Codes (WMO-No. 306)," WMO, Switzerland. 2015. Accessed 19 December 2018. http://www.wmo.int/pages/prog/www/WMOCodes/WMO306_v11/Publications/2015update/306_vol_I1_2015_en.pdf.
- Yamashita, M., M. Yoshimura, and T. Nakashizuka. 2004. "Cloud Cover Estimation Using Multitemporal Hemisphere Imageries." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 35: 826–829.

Appendices

Appendix A. Midpoint displacement algorithm

```

N ← 2 ^ maxLevel
delta ← sigma
F([1, N + 1], [1, N + 1]) ← sigma * randn(2)
d = N / 2;
For stage = 1 to maxLevel do
    /*From grid Type I to Type II*/
    delta ← delta*(0.5^(H*0.5))
    x ← d + 1: 2 * d:N - d + 1
    y ← d + 1: 2 * d:N - d + 1
    F(x, y) ← (F(x + d, y + d) + F(x + d, y - d) + F(x - d, y + d) + F(x - d, y - d))
    / 4 + delta * randn([length(x) length(y)])
    /*From grid Type II to Type I*/
    delta ← delta*(0.5^(H*0.5))
    x ← d + 1: 2 * d:N - d + 1
    y ← size(x)
    F(x, 1) ← (F(x + d, 1) + F(x - d, 1) + F(x, d)) / 3 + delta * randn(y)
    F(x, N + 1) ← (F(x + d, N + 1) + F(x - d, N + 1) + F(x, N + 1 - d)) / 3 + delta * randn(y)

```



```

F(1, x) ← (F(1, x + d) + F(1, x - d) + F(d, x)) / 3 + delta * randn(y)
F(N + 1, x) ← (F(N + 1, x + d) + F(N + 1, x - d) + F(N + 1 - d, x)) / 3
+ delta * randn(y)
x ← d + 1: 2 * d:N - d + 1
y ← 2 * d + 1: 2 * d:N - d + 1
F(x, y) ← (F(x + d, y) + F(x - d, y) + F(x, y + d) + F(x, y - d))
/ 4 + delta * randn([length(x) length(y)])
F(y, x) ← (F(y + d, x) + F(y - d, x) + F(y, x + d) + F(y, x - d))
/ 4 + delta * randn([length(x) length(y)])
d ← d / 2
end For
    
```

Appendix B. Confusion matrices for the evaluation of the per-okta accuracy of the estimated cloud cover for each scenario

Table A1. Confusion matrix for the evaluation of the per-okta accuracy of the estimated cloud cover using 30 PVs when compared with the simulated one.

		Simulated cloud cover in oktas – 30 PV stations								
		0	1	2	3	4	5	6	7	8
Predicted cloud cover in oktas	0	—	37%							
	1		58%	20%						
	2		5%	62%	16%					
	3			17%	52%	12%	1%			
	4				31%	64%	35%	1%		
	5					24%	49%	21%		
	6						16%	57%	5%	
	7							21%	55%	
	8								39%	—

Note: Percentage indicates the number of correctly identified cloud cover in oktas, over the simulated one.

Table A2. Confusion matrix for the evaluation of the per-okta accuracy of the estimated cloud coverage using 50 PVs when compared with the simulated one.

		Simulated cloud cover in oktas – 50 PV stations								
		0	1	2	3	4	5	6	7	8
Predicted cloud cover in oktas	0	—	31%							
	1		63%	24%						
	2		6%	66%	23%					
	3			9%	67%	27%				
	4				11%	61%	12%			
	5					12%	66%	10%		
	6						22%	70%	4%	
	7							20%	61%	
	8								35%	—

Note: Percentage indicates the number of correctly identified cloud cover in oktas, over the simulated one.

Table A3. Confusion matrix for the evaluation of the per-okta accuracy of the estimated cloud coverage using 75 PVs when compared with the simulated one.

		Simulated cloud cover in oktas – 75 PV stations								
		0	1	2	3	4	5	6	7	8
Predicted cloud cover in oktas	0	—	33%							
	1		64%	10%						
	2		4%	80%	12%					
	3			10%	79%	14%				
	4				9%	76%	11%			
	5					10%	79%	9%		
	6						10%	82%	2%	
	7							9%	63%	
	8								35%	—

Note: Percentage indicates the number of correctly identified cloud cover in oktas, over the simulated one.

Table A4. Confusion matrix for the evaluation of the per-okta accuracy of the estimated cloud coverage using 100 PVs when compared with the simulated one.

		Simulated cloud cover in oktas – 100 PV stations								
		0	1	2	3	4	5	6	7	8
Predicted cloud cover in oktas	0	—	15%							
	1		82%	11%						
	2		3%	80%	13%					
	3			8%	77%	10%				
	4				10%	81%	12%			
	5					9%	77%	9%		
	6						12%	81%	3%	
	7							10%	81%	
	8								16%	—

Note: Percentage indicates the number of correctly identified cloud cover in oktas, over the simulated one.

Table A5. Confusion matrix for the evaluation of the per-okta accuracy of the estimated cloud coverage using 500 PVs when compared with the simulated one.

		Simulated cloud cover in oktas – 500 PV stations								
		0	1	2	3	4	5	6	7	8
Predicted cloud cover in oktas	0	—	3%							
	1		95%	7%						
	2		2%	90%	9%					
	3			3%	85%	4%				
	4				6%	89%	6%			
	5					6%	88%	4%		
	6						6%	85%	2%	
	7							11%	97%	
	8								1%	—

Note: Percentage indicates the number of correctly identified cloud cover in oktas, over the simulated one.

Table A6. Confusion matrix for the evaluation of the per-okta accuracy of the estimated cloud coverage using 1000 PVs when compared with the simulated one.

		Simulated cloud cover in oktas – 1000 PV stations								
		0	1	2	3	4	5	6	7	8
Predicted cloud cover in oktas	0	—								
	1		97%	5%						
	2		3%	91%	5%					
	3			4%	92%	4%				
	4				3%	94%	5%			
	5					2%	91%	3%		
	6						3%	92%	1%	
	7							5%	99%	
	8									—

Note: Percentage indicates the number of correctly identified cloud cover in oktas, over the simulated one.

Table A7. Confusion matrix for the evaluation of the per-okta accuracy of the estimated cloud coverage using 1000 PVs distributed in areas excluding parks, roads and waterways, when compared with the simulated one.

		Simulated cloud cover in oktas – 1000 PV stations (excluding parks, roads and waterways)								
		0	1	2	3	4	5	6	7	8
Predicted cloud cover in oktas	0	—								
	1		95%	9%						
	2		5%	80%	14%					
	3			12%	75%	19%				
	4				11%	73%	11%			
	5					9%	76%	15%		
	6						13%	79%	4%	
	7							7%	96%	
	8									—

Note: Percentage indicates the number of correctly identified cloud cover in oktas, over the simulated one.