

Characterization of surface solar-irradiance variability using cloud properties based on satellite observations



Takeshi Watanabe^{*}, Yu Oishi, Takashi Y. Nakajima

Research and Information Center, Tokai University, Tokyo, Japan

ARTICLE INFO

Article history:

Received 2 June 2016

Received in revised form 25 October 2016

Accepted 26 October 2016

Available online 4 November 2016

Keywords:

Variability of surface solar irradiance

Cloud property

Discriminant analysis

ABSTRACT

The variation in surface solar irradiance (SSI) on short timescales has been investigated previously in relation to ground-based observations. Such results are limited to the locality of the observation stations, leading to insufficient knowledge about the spatial distribution of variation features. We propose a method for characterizing variations in SSI using cloud properties obtained from satellite observations. Datasets of cloud properties from satellite observation and SSI from ground-based observation are combined at simultaneous observation points to investigate their relations. The SSI variations are classified statistically into six categories. The cloud properties related to the categorized variation features are then analyzed. From such relations, a statistical discriminant method is used to design a classifier to assign a category to the SSI variation over an area from the cloud properties obtained by satellite observation. The accuracy of classification and feature selection is discussed.

© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Solar energy is expected to be part of the solution to the problem of global warming. Variation in solar irradiance at ground level causes fluctuation in the power output from solar power systems, which is a disadvantage of generating power that way. This work focuses on variation over timescales of no more than a few hours, which is caused mainly by clouds. The effects of aerosol and water vapor are also important, but these contribute primarily to slower variation over more than a few hours. Variation in surface solar irradiance (SSI) occurs in two ways: interception by clouds between observation stations and the sun, and reflection and scattering by cloud particles.

Observation using ground-based equipment is the main method for obtaining temporal resolutions shorter than a few minutes. An advantage of ground observations is that they can allow continuous high-temporal-resolution data at a single position. However, they are disadvantaged by their narrow (and thus limited) field of view. In contrast, satellite observations provide a large field of view, but the frequency of observations over a single location is lower than that with ground-based observation, and spatial resolutions are also coarser. However, satellite observations also provide information about cloud properties. Combining ground and satel-

lite observations should therefore be a good way to investigate the relation between clouds and SSI.

Some metrics relevant to SSI are used to analyze its short-term variation. Lave and Kleissl (2010) and Lave et al. (2012) analyzed the ramp rate (RR) to investigate geographic smoothing effects. The RR is defined as the change in magnitude of solar irradiance over a given period. Tomson and Tamm (2006) investigated the stability of SSI by using absolute values of its increments for given periods. Woyte et al. (2007) applied wavelet spectrum analysis to classify fluctuations in solar irradiance. Watanabe et al. (2016) used three metrics—the mean, standard deviation, and sample entropy—to evaluate regional features of variation in SSI over Japan.

The relation between SSI and clouds has also been investigated using metrics related to SSI. These studies are based fundamentally on measurements of solar irradiance at ground level integrated with cloud effects. Duchon and O'Malley (1999) used a 21-min window mean of solar-irradiance data with 1-min resolution and the corresponding standard deviation to develop a method for classifying cloud type according to these two metrics. Orniśi et al. (2002) also proposed cloud classification using metrics similar to those used by Duchon and O'Malley (1999) and improved the classification accuracy. Martínez-Chico et al. (2011) performed cloud classification by considering an index for direct solar irradiance at the ground. Their index is defined as the ratio of direct solar irradiance to extraterrestrial irradiance. Pages et al. (2003) classified cloud type using temperature, wind speed, and air relative humidity data in addition to solar-irradiance data.

^{*} Corresponding author at: Research and Information Center, Tokai University, 2-28-4 Tomigaya, Shibuya-ku, Tokyo 151-0063, Japan.

E-mail address: nabetake@ees.hokudai.ac.jp (T. Watanabe).

Previous work has thus deepened our understanding of SSI variation. However, the results of these studies are based mainly on analyses using ground-based observation data of the area around observation stations, leading to insufficient knowledge about the spatial distribution of variation features. This work aims at filling such gaps. We first investigate the relation between SSI variation and cloud properties from satellite observations. We then characterize the SSI variability by cloud properties. By applying such relations, we propose a method for estimating the variability of SSI using cloud properties as retrieved from satellite observation. The spatial distribution of the variability will contribute to new understanding of the surface solar variation and aid the development of applications to solar energy engineering. For example, the operators of a grid system could anticipate likely regions of strong variability and consider alternative operational measures. Seasonal and regional features of SSI variation would also be useful support information when planning to construct solar power plants.

We used the Moderate Resolution Imaging Spectroradiometer (MODIS) cloud products for the analysis in this study. Cloud properties from MODIS data are available for long periods. However, only one or two images can be obtained in a day for any particular location. This is a disadvantage in solar energy engineering. Recently a new-generation geostationary meteorological satellite, HIMAWARI-8 of the Japan Meteorological Agency (JMA), was launched and is now in service (Bessho et al., 2016). Other such satellites (e.g., GOES-R of NOAA/NASA, Meteosat Third Generation (MTG) of EUMETSAT) are scheduled for launch in the next few years (Mohr, 2014). These will have more observation bands and higher

observation frequencies than previously launched geostationary satellites. Abundant information about cloud, aerosol, and solar irradiance will be obtained from geostationary meteorological satellite observation. At present, practical applications based on using MODIS data in solar energy engineering may be limited, but we expect that in the future the proposed approach can be applied to cloud products based on geostationary satellite observation.

The remainder of this paper is structured as follows. Sections 2 and 3 describe our data and methods, respectively. Section 4 describes the processing of data from ground- and satellite-based observations for analysis. Section 5 discusses cloud properties in relation to variations in SSI. Section 6 develops a method for classifying SSI variability that is designed using statistical discriminant methods. Section 7 discusses and summarizes this work.

2. Data

2.1. Surface solar irradiance

We use existing SSI data from Japan. The JMA maintains ground-based observation stations and performs quality control and routine maintenance of their equipment. Solar irradiance is defined as the total radiation measured over 1 min of data sampled at 10 s intervals, and its temporal interval is 1 min. Pyranometers were replaced at most stations in the middle of 2011 (Ohtake et al., 2015). Forty-seven observation sites are selected based on availability of solar irradiance data for the five years from 2010 to 2014. Data mainly from six observation stations in the Kanto region, which is on the Pacific side of eastern Japan (Fig. 1) are ana-

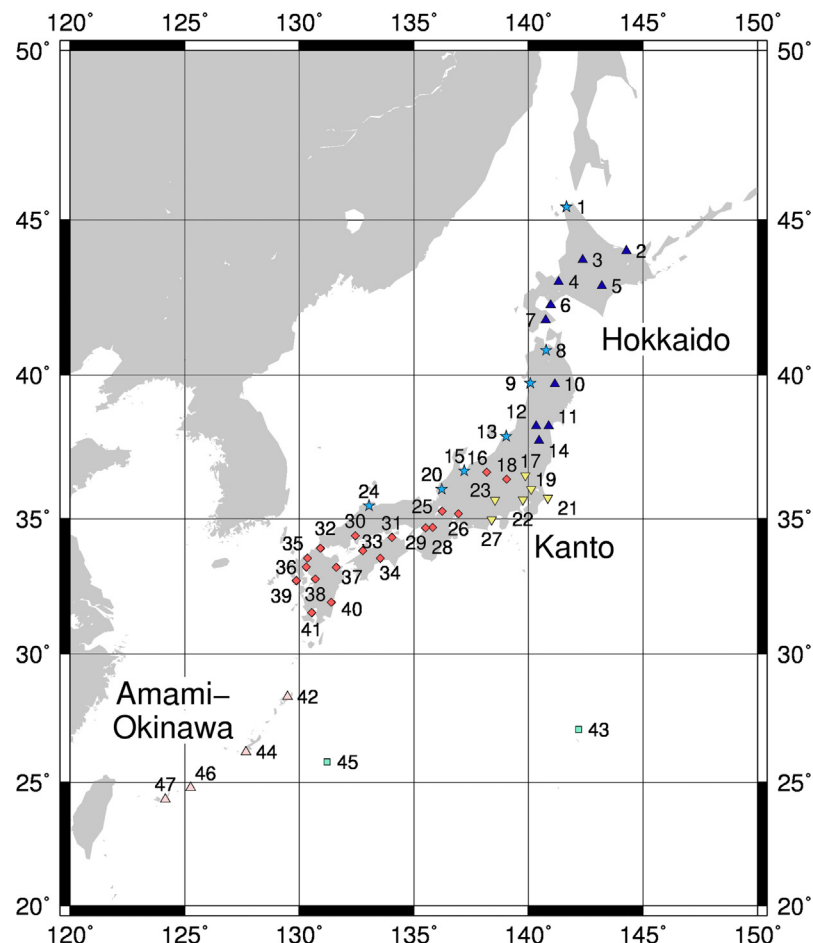


Fig. 1. Observation stations throughout Japan. Stations 17, 19, 21, 22, 27, and 27 are located in the Kanto region. Colors and marks indicate classes with similar variation features, as determined by Watanabe et al. (2016).

lyzed. Watanabe et al. (2016) classifies these stations into the same cluster based on similarity of variation features of SSI.

In this study, we analyze variation on a 2-h timescale while simultaneously analyzing the solar irradiance data at some stations. For such analyses, diurnal trends and latitudinal effects on the magnitude of the SSI are removed, so the clearness index (CI) as defined by Woyte et al. (2007) is used. The CI at time t is defined as the ratio of the observed SSI (I_g) to the downward shortwave irradiance at the top of the atmosphere (I_t):

$$CI(t) = I_g(t)/I_t(t).$$

Here, I_t is calculated as

$$I_t(t) = I_0 E(t) \cos Z(t, l),$$

where $I_0 = 1367 \text{ W/m}^2$ is the solar constant (Iqbal, 1983), $E(t)$ is the eccentricity factor at time t , and $Z(t, l)$ is the solar zenith angle at time t and latitude l . The value of CI indicates the availability of SSI at a given time and location.

2.2. Cloud properties

We use spatially distributed cloud properties based on MODIS observations. Level-2 MODIS cloud products from the Terra (MOD06) and Aqua (MYD06) polar-orbital satellites are available for the same period as the JMA SSI data, from 2010 to 2014. Collection 6 (the latest dataset) is selected, improving the accuracy of the algorithm used to detect clouds beyond that of previous datasets (Platnick et al., 2015a,b). Both satellites make daytime observations of Japan, with Terra passing over eastern Japan at roughly 11:00 local time and Aqua at 13:00.

The cloud properties used in the analysis are cloud fraction (FR), cloud top height in pressure level (CTH), cloud optical thickness (COT), and the effective cloud particle radius (ER). The FR data have a 5-km spatial resolution. Pixel locations can be obtained from the same data file as that for the L2 cloud product. The COT, CTH, and ER data are for a grid with 1-km spatial resolution. The location of a pixel on this grid is obtained from the L1 MODIS product (MOD03 and MYD03). Information on the cloud mask is also obtained from the MODIS cloud product. The cloud mask data indicate whether a given view of the earth surface is unobstructed by cloud or thick aerosol, expressed in four levels of confidence regarding whether a pixel is regarded as cloudy, uncertain/probably cloudy, probably clear, or clear. Lower confidence levels are associated with pixels of cirrus cloud, snow and ice cover, and the edges of cloudy regions (Ackerman et al., 2010).

3. Methods

3.1. Cluster analysis: k -means method

The k -means method is a major nonhierarchical clustering method (Hartigan and Wong, 1979; Wilks, 2011). In the k -means method, M points in N dimensions are divided into k clusters so that the within-cluster sum of squares is minimized. The clustering algorithm requires k initial cluster centers, which are randomly determined. Next, each point is placed into its nearest cluster, based on the Euclidian distance between the point and the cluster center. Cluster centers are then updated, and each point is reassigned to the closest updated cluster. This procedure is repeated until no points require reassignment. The “stats” package of the R software (R Development Core Team, 2015) is used to perform k -means cluster analysis.

The number k of clusters to use is determined by maximizing the Calinski–Harabasz pseudo- F statistic (Calinski and Harabasz, 1974). This statistic is given by the formula

$$\text{Pseudo-}F = (A/W)[(n - k)/(k - 1)],$$

where A and W are the among- and within-cluster variances, respectively, n is the number of objects, and k is the number of existing clusters.

3.2. Multiple discriminant method

A statistical discriminant method is used to develop discriminant functions, also called a classifier, from training data (Wilks, 2011). The training data used in this work are composed from more than two classes, and the multiple discriminant method is then performed. The performance of the discriminant method is affected by factors such as the sample size and the normality of the sample distribution (Bayne et al., 1983; Lachenbruch et al., 1973). It varies between models even with the same training data. Three types of discriminant method are used: Fisher's linear and quadratic discriminant methods, and the linear logistic discriminant method.

We assume that the covariance of each pair of classes is the same for the linear discriminant analysis and is different for the quadratic discriminant analysis. The Mahalanobis distance is used as the distance between a point and the mean of a class in these discriminant methods. Points are classified into classes according to closeness of mean. The quadratic method has the advantage of providing a more detailed classification. The logistic discriminant is based on the logistic regression function. This method is considered robust for various underlying distributions (Bayne et al., 1983). The “MASS” (Venables and Ripley, 2002) and “nnet” (Venables and Ripley, 2002) packages of the R data analysis software are used for calculation.

To evaluate the performance of the classifier, two correct-answer ratings are used as a measure of accuracy: the overall rate of correct answers (defined as the number of correctly classified points divided by the total number of points) and the class-level mean of the rate of correct answers (defined as the simple average of rates of correct answers in each class).

There are other classifications based on recently developed mathematical methods, such as neural networks and machine learning (Tapakis and Charalambides, 2013). Although these methods have some advantages, classical statistical discrimination methods are selected because they simply and clearly reflect the features of physical properties. Section 5 describes the discriminant analysis.

3.3. Textural features

Textural features are used to evaluate the spatial distribution features of cloud pixels. As shown previously (Ameur et al., 2004; Haralick et al., 1973), textural features are useful for cloud detection and cloud-type classification. Five textural features are selected: angular second moment (ASM), contrast (CNT), correlation (CRR), entropy (ENT), and local homogeneity (LHM). Texture features are defined following Haralick et al. (1973). Descriptions of these variables are as follows.

ASM: measure of image homogeneity

CNT: measure of the contrast or amount of local variation present in the image

CRR: grayscale linear dependencies in the image

ENT: measure of image randomness

LHM: similarity of adjacent gray tones

COT is an important factor affecting the SSI magnitude. We assume that the spatial distribution of COT is related to temporal fluctuation of the SSI. Texture features are computed using the

base-10 logarithm of COT in the definition domain, which ranges from -2 to 2 because COT for cloudy pixels ranges from 0.01 to 100.0 . The COT of the pixel assigned to clear is 0 or not defined, but clear pixels in the domain have to be assigned with values in order to compute the textural features. Therefore, a clear pixel is treated as a cloudy pixel with the minimum COT value of 0.01 .

Because textural features are based on the relation between grayscale values at two nearest neighboring grid points, textural features are functions of the azimuthal angle between two grids. Azimuthal angles of 0° , 45° , 90° , and 135° are selected, providing four types of averaged textural feature.

3.4. Metrics of variation in solar irradiance

To investigate the variation in SSI, its features are evaluated using metrics. Watanabe et al. (2016) used the mean, standard deviation, and sample entropy (Pincus, 1991; Richman and Moorman, 2000) to evaluate features of SSI variation. These metrics represent the availability of solar irradiance, strength of variation, and manner of fluctuation, respectively. Sample entropy is a metric that represents time-series complexity. When sample entropy increases, CI fluctuates at higher frequency.

3.5. Cloud confidence index

The confidence of cloud detection in the defined domain is evaluated using MODIS cloud mask data. The index, called the cloud confidence index (CCI), is defined as the ratio of the number of pixels categorized as uncertain/probably cloudy to the total number of pixels categorized as either of cloudy and uncertain/probably cloudy. A larger CCI value indicates that more pixels are assigned to clouds that are detected at lower confidence levels in the domain.

4. Processing of the simultaneous observation dataset

To investigate the relation between SSI variation and cloud properties, the dataset is prepared from ground-based and satellite observations for the five years from 2010 to 2014. A simultaneous observation is defined as one for which the MODIS sensor made an observation over the ground-based observation station. A simultaneous observation point is characterized by three variation metrics of SSI, four cloud properties, and five textural features from the MODIS observation.

The temporal window and spatial domain have to be determined to compute the variation metrics and cloud properties, respectively. Considering cloud movement and cloudy areas, SSI variation in the given period is related to not only clouds over the observation station but also those over the entire domain. Therefore, a temporal window that provides sufficient length to calculate the three variation metrics is first determined. Approximately 100 points are enough to obtain a significant value for sample entropy (Richman and Moorman, 2000). The temporal window is determined as 121 min, and its center is at the simultaneous observation point. The spatial domain is determined as a domain of about 45×45 km, and its center is located at the observation station, considering the speed of synoptic-scale disturbances. In mid-latitude over Japan, disturbances tend to move eastward at about 10 km/h (see Chang et al., 2002). We assume that clouds accompany the synoptic disturbance. Clouds within 20 km of the observation station probably cross the path between the observation station and the sun for a period of 2 h, which causes SSI variation. The three cases of 25, 45, and 65 km are analyzed using the three steps discussed below, and the results do not change the conclusions. This does not mean that the selection of the spatial and

temporal domains for the satellite and ground data is not important. The movement and size of synoptic disturbances vary daily. In this study, we treat different synoptic weather conditions and weather in different seasons in the same manner. Hence, selection of the spatial domains may not influence the results in this work. Several cloud types are likely to be present simultaneously in the domain. Because area-averaged cloud properties are used, we do not make cloud type analysis the central focus of this work. In addition, we assume clouds to be single-layered, hence multilayered clouds are not distinguished. Note that not every cloud in the definition domain affects the SSI at an observation station. Processing of the simultaneous observation dataset involves the following three steps.

- (1) Cloud properties and textural features are averaged over areas.

Each cloud-property variable is averaged over the domain centered at the ground-based observation station. To compute the area-averaged COT, CTH, and ER, only data on grids containing clouds are used. The area-averaged FR is computed using all grids over the domain. If the domain is perfectly cloud-free, a simultaneous observation point is not defined. Textural features are computed using COT over the domain.

- (2) Variation in surface solar irradiance is characterized.

A local time series is obtained from a 121-min window in the CI time-series. Three variation metrics—the mean, standard deviation, and sample entropy—are computed from the local time series. Fig. 2 shows a three-dimensional plot of the variation metrics of simultaneous observation points.

- (3) Simultaneous observation points are categorized by the k -means method applied to three-dimensional variation features

The simultaneous observation points are categorized according to variation features. The Calinski–Harabasz pseudo-F statistic (Fig. 3) is used to determine how many categories should be used to characterize the simultaneous observation points. A local maximum of the index is seen in the 4- and 6-cluster cases. Although the pseudo-F statistics in the 4-cluster case are larger than those in the 6-cluster case, the 6-cluster case is selected because a more detailed categorization is useful for understanding the features of SSI variation. The simultaneous observation points are thus divided into six variation feature categories, C_1 to C_6 .

It is assumed that nearby points of variation features have similar cloud properties. To obtain a clear relation between SSI variation and cloud properties, points that are far from the class mean are removed. These outliers are removed according to the criterion that the Mahalanobis distance from the class mean must be less than 1.5. This threshold was selected subjectively. To judge whether this threshold is fit for analysis, Hotelling's T^2 statistic was used to check whether the cloud properties are equal, checking all pairs of variation classes. We consider the cloud properties to be related to SSI variation wherever the test shows a difference.

5. Results

5.1. Categorization of variation in surface solar irradiance

Fig. 2 shows the resulting clusters considering variation features, and Table 1 summarizes the number of simultaneous points in each class. Fig. 4 shows part of the time series of the CI in August 2011 at observation stations across the Kanto region as an example

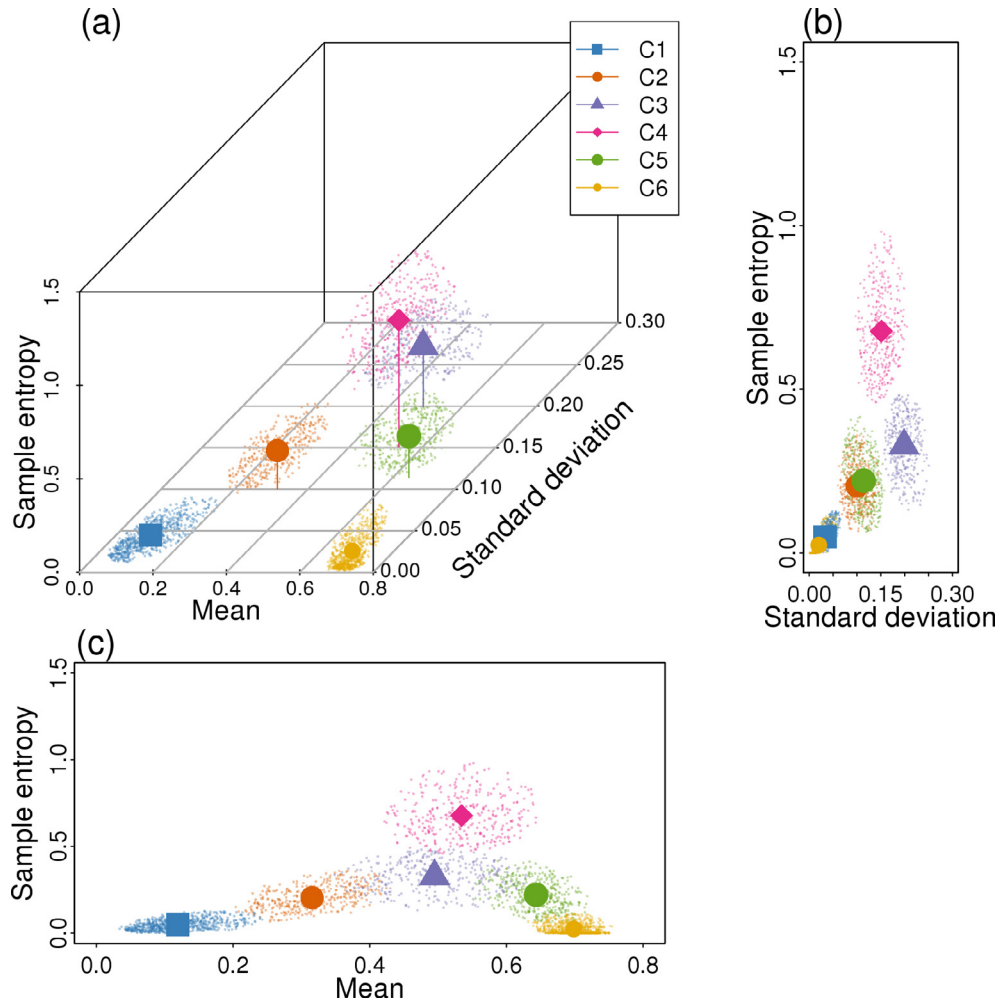


Fig. 2. (a) Simultaneous plot of three variation metrics. Colors represent the resultant class as classified by the k-means method. Large marks represent the class center; (b) and (c) are two-dimensional diagrams of the standard-deviation-sample entropy, and mean-sample entropy, respectively.

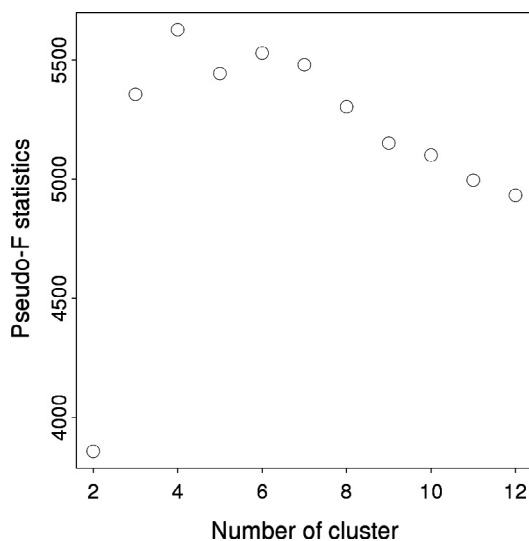


Fig. 3. Calinski-Harabasz pseudo-F statistic for number of clusters for the k-means method.

of the clustering results. Although the clustering is mathematically determined, each cluster shows distinctive variation features (Figs. 2 and 4). There are three main clusters, each with two sub-

clusters: those with small (C_1 and C_2), moderate (C_3 and C_4), and large (C_5 and C_6) mean CI. Sub-clusters C_1 and C_2 have small solar-irradiance availability. The variability of C_1 is smallest because its standard deviation and sample entropy are small. The standard deviation of C_2 is relatively large, while its sample entropy is relatively small, indicating that solar irradiance varies strongly with longer period. The magnitude of CI in C_3 and C_4 is moderate, and their standard deviations are large. The difference between these two classes is the sample entropy. C_3 has smaller sample entropy, and so the SSI variation fluctuates strongly with a longer period, while the variation of C_4 is strong and rapid. C_6 corresponds to clear or almost clear conditions. C_5 also has high solar-irradiance availability, but it is more variable than that of C_6 .

5.2. Cloud properties related to variation in surface solar irradiance

Cloud properties associated with variation features can be clarified according to the results of the above cluster analysis. Fig. 5 shows the distribution of each cloud property in each class using a boxplot diagram (see McGill et al. (1978) and Wilks (2011) for a description of the boxplot diagrams as used here). Each cloud property is standardized using its mean and standard deviation. The null hypothesis that the cloud properties of two classes are equal is rejected for all pairs of classes at the 1% significance level or better, suggesting that variation features in the SSI are related to cloud properties from satellite observations with moderate spatial

Table 1
Number of simultaneous points in each class.

Stage of analysis procedure	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	Total
Original (<i>k</i> -means)	1228	901	822	743	998	1092	5784
Outlier	703	370	342	339	405	712	2871
CCI	473	262	221	248	190	163	1557

The numbers in the “Original” row are obtained after *k*-means classification analysis.

The numbers in the “Outlier” row are obtained after filtering based on the Mahalanobis distance as mentioned in Section 4.

The numbers in the “CCI” row are obtained after filtering based on the CCI mentioned in Section 5.

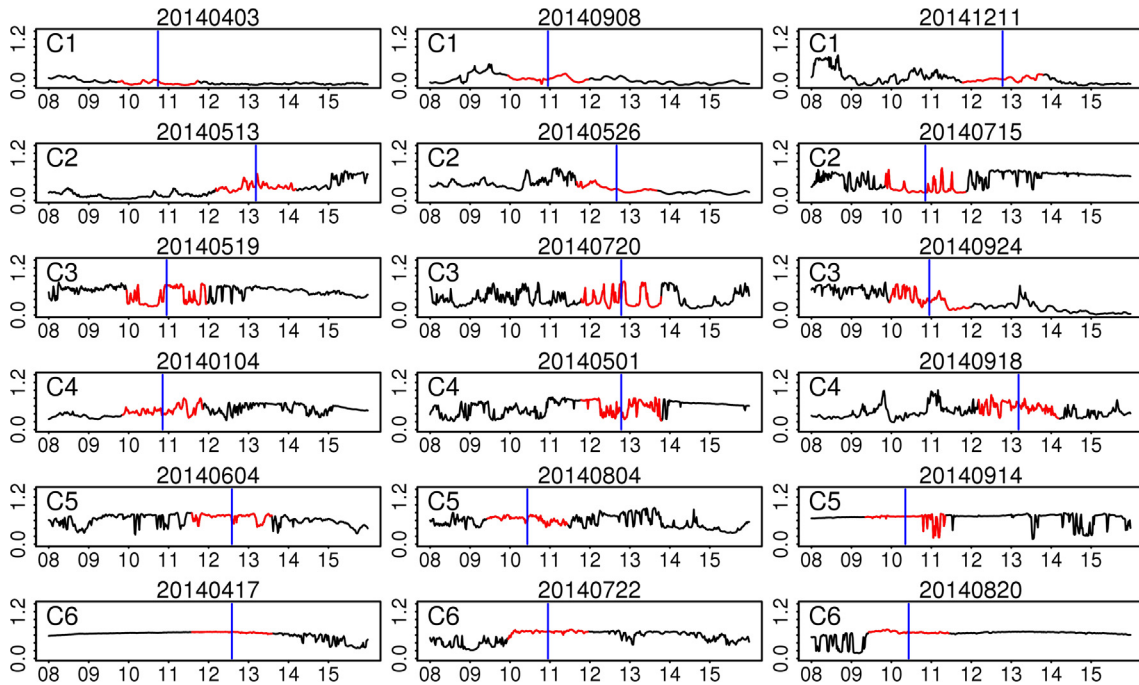


Fig. 4. Partial time series of CI at observation station 22 (Tokyo) in 2014: horizontal axis represents hours in Japan Standard Time (JST). Blue lines represent the simultaneous observation points. Red lines are the local time series within the 121-min temporal window. Characters C₁–C₆ (top-left of each panel) indicate the variation class.

resolution. However, note that it is difficult to justify assuming a normal distribution for some cloud properties. For example, the FR distributions in C₁–C₄ are clearly skewed toward larger values. The cloud properties of each variation class in Fig. 5 are summarized as follows.

- C₁ corresponds to overcast skies with whole-sky thick cloud cover because COT and FR are largest of all classes. Small CNT indicates that clouds cover the whole area, although the CI variability is small.
- C₂ also corresponds to overcast skies, but the COT is smaller than in C₁. The spatial distribution of C₂ tends to be more disordered and less homogenous than in C₁, which is judged from the LHM and ENT of textural features. This causes more variability in C₂ than in C₁. We note that C₁ and C₂ are seen in the inner regions of vast areas of thick cloud (Fig. 6).
- C₃ and C₄ correspond to moderate CI, so it is reasonable to conclude that COT is also moderate. The remarkable feature of these two classes is that the CNT is large. Hence, these two classes tend to be seen at the margins of optically thick cloudy areas or at the boundary between cloudy and cloud-free areas (Fig. 6). The cloud properties of these two classes are similar, but several differences are seen. We note that C₃ has lower CTH and smaller FR, while C₄ has larger ER and ENT (smaller LHM). The variation of C₄ is characterized as a larger sample entropy, which indicates stronger fluctuations with higher fre-

quencies. Such variation features are related to an open and unordered cloud distribution that includes broken clouds of various sizes (Martínez-Chico et al., 2011). Clouds smaller than the spatial resolution of MODIS cannot be correctly resolved. However, it seems that the cloud properties from MODIS observations reflect such disordered spatial distribution of cloud in C₄.

- C₅ has cloud properties that are intermediate between clear and other cloudy classes, and is characterized as having large CI and moderate CI variability. The COT is smaller than for other cloudy classes and FR is not small, so it seems that C₅ contains cloudy skies with optically thin clouds. The range of distribution of cloud properties of C₅ tends to be wide, so features of cloud properties in C₅ are somewhat unclear.
- C₆ corresponds to clear or almost clear sky, where FR and COT are small. The textural features of C₆ may not be meaningful because there are fewer clouds in the definition domain.

5.3. Robustness of the relation between variation in surface solar irradiance and cloud properties

The distributions of some cloud properties are widely spread and skewed. In addition, some outliers are seen. Such distributions may cause the relation between SSI variation and cloud properties to be unclear and unstable. We consider one of the causes for such distributions to be the confidence level of cloud detection. The

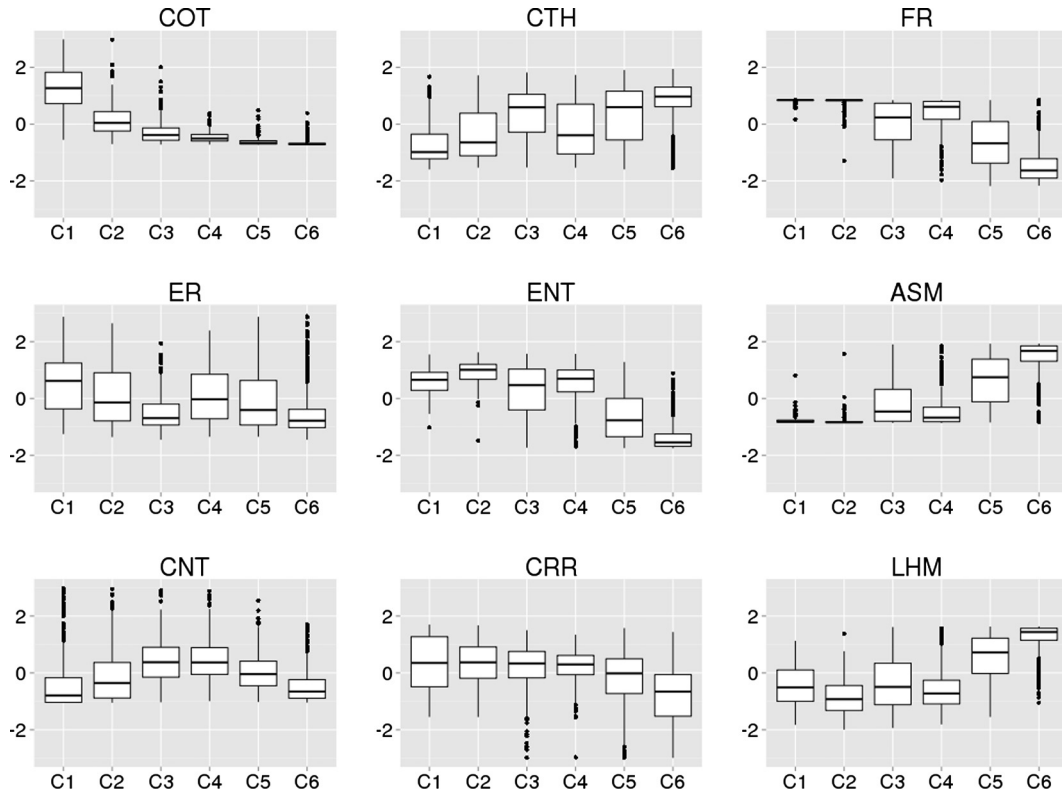


Fig. 5. Distribution of cloud properties in each class. The horizontal line in each box represents the median. The upper and lower box sides are defined as the 25th and 75th percentiles, respectively. The upper (resp., lower) whisker is plotted at the highest (resp., lowest) point at $+1.5$ (resp., -1.5 IQR) times the upper side (resp., lower side). Points represent outliers.

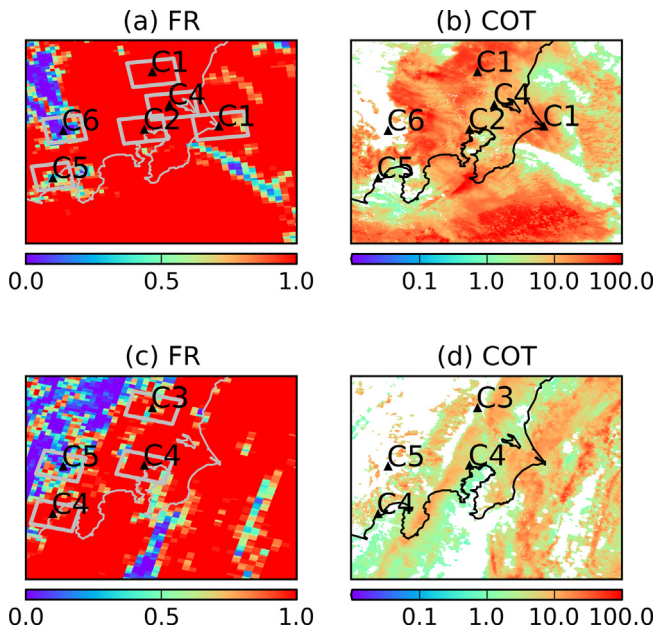


Fig. 6. Cloud properties and variation classes, (a) and (c) show FR and (b) and (d) show COT. Filled triangles represent observation stations. Gray rectangles represent the defined domain. White areas in (b) and (d) indicate pixels assigned to clear, where COT is 0 or not defined, (a) and (c) are drawn from MODIS/Aqua cloud product (21 April 2012) and (c) and (d) from MODIS/Terra (17 June 2013).

robustness of the relation is verified using the CCI index (defined in Section 3.5).

Fig. 7 shows the distributions of cloud data for CCI below 0.25. Although this cutoff is selected subjectively, a null hypothesis stat-

ing that the cloud properties of two classes are equal is rejected for all pairs at significance levels of 1% or better. Most variables show a shift of their median and a reduction in outliers after this filtering procedure (Figs. 5 and 7). The distributions of cloud properties in C_3 – C_6 show particularly marked changes. The reduction in outliers suggests an increasing robustness of the relation between SSI variation and cloud properties. The discussion below focuses on changes in FR because this is related directly to the cloud mask. The distribution ranges of C_3 and C_4 are reduced and the medians are shifted to larger values, as are the medians in C_5 and C_6 . These changes are due to the removal of points with smaller FR. This result suggests that a high confidence of cloud detection is useful for finding a robust relationship between SSI variation and cloud properties. A large reduction in the number of points due to this filtering is seen in C_5 and C_6 , but there is less reduction in C_1 and C_2 (Table 1), which correspond to overcast skies with thick clouds. Referring to Duchon and O'Malley (1999) and Ornisi et al. (2002), one of the major cloud types corresponding to C_5 and C_6 is cirrus. Specific cloud types may thus be filtered out by using the above approach.

6. Classification of variability in surface solar irradiance according to cloud properties as observed by satellite

The results in the previous section indicate that we can predict which category the SSI variation over the area belongs to from the cloud properties as obtained from satellite observations. A classifier to do so is designed and its performance is discussed below.

6.1. Classifier design

The classifier is designed using Fisher's linear and quadratic discriminant methods and the linear logistic discriminant method.

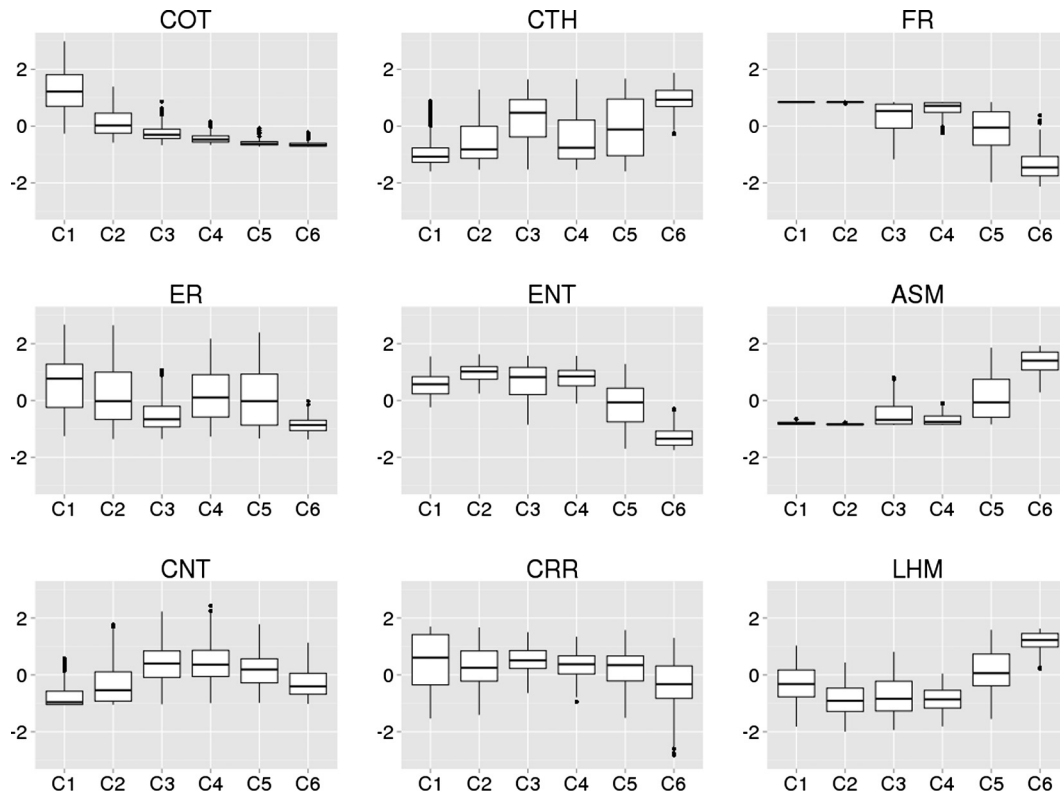


Fig. 7. Same as Fig. 5 but after filtering based on the CCI criterion.

Table 2
Classifier performance using training data.

			Results						Rate of correct answer
		Total	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	
Training	C ₁	473	403/435	69/38	0/0	1/0	0/0	0/0	0.852/0.920
	C ₂	262	17/36	209/203	5/8	31/15	0/0	0/0	0.798/0.775
	C ₃	221	0/0	42/29	94/105	43/56	40/26	2/5	0.425/0.475
	C ₄	248	0/0	45/29	55/34	137/166	11/19	0/0	0.552/0.669
	C ₅	190	0/0	11/4	27/19	47/45	64/97	41/25	0.337/0.511
	C ₆	163	0/0	0/0	0/6	0/0	19/19	138/138	0.883/0.847

Numbers before and after the solidus are from Fisher's linear discriminant method and the linear logistic discriminant method, respectively.

The training data come from a simultaneous observation dataset covering the five years from 2010 to 2014 at six observation stations in the Kanto region. The discussion in the previous section suggests that a raw simultaneous observation dataset will be too noisy for classifier design. The dataset is therefore pre-processed to provide training data. First, data for which the CCI exceeds 0.25 are removed, and then outliers are removed.

6.2. Classifier validation (performance)

The performance of the classifier is evaluated using training data and all simultaneous observation data. This approach to the use of training data is known to bias the outcome toward higher accuracy. Table 2 summarizes the results of classification in the case of the training data. Classifiers using the quadratic discriminant method cannot be defined because the covariance matrix for C₁ becomes singular. The overall rates of correct answers for Fisher's linear and the linear logistic discriminant method are 0.675 and 0.735, respectively, and the class-level mean of the rate of correct answers are 0.641 and 0.699. The results of classification for C₁, C₂, and C₆ show higher hit rates. In contrast, C₃, C₄, and C₅ are difficult to classify accurately and confusion often occurs

between neighboring classes. This is possibly because neighboring classes tend to have similar cloud properties. In addition, the spatial distribution of cloud properties varies continuously. There are often different cloud types present simultaneously in the defined domain. The disadvantage of this classification procedure is that cloud motion and migration of cloudy regions are not considered. Thus, it is difficult to identify which cloud properties dominate at an observation station in a 2-h temporal window from snapshot-like satellite observations alone.

Table 3 summarizes the results of classification in the case of all simultaneous observation data. The overall rates of correct answers for Fisher's linear and the linear logistic discriminant methods are 0.627 and 0.664, respectively, and the average rates of correct answers is 0.560 and 0.608. The accuracy thus declines in each case compared with that of the training data, partly because of the lower confidence of cloud detection.

6.3. Feature selection

Although various features are useful for investigating cloud properties in detail, all features may not be necessary for satisfactory classification. A classifier that uses fewer features is expected

Table 3

Classifier performance using all simultaneous observation points.

			Results						Rate of correct answer
	Total		C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	
Training	C ₁	703	549/614	99/77	31/9	23/2	1/1	0/0	0.781/0.873
	C ₂	370	37/57	227/219	53/57	51/32	1/4	1/1	0.614/0.592
	C ₃	342	4/2	48/34	114/131	59/68	63/53	54/54	0.333/0.383
	C ₄	339	0/0	47/31	72/49	150/182	48/67	22/10	0.442/0.537
	C ₅	405	0/0	11/4	37/38	49/48	116/184	192/131	0.286/0.454
	C ₆	712	1/1	3/1	11/18	3/2	49/114	645/576	0.906/0.809

Numbers before and after the solidus are from Fisher's linear discriminant method and the linear logistic discriminant method, respectively.

Table 4

Feature selection by number of features.

Number of features	Accuracy	Selected features
3	0.689	COT, FR, ENT
4	0.696	COT, FR, ENT, LHM
5	0.703	COT, FR, ER, ENT, LHM
6	0.701	COT, CTH, FR, ER, CNT, LHM
7	0.706	COT, CTH, FR, ER, ENT, CRR, LHM
8	0.702	COT, CTH, FR, ENT, ASM CNT, CRR, LHM

to function with better robustness against noise and to be easier to compute.

Feature selection is performed in a simple way. A classifier is designed using a subset of features chosen from among the nine cloud features, and performance is evaluated using the training data. This procedure is repeated for all possible combinations of cloud properties. The average rate of correct answers is used to measure accuracy. It is assumed that a classifier with higher accuracy is designed with more suitable features.

Accuracy higher than 70% is maintained when more than four features are chosen (Table 4). The accuracy has peaks at six and seven features, likely because of reduced redundancy of the training data. Four cloud properties—COT, FR, CTH, and ER—are good features for classification, although CTH and ER have the lower priority of those four variables. As indicated in Fig. 5, a textural feature represents a positive or negative relation with the others. For example, ENT is negatively correlated with LHM. To reduce data redundancy, it is better to select the minimum number of textural variables or compress the original data into lower-dimensional data (Ameur et al., 2004). COT, FR, and ENT seem to be the most important variables for classification because these are selected for all cases.

7. Discussions and conclusions

To compensate for the disadvantages of ground-based observation, we proposed a method for predicting the variability of SSI. Fig. 8 shows the spatial distribution of variation categories classified using a classifier designed with the linear logistic discriminant method and seven features (COT, CTH, FR, ER, ENT, CRR, and LHM). The spatial distribution and the extent of variation categories can be found from this figure. The classifier worked adequately over the Kanto region (the black rectangle in Fig. 8) although adequacy could not be ensured when the classifier was applied to other regions.

For practical use in solar engineering, a general classifier that can be applied to the whole region of a satellite image should be developed. However, the method proposed in this work is still at the stage of feasibility testing for such a goal because several important problems remain. One is that of regional features of the relation between SSI variation and cloud properties. According to Watanabe et al. (2016), features of the SSI variation differ between regions in Japan (Fig. 1). Table 5 compares the accuracies

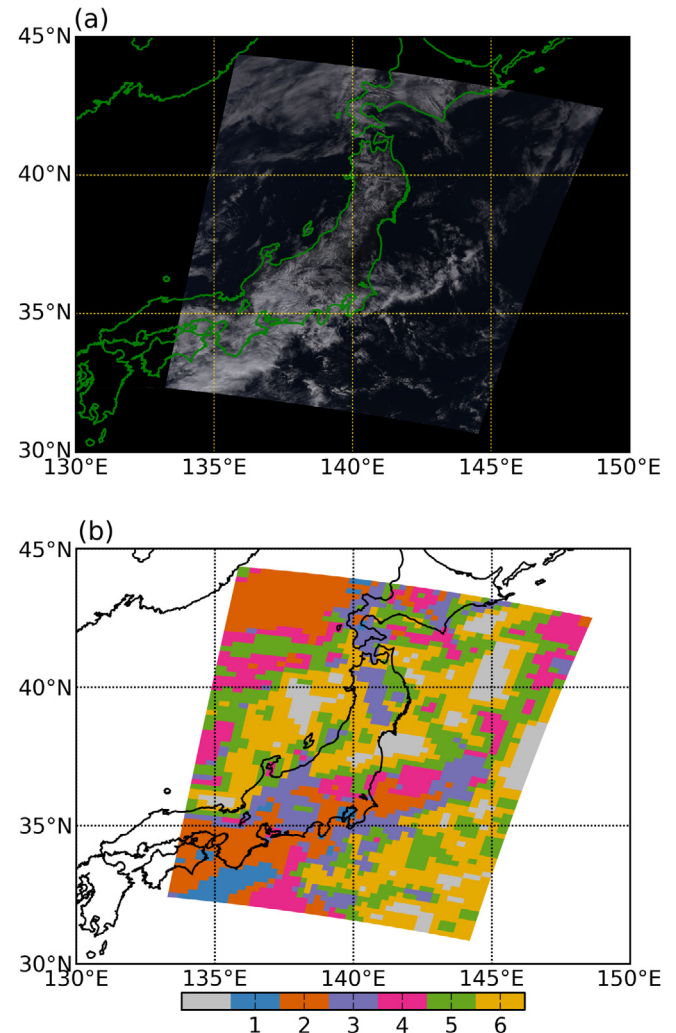


Fig. 8. (a) True-color composite from MODIS/Terra L1B products at 1:30 UTC on 24 February 2011. (b) Spatial distribution of variation classes as classified from cloud properties. Colors represent variation classes corresponding to Fig. 2. Gray represents cloud-free areas. Classification is performed for two thirds of the image. Cloud properties are obtained from the MODIS/Terra L2 cloud products for 1:30 UTC on 24 February 2011.

Table 5

Comparison between classifiers.

Test data	Training data	Accuracy
Hokkaido	Hokkaido	0.671
Kanto	Kanto	0.611
Amami–Okinawa	Amami–Okinawa	0.632
	Kanto	0.578

Accuracy is measured using the mean of the rate of correct answers in each class.

of classifiers designed using different training data. Using the same procedure as above, classifiers were designed based on simultaneous observation points over the Hokkaido (Stations 2–7) and Amami–Okinawa (Stations 42, 44, 46, and 47) regions. The accuracies for both test datasets were significantly reduced when a classifier for the Kanto region was used.

The classification accuracy is not particularly high. There are several possible solutions for improving the classifier. More cloud property and irradiance variation features should be evaluated and tested. This work used three variation metrics, but several variation metrics were proposed (see the Introduction). The selection of metrics that better characterize variability and cloud properties would result in better associations between clouds and SSI variation. The effect of cloud-detection confidence was discussed in Section 5. Low confidence causes inconsistency between data from ground-based and satellite observations. Improved cloud detection, especially of thin clouds and the edges of cloudy regions, is thus also desired. We suggest that multilayered clouds should be distinguished from single-layer clouds because it seems that multilayered clouds affect the variation in solar irradiance in a different way than single-layer clouds do. In addition, the retrieval of ER of multilayered clouds tends to be influenced by the assumption of single-layer clouds (Wind et al., 2010). There are also other classification methods that were not investigated in this work. More suitable classification methods should be chosen after more testing.

We suggest that the proposed method could be applied to every area in which ground-based observations of solar irradiance are made. The relation between SSI variation and cloud properties differs between regions. Hence, a classifier designed with the proposed approach needs to be determined for each region. Whether it is better to design classifiers globally or regionally is an important and interesting question. To answer this question, a clearer understanding of the relation between cloud and SSI variability is necessary. Nevertheless, a globally designed classifier or a classification algorithm that can be applied everywhere would be useful for solar energy engineering.

Cloud properties from MODIS observations were used in this work. Hence, practical use of this approach for solar energy engineering is limited. Newer geostationary satellites, such as Himawari-8 and -9, have more observation bands and can generate much more information about cloud properties (Bessho et al., 2016). This will allow us to measure the variability of SSI continuously on shorter timescales.

Acknowledgement

The Terra and Aqua/MODIS Level-2 cloud products datasets were acquired from the Level-1 & Atmosphere Archive and Distribution System (LAADS) Distributed Active Archive Center (DAAC), located in the Goddard Space Flight Center in Greenbelt, Maryland (<https://ladsweb.nascom.nasa.gov/>). This work was partly supported by the Japan Science and Technology Agency through the CREST/EMS funding program.

References

Ackerman, S., Frey, R., Strabala, K., Liu, Y., Gumley, Baum, L., Menzel, P., 2010. Discriminating Clear-Sky From Cloud With MODIS - Algorithm Theoretical Basis

- Document (MOD35). MODIS Cloud Mask Team. Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin, Madison.
- Ameur, Z., Ameur, S., Adane, A., Sauvageot, H., Bara, K., 2004. Cloud classification using the textural features of Meteosat images. *Int. J. Remote Sensing* 25, 4491–4503.
- Bayne, C.K., Beauchamp, J.J., Kane, V.E., 1983. Assessment of Fisher and logistic linear and quadratic discrimination models. *Comput. Stat. Data Anal.* 1, 257–273.
- Bessho, K. et al., 2016. An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. *J. Meteor. Soc. Jpn.* 94, 151–183. <http://dx.doi.org/10.2151/jmsj.2016-009>.
- Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. *Commun. Stat.* 3, 1–27. <http://dx.doi.org/10.1080/03610927408827101>.
- Chang, E.K.M., Lee, S., Swanson, K.L., 2002. Storm track dynamics. *J. Climate* 15, 2163–2183.
- Duchon, C.E., O'Malley, M.S., 1999. Estimating cloud type from pyranometer observation. *J. Appl. Meteor.* 38, 132–141.
- Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textual features for image classification. *IEEE Trans. Syst., Man, Cybernetics*, SMC-3 6, 610–621.
- Hartigan, J.A., Wong, M.A., 1979. A K-means clustering algorithm. *Appl. Stat.* 28, 100–108.
- Iqbal, W., 1983. An introduction to solar radiation. Academic Press, Oxford.
- Lachenbruch, P.A., Sneeringer, C., Revo, L.T., 1973. Robustness of the linear and quadratic discriminant function to certain types of non-normality. *Commun. Stat.* 1, 39–56.
- Lave, M., Kleissl, J., 2010. Solar variability of four sites across the state of Colorado. *Renewable Energy* 35, 2867–2873.
- Lave, M., Kleissl, J., Arias-Castro, E., 2012. High-frequency irradiance fluctuations and geographic smoothing. *Sol. Energy* 86, 2190–2199.
- Martínez-Chico, M., Batlles, F.J., Bosch, J.L., 2011. Cloud classification in a mediterranean location using radiation data and sky images. *Energy* 36, 4055–4062.
- McGill, R., Tukey, J.W., Larsen, W.A., 1978. Variations of box plots. *Am. Stat.* 32, 12–16.
- Mohr, T., 2014. Preparing the use of new generation geostationary meteorological satellite. *WMO Bull.* 63, 42–44.
- Ohtake, H., Fonseca Jr., J.G.S., Takashima, T., Oozeki, T., Shimose, K., Yamada, Y., 2015. Regional and seasonal characteristics of global horizontal irradiance forecasts obtained from the Japan meteorological agency mesoscale model. *Sol. Energy* 116, 83–99. <http://dx.doi.org/10.1016/j.solener.2015.03.020>.
- Ornisi, A., Tomasi, C., Calzolari, F., Nardino, M., Cacciari, A., Geogiadi, T., 2002. Cloud cover classification through simultaneous ground-based measurements of solar and infrared radiation. *Atmos. Res.* 61, 251–275.
- Pages, D., Calbo, J., González, J.A., 2003. Using routine meteorological data to derive sky conditions. *Ann. Geophys.* 21, 649–654.
- Pincus, S.M., 1991. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. U.S.A.* 88, 2297–2301.
- Platnick, S. et al., 2015a. MODIS atmosphere L2 cloud product (06_L2). In: NASA MODIS Adaptive Processing System. Goddard Space Flight Center, USA. http://dx.doi.org/10.5067/MODIS/MOD06_L2.006.
- Platnick, S. et al., 2015b. MODIS atmosphere L2 cloud product (06_L2). In: NASA MODIS Adaptive Processing System. Goddard Space Flight Center, USA. http://dx.doi.org/10.5067/MODIS/MYD06_L2.006.
- R Development Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Richman, J.S., Moorman, J.R., 2000. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039–H2049.
- Tapakis, R., Charalambides, A.G., 2013. Equipment and methodologies for cloud detection and classification: a review. *Sol. Energy* 95, 392–430.
- Tomson, T., Tamm, G., 2006. Short-term variation of solar radiation. *Sol. Energy* 80, 600–606.
- Venables, W.N., Ripley, B.D., 2002. *Modern Applied Statistics with S*. Springer, New York.
- Watanabe, T., Takamatsu, T., Nakajima, T., 2016. Evaluation of variation in surface solar irradiance and clustering of observation stations in Japan. *J. Appl. Meteor. Climatol.* 55, 2165–2180.
- Wilks, D.S., 2011. *Statistical Methods in the Atmospheric Science*. Academic Press, Oxford.
- Wind, G. et al., 2010. Multilayer cloud detection with the MODIS near-infrared water vapor absorption band. *J. Appl. Meteor. Climatol.* 49, 2315–2333.
- Woyte, A., Belmans, R., Nijs, J., 2007. Fluctuation in instantaneous clearness index: analysis and statistics. *Sol. Energy* 81, 195–206.