RESEARCH ARTICLE

Correspondence to:
D. Pozo-Vázquez,
dpozo@ujaen.es

Automatic Cloud-Type Classification Based On the Combined Use of a Sky Camera and a Ceilometer

J. Huertas-Tato[1] [iD], F. J. Rodríguez-Benítez[2], C. Arbizu-Barrena[2] [iD], R. Aler-Mur[1] [iD], I. Galvan-Leon[1] [iD], and D. Pozo-Vázquez[2] [iD]

[1]EVANNAI Research Group, Department of Computing Science, University Carlos III, Madrid, Spain, [2]MATRAS Research Group, Department of Physics, University of Jaén, Jaén, Spain

Abstract A methodology, aimed to be fully operational, for automatic cloud classification based on the synergetic use of a sky camera and a ceilometer is presented. The random forest machine learning algorithm was used to train the classifier with 19 input features: 12 extracted from the sky camera images and 7 from the ceilometer. The method was developed and tested based on a set of 717 images collected at the radiometric stations of the Univ. of Jaén (Spain). Up to nine different types of clouds (plus clear sky) were considered (clear sky, cumulus, stratocumulus, nimbostratus, altocumulus, altostratus, stratus, cirrocumulus, cirrostratus, and cirrus) plus an additional category multicloud, aiming to account for the frequent cases in which the sky is covered by several cloud types. A total of eight experiments was conducted by (1) excluding/including the ceilometer information, (2) including/excluding the multicloud category, and (3) using six or nine different cloud types, aside from the clear-sky and multicloud category. The method provided accuracies ranging from 45% to 78%, being highly dependent on the use of the ceilometer information. This information showed to be particularly relevant for accurately classifying "cumuliform" clouds and to account for the multicloud category. In this regard, the camera information alone was found to be not suitable to deal with this category. Finally, while the use of the ceilometer provided an overall superior performance, some limitations were found, mainly related to the classification of clouds with similar cloud base height and geometric thickness.

Plain Language Summary The different cloud types are the results of different atmospheric processes. In addition, cloud types have a varied interaction with the solar radiation. Therefore, cloud monitoring, have interest in a varied of fields, ranging from the study of the atmospheric thermodynamic processes to solar energy. So far, cloud monitoring is conducted based on human observation, making cloud type databases scarce and, in general, low reliable. A procedure for automatic cloud classification is conducted here using information from a sky camera and a ceilometer. The information derived from these two instrument is showed provide an enhanced performance.

1. Introduction

Scientific interest in retrieving cloud information dates many decades back and was mainly related to civil and military aviation. The attention to cloud information has grown in the framework of climate studies, since clouds play a key role in Earth energy balance (Li et al., 2014; Wild et al., 2013). More recently, in the field of weather forecasting, the improvement in cloud representation has emerged as a significant research field. Mainly, because clouds are involved in multiple and strong interactions, their misrepresentation may have large impacts and implications in the atmospheric dynamics and, then, in the accuracy of the simulations of the numerical weather prediction models (Haiden et al., 2015; Pincus et al., 2011). Lastly, the growing penetration of the solar energy around the world has fostered a great interest in cloud information, since clouds are the main source of variability of the solar energy (Martínez-Chico et al., 2011; Mateos et al., 2014; Tzoumanikas et al., 2016). In all the previous fields of science, the establishment of proper, accurate, and cheap cloud monitoring systems is crucial (World Meteorological Organization (WMO), 2012). Accurate and consistent cloud observations, which are globally standardized, remain an important need (WMO, 2017). Nevertheless, the type of cloud information (cloud parameters, temporal and spatial resolution, etc.) needed greatly varies depending on the application. In some of the above mentioned applications, information about the type of cloud is crucial. Human-reported information was the first available continuous source of information on cloud type. But the high associated cost, the low accuracy, and issues such as the representativeness

make this source of information under menace in many countries. The use of satellite retrieval for cloud classification is a promising tool because of their spatial coverage. For instance, the "cloud-type" product of EUMETSAT Application Facility on Climate Monitoring or the APOLLO project (Kriebel et al., 2003; Weya & Schroedter-Homscheid, 2014) reports operationally a coarse cloud classification. Nevertheless, the performance of these operational cloud-type monitoring systems is still limited due to limitations of the satellite platforms.

The other alternative is the use of ground-based sky cameras systems. These systems, which basically date one decade back, are now considered the reference for cloud cover estimates (Boers et al., 2010; Cazorla et al., 2008; Long et al., 2006). More recently, the automatic recognition of cloud types has emerged as possible product of these instruments.

There are two basic steps for automatic cloud classification. First, the extraction from the camera images of appropriate and distinctive information of the different sky conditions and cloud types. To this end, different features can be computed on the information from camera channels. Particularly, these features account for characteristics such as cloud shape, texture, or the color of the sky/clouds. Second, once a set of distinctive features are obtained, cloud classification relies on the use of automatic classification algorithms. Ultimately, these algorithms are trained and tested with human-supervised cloud-type databases.

The type and number of features have increased enormously in the last years, benefiting from other fields of research, such as automatic pattern recognition. For instance, Calbo and Sabburg (2008) used texture properties and the Fourier transform of the camera visible channels to classify up to eight classes of sky conditions. The methodology achieved an accuracy of about 62%. Heinle et al. (2010) proposed the use of a combined set of textural and color features for the classification of up to seven cloud types, with a classification success rate of about 75%. Rumi et al. (2013) proposed the use of features from the infrared channels of a camera, obtaining an accuracy of 90% in the estimation of towering cumulus and cumulonimbus cloud types. Kazantzidis et al. (2012) proposed the use of a multicolor criterion on sky images, showing an average performance of about 87% using seven cloud categories. Kliangsuwan and Heednacram (2015) used a new methodology, based on the fast Fourier transform, for feature extraction for cloud classification. The overall accuracy of this methodology was shown to be 90% for the automatic classification of seven clouds types. Wacker et al. (2015) used, as ancillary information for cloud classification, the measured longwave radiation. They reported an improvement of up to 10%, compared to the use of just the sky camera information. The reported mean accuracy ranged from 80 to 90%. Cheng and Yu (2015) have proposed a cloud classification method based on division of the image in different blocks. In this way the authors were able to account for mixed clouds types in one image, obtaining an improved classification accuracy. Recently, Li et al. (2016) used a novel approach for cloud-type recognition, based on the analysis of image as a collection of patches, rather than a collection of pixels. The method showed an accuracy of 90% for five classes of sky conditions.

Regarding classification machine learning algorithms, the literature contains proposals ranging from artificial neural networks (Kliangsuwan & Heednacram, 2015; Lee et al., 1990; Singh & Glennen, 2005), to k-nearest neighbor (KNN) (Cheng & Yu, 2015; Heinle et al., 2010; Kazantzidis et al., 2012; Wacker et al., 2015) and support vector machines (SVM) (Schmidt et al., 2015; Taravat et al., 2015; Zhen et al., 2015). ANNs are a commonly machine learning technique used in cloud classification. It is actually a nonlinear regression technique that can be used for classification by setting a threshold on the output(s). Typically, standard architectures with three layers are used (input/hidden/output) and in a multiclass classification context, like cloud classification, there are as many output neurons as classes. KNN does not need to fit a model to the data; rather, it stores all data and classifies new instances by looking for the closest stored data instance(s). KNN does not require any adaptation for multiclass problems. The basic version of KNN may suffer more than other methods when there are many features, or some of them are irrelevant. It is also very slow for real use if the data set is large. However, there are methods for KNN that can improve both accuracy (like Weinberger & Saul, 2009) and speed (like kd-trees, Wess et al., 1993). SVMs aim to maximize the generalization capabilities by finding separation boundaries between classes that maximize the margin. They have fewer local minima issues compared to ANNs, because SVMs solve a constrained convex optimization problem, with a single global optimum. On the other hand, the most common approach to SVM trains binary classifiers, hence requiring to train as many models as classes (one-versus-rest approach) or as many as pairs of classes

(one-versus-one approach), although there are also approaches that deal with multiple classes directly (Crammer & Singer, 2001).

To sum up, performance of the different approaches and studies for automatic cloud classification varies greatly and can hardly be compared due to several reasons: different cameras, different cloud data sets, different time representativeness, different experimental setups and evaluation methods, and different cloud classes.

Aside from the sky cameras, the use of ceilometers for cloud property retrieval has emerged in the last few years (Illingworth et al., 2007). Ceilometers are single-wavelength low-powered lidars (light detection and ranging), which can provide high-frequency observations of cloud profiles, including parameters such as the cloud base height (CBH) and cloud top height (e.g., cloud cover). Unlike satellite imagery, which generally provides low-reliability CBH estimates at relatively low temporal resolution (very few samples per hour), ceilometers are able to provide an accurate description of the location of the cloud vertical boundaries with even several samples per minute (Arbizu-Barrena et al., 2015; Costa-Suros et al., 2014; Martucci et al., 2010; Viúdez-Mora et al., 2015).

In this work, we propose and evaluate a methodology for automatic cloud classification based on the synergetic use of the information reported by a sky camera and a ceilometer. So far, ceilometer has been not used in automatic cloud classification, so here the added value of this instrument is evaluated. Following recent bibliography, different features were derived from the sky camera images. These features, along with the information reported by the ceilometer, were used as input for a state-of-the-art machine learning classification system: random forests (RFs) (Breiman, 2001). This is a machine learning technique that has seldom been used for cloud classification but which is known to be among the best performers in classification tasks, according to some empirical studies (Caruana et al., 2008; Caruana & Niculescu-Mizil, 2006). In a recent work (Cheng & Lin, 2016), RFs have been used together with other algorithms (such as SVM and a Bayesian classifier) to develop a voting scheme for classifying each pixel in the image as cloud or noncloud. This is a related but different issue as the one addressed in the present paper, where whole images are classified, rather than individual pixels. RF belongs to the ensemble of decision trees family of algorithms. Ensemble techniques build models by training not only one but many different submodels whose outputs are combined. In RF, randomization techniques are used to build a varied set of submodels (decision trees). Classification is done by majority voting. RF deals with multiclass problems with no further adaptation. Also, RF training algorithms can easily take advantage of parallel computing.

The methodology is evaluated on a data set recorded from a camera and a ceilometer located at the radiometric station of the University of Jaén (Spain) over a set of days corresponding to the period 2013–2015. The procedure here proposed aims to mimic a fully operational one. As a consequence, skies with multiple cloud types and layers at the same time are considered and accounted for. In a recent work Wacker et al. (2015) reported this kind of skies to be highly challenging in automatic cloud classification. Three analyses were conducted: the first one to evaluate the role of the ceilometer and the camera information, the second to analyze the performance of the method when skies with several clouds types are included and, finally, to evaluate the performance of the model when using an increased number of cloud types. Evaluation was conducted on the light of the different cloud characteristics and the nature of the camera and ceilometer information.

2. Data Description

In this section, issues concerning to data used in this work are explained. Particularly, the camera and ceilometer hardware characteristics, data and preprocessing procedures, and the different types of clouds used in the classification are described.

2.1. Camera and Ceilometer Hardware Description

All the measurements used in this study were collected at the meteorological station of the University of Jaén, Andalucía (southern Spain), at coordinates 37.7877°N and 3.7782°W, and 454 m above mean sea level (Figure 1).

A total sky imager model Yesdas TSI-880 and a Jenoptik CHM 15k Nimbus ceilometer were installed in September 2012 (Figure 2). The TSI-880 is composed by solid-state CCD pointed downward at a hemispheric mirror, which reflects the whole hemisphere (fish-eye vision). Reflection of the Sun is blocked by a dark strip
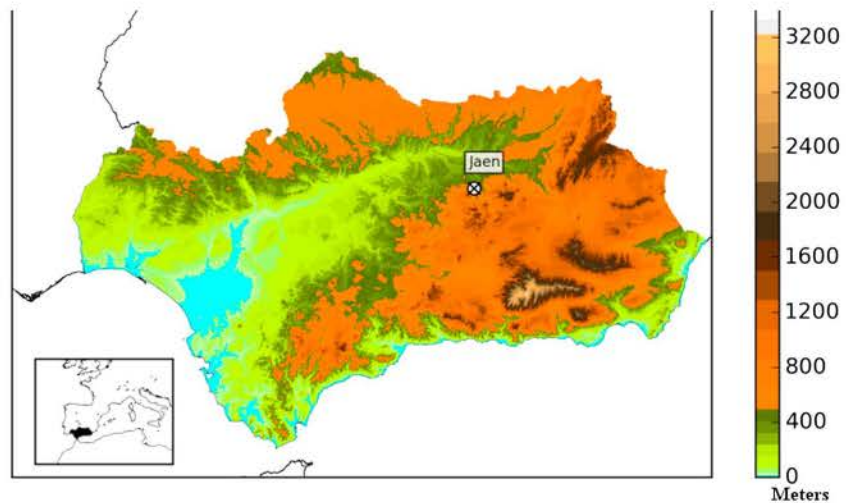
Figure 1. Study region and location of the meteorological station at the University of Jaén.

(shadow band), thereby protecting the imager optics. The TSI provides 352×288 pixels images every 30 s and has been designed for climate/weather applications, showing to be robust regarding environmental conditions (Long & DeLuisi, 1998). Notably, this camera has been proven to be accurate for the estimation of the cloud cover (Boers et al., 2010; Kreuter et al., 2009; Long et al., 2006; Mannstein et al., 2010). In the case of high clouds, it is able to report the sky conditions over a spatial domain of about $38 \text{ km} \times 38 \text{ km}$ (Mannstein et al., 2010). In the last years, this sky camera has been used as reference instrument in solar energy applications (Chow et al., 2011; Martínez-Chico et al., 2011; Quesada-Ruiz et al., 2014).

The Jenoptik CHM 15k nimbus ceilometer uses laser pulses at wavelength of 1,064 nm, receiving the back-scattered signal over a field of view of 0.45 mrad. This instrument is able to detect up to five cloud layers simultaneously and to provide their altitude with an accuracy of ± 5 m, being its vertical cloud detection range from 5 m to 15 km. The sample rate is 15 s. This particular ceilometer is one of the very few ones able to detect

Figure 2. Meteorological station of the University of Jaen. (left) Ceilometer and Sun tracker in the background. (right) The TSI-880 sky camera.
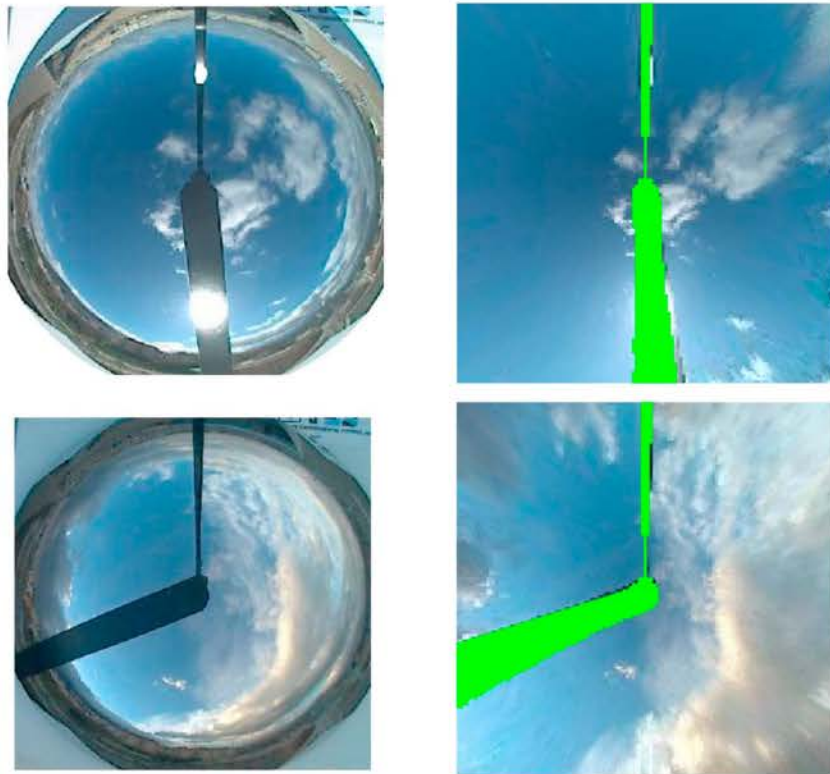
Figure 3. Two examples of raw and processed TSI images. (top row) Raw/processed image corresponding to day 2015-01-22 at 12:27:33 UTC. This image was classified as cumulus cloud according to Table 1. (Bottom row) Raw/processed image corresponding to the day 2015-02-14 at 17:22:18 UTC. This image was classified containing several cloud layers, that is, multicloud type according to Table 1.

clouds above 7.5 km and with less spurious values and better resolution in the upper cloud boundary than other similar instruments (Boers et al., 2010; Martucci et al., 2010; Wiegner et al., 2014).

2.2. Sky Camera Images and Ceilometer Data Preprocessing

A total of 717 TSI images, and the corresponding ceilometer estimates, was processed for this study. The images, corresponding to a total of 131 days of the years 2013 to 2015, were selected in order to have a representative sample, with different solar zenithal angles, of the 11 categories described in the following section. Every sample was meant to be representative of 5 min intervals, that is, images of each of the 11 categories were carefully selected to ensure that during the five previous minutes period exactly the same category was presented. First, the TSI images were masked in order to highlight the border, buildings, and band in the images. Second, the images were projected following Marquez and Coimbra (2013). This procedure transforms the images from a spherical to a rectangular grid. In order to prevent horizon distortion effects, this transformation was conducted only for zenithal angles below 65°, that is, a 130° field of view of the camera. Figure 3 shows some examples of the TSI raw and processed images.

The ceilometer reports every 15 s cloud profiles representative of the column at the ceilometer location, namely, cloud base height (CBH) and cloud penetration depth (CPD). In this work, up to three different cloud layers were considered. The CPD can be regarded as a proxy of the cloud geometrical thickness. Due to the nature of the clouds (high variability in space and time), ceilometer data should be properly processed in order to provide meaningful information linked to the TSI images, at the 5 min time interval used here. This is particularly relevant for some cloud types, such as cumulus, stratocumulus and cirrocumulus and, in general, cumuliform clouds. These clouds form patches, and therefore, the ceilometer may not report cloud information in some of the 20 samples of the 5 min evaluation period used here. In addition, the ceilometer sometimes provides spurious measures, or out of range values, which are related to the nature of the
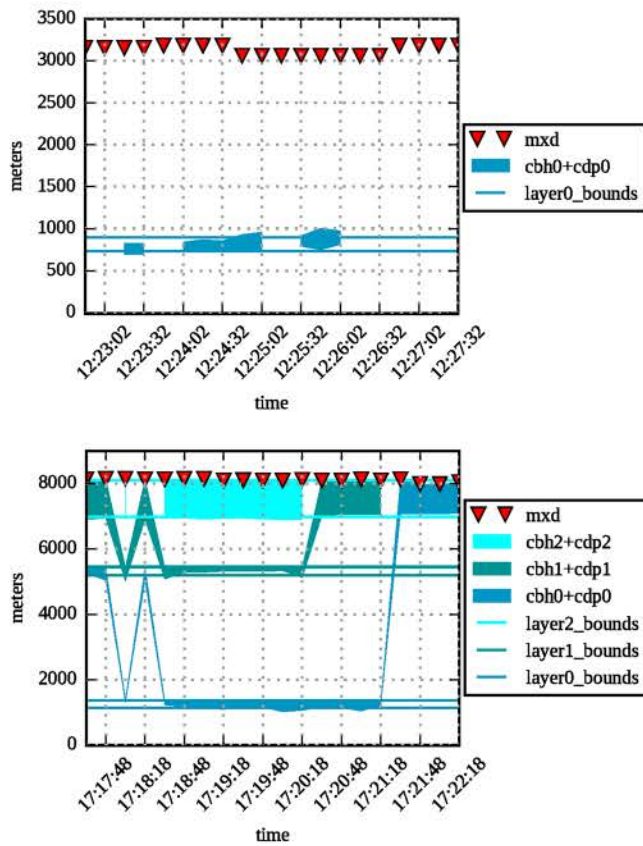
Figure 4. (top) Raw and processed ceilometer data corresponding to the image in Figure 3, top. The blue shaded area represents the range between the measured cloud base height and this values plus the cloud penetration depth. Values correspond to the five previous minutes at which the image was obtained. The bottom straight line shows the final estimate of the cloud base height for this sample, while the difference between the top and bottom straight lines shows the final estimate of the cloud penetration depth. Note that only some measurements were available during the 5 min interval. Triangular points at the top indicate the maximum detection range of the ceilometer for this particular measurement interval. (bottom) As in Figure 4, top but for the bottom image in Figure 3. Similarly to the previous case, the shaded areas indicate the measurements and the straight lines, the final cloud base height and cloud penetration depth estimates. Note that, in this case, three cloud layers were detected.

backscattering signal processed by these instruments. Given these issues, and since the methodology used in this work aims at emulating a fully operational system, ceilometer data were processed to provide meaningful cloud profile information. First, based on the 20 collected ceilometer samples, a number of candidate group of measurements are selected according to the active layers (up to three). Second, clear-sky values were removed from the 20 samples. Then, based on the CBH values of the remaining samples, a cluster analysis was carried out. The number of centroids in this cluster analysis provided the number of cloud layers, up to a maximum of 3. Finally, for each centroid, a mean CBH and CPD were computed, after applying a filter for outliers. If the 20 measurement are reported as clear sky (i.e., no clouds are detected), the ceilometer procedure final output is the presence of "0" cloud layers. Figure 4 shows an example of outputs of this processing procedure.

Since this process has been evaluated trying to mimic an operational system, some problems have been found. Particularly, in about 3% of the samples (24 images), the ceilometer reported no cloud information in cases for which the TSI-880 image was classified in some cloud category different from clear sky. A further analysis confirmed that 15 of these cases corresponded to cirrocumulus and cumulus. These clouds, in many cases, do not cover the whole sky dome and may not overpass the ceilometer column with the 5 min window here used. The other nine cases correspond to cloud types such as cirrus and nimbostratus. In these cases, the ceilometer was not able to provide the proper cloud information due to technical issues, reporting a very low detection range.

2.3. Sky Conditions and Cloud Classes

The sky images and the ceilometer information were used to manually classify the 717 samples according to classes displayed in Table 1. Particularly, two types of classification experiments were conducted. In the first ones, up to seven cloud types were used (first column in Table 1). These cloud categories are most commonly used in the bibliography (Heinle et al., 2010; Kazantzidis et al., 2012) and try to group cloud types with similar characteristics. In the second one, compound categories are decomposed into the individual cloud types, resulting in 10 cloud types (second column in Table 1). In both cases we have added the multicloud category, which aims to represent cases in which the sky is covered by several cloud types at the same time, including the case of several cloud layers. This category is commonly found and should be considered in fully operational systems. The multicloud category has been scarcely addressed in the literature. Wacker et al. (2015) describe the problems for automatic recognition of this category, but no attempt for classification was made. Only in Li et al. (2016) the multicloud case is considered in an automatic cloud classification procedure. This multicloud category is described as a mix of the sky conditions considered in this work and covering more than 20% of the sky.

3. Methods and Evaluation

In this section, automatic cloud classification is addressed. First, the features extracted from the images and ceilometer information, to be used as inputs to machine learning algorithm, are described. Next, a short description of random forest algorithm is included. Finally, the metrics and procedure used in this work to evaluate the performance of the classifier are also presented.

Table 1
Categories Used for the Cloud Classification and Main Characteristics Derived From the Ceilometer

Cloud types		Number of images	CBH Mean (SD)	CPD Mean (SD)
Seven cloud types + multicloud	10 cloud types + multicloud			
Clear sky	Clear sky (CLS)	48		
Cirrus and cirrostratus	Cirrus (ci)	131	9,086 (1,515)	951 (501)
	Cirrostratus (cs)	39	7,684 (676)	1,829 (422)
Cirrocumulus and altocumulus	Cirrocumulus (cc)	13	6,832 (2,023)	469 (238)
	Altocumulus (ac)	75	4,494 (2,257)	726 (516)
Altostratus and stratus	Altostratus (as)	57	6,701 (1,751)	1,858 (607)
	Stratus (st)	53	833 (485)	295 (276)
Stratocumulus	Stratocumulus (sc)	49	1,358 (372)	275 (126)
Cumulus	Cumulus (cu)	54	1,121 (513)	176 (32)
Nimbostratus	Nimbostratus (ns)	42	702 (345)	448 (424)
Multicloud	Multicloud (MC)	156		

Note. In the first experiments, a total of eight classes (first column) was distinguished. In the second one, classes increased to 11 (second column). In both cases, the category multicloud is included, indicating the presence of several layers and/or different cloud types in the same image. The mean and standard deviation (in parenthesis at the right) of the CBH and CPD (in meters) are displayed in the last two columns. Values correspond to the whole experimental database.

3.1. Features for Cloud Automatic Classification

In this work, we employ a wide set of features (see Table 2) as inputs to the cloud classifier. They are divided into two main groups, depending on which instrument was used to compute them: image features (extracted from the ground camera, features 1 to 12) and cloud layer features (extracted from the ceilometer, features 13 to 19).

3.1.1. Features From the Camera

Most of the image features used in this work are based on Heinle et al. (2010), and they have been obtained from the red, green, and blue channels of images. These channels are represented using three matrices M^r, M^g, and M^b, red, green, and blue respectively. Each (i, j) location in the matrices corresponds to a pixel in the image, with integer values between 0 and 255. There are several types of image features: spectral features, textural features, and cloud coverage.

The spectral features (rows from 1 to 7 in Table 2) use the color matrix M^c exclusively (where $c = r, g$, or b), extracting statistical measures directly from it. These are the simplest from the feature set and require very little processing.

The textural features (rows from 8 to 11 in Table 2) make use of a gray level co-occurrence matrix (GLCM). This is a transformation over one of the color channels. The result is a $g \times g$ matrix, g being the number of gray levels considered in the image. Thus, every element of the GLCMs in rows from 8 to 11 in Table 2 ($p_{i,j}^c$) represents the relative frequency of two adjacent pixel values i and j. c represents the color of the source channel. Here we use $g = 256$ *value levels*. GLCMs represent the relative frequency of two pixel values appearing together in the image, at a given offset (in this case, $x' = x + 1$, $y' = y + 0$). This matrix is commonly used in image analysis for detecting textures in gray images or in a given color channel and are supposed to give information on the spatial distribution of color, which spectral features are unable to provide. Textures are relevant in the detection of cloud types. There are several other textural features, as proposed by Haralick and Shanmugam (1973). However, the four used in the article are the subset of features proposed by Heinle et al. (2010). These measure different properties of the GLCM and are the following: energy (it measures the homogeneity of gray level differences), entropy (it measures the randomness of gray level differences), contrast (it measures local variation within the gray level matrix), and homogeneity (it measures similarity of adjacent gray levels within the matrix).

Finally, a cloud coverage statistic (row 12 in Table 2) is used in the procedure. To obtain this cloud coverage, first, the original red-green-blue image was converted to hue-saturation-value (HSV) color space following Smith (1978) and Jayadevan et al. (2015). Hue describes the color itself, while saturation denotes the degree of difference between a color and gray and value represents the brightness. Saturation ("Sat" in Table 2) fits

Table 2
Table Listing All of the Features Used as Input to the Classifier

|	Feature	Type	Formula
1	μ^r; Red average	Image-spectral	$\mu^{c_1} = \frac{1}{n^2} \sum_{j=0}^{n} \sum_{i=0}^{n} M_{i,j}^{c_1}$
2	μ^b; Blue average	Image-spectral	Same as above
3	σ^b; Blue deviation	Image-spectral	$\sigma^{c_3} = \sqrt{\frac{1}{n^2} \sum_{j=0}^{n} \sum_{i=0}^{n} \left(M_{i,j}^{c_3} - \mu^{c_3}\right)^2}$
4	γ^b; Blue skewness	Image-spectral	$\gamma^{c_3} = \frac{1}{n^2} \sum_{j=0}^{n} \sum_{i=0}^{n} \left(\frac{M_{i,j}^{c_3} - \mu^{c_3}}{\sigma^c}\right)^3$
5	D^{rg}; Red-green mean difference	Image-spectral	$D^{c_2} = \mu^{c_1} - \mu^{c_2}$
6	D^{rb}; Red-blue difference	Image-spectral	Same as above
7	D^{gb}; Green-blue mean difference	Image-spectral	Same as above
8	EN^b; Blue	Image-textural	$EN^{c_3} = \sum_{j=0}^{g} \sum_{i=0}^{g} \left[p_{i,j}^{c_3}\right]^2$
9	ENT^b; Blue	Image-textural	$ENT^{c_3} = \sum_{j=0}^{g} \sum_{i=0}^{g} p_{i,j}^{c_3} \log_2 p_{i,j}^{c_3}$
10	CON^b; Blue	Image-textural	$CON^{c_3} = \sum_{j=0}^{g} \sum_{i=0}^{g} (i-j)^2 p_{i,j}^{c_3}$
11	HOM^b; Blue	Image-textural	$HOM^{c_3} = \sum_{j=0}^{g} \sum_{i=0}^{g} \frac{p_{i,j}^{c_3}}{1 + |i-j|}$
12	C; % cloud coverage	Image-coverage	$Sat_{i,j} > T; T = 0.41$ (1) $C = \frac{cp}{tp}$ (2)
13	h_μ^1; Mean height from layer 1	Ceilometer-height	From CBH layer 1
14	h_μ^2; Mean height from layer 2	Ceilometer-height	From CBH layer 2
15	h_μ^3; Mean height from layer 3	Ceilometer-height	From CBH layer 3
16	t_μ^1; Mean thickness from layer 1	Ceilometer-thickness	From CPD layer 1
17	t_μ^2; Mean thickness from layer 2	Ceilometer-thickness	From CPD layer 2
18	t_μ^3; Mean thickness from layer 3	Ceilometer-thickness	From CPD layer 3
19	l; Present layers	Ceilometer-layers	Number of detected layers

Note. The last column shows how the feature is obtained according to the description in section 3.1. c_k indicates the color channel, using $k = 1$, 2, and 3 for red, green, and blue, respectively; n is the size of the $n \times n$ image, and g indicates the gray value of the pixel ($g = 256$ levels).

into the range [0, 1], from white, through the grays, to the most colorful hue. In this work, cloudy pixels are detected based on a threshold value $T = 0.41$ for the saturation value. Pixels (i, j) with a saturation greater or equal than this threshold are detected as clear sky; otherwise, the pixels are classified as cloudy. The percentage of sky covered is calculated using the formula labeled as (2) in Table 2 (row 12), where cp and tp are the amount of cloudy pixels and the total amount of pixels, respectively.

3.1.2. Features From the Ceilometer

The ceilometer offers height and thickness information about the cloud type (CBH and CPD) that can help discern differences between similar-looking clouds, which would be impossible to recognize otherwise.

Layers in cloud formations are numbered in order of distance from the ground. Layer 1 is the closest to the ground, then layers 2 and 3. Given this, we define six new features (CBH and CPD of each layer) plus an extra feature indicating how many actual layers (out of three) have been detected (rows from 13 to 19 in Table 2). We represent the information for each layer as h_μ^n or t_μ^n to indicate the mean CBH or CPD of layer n ($n = 1,2,3$),
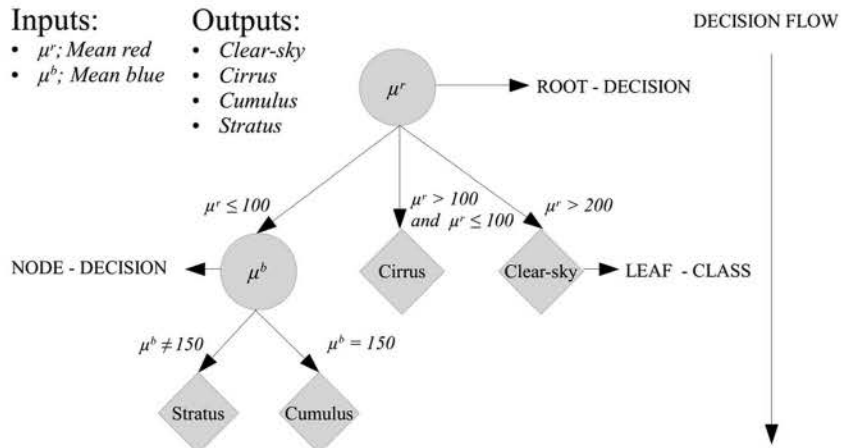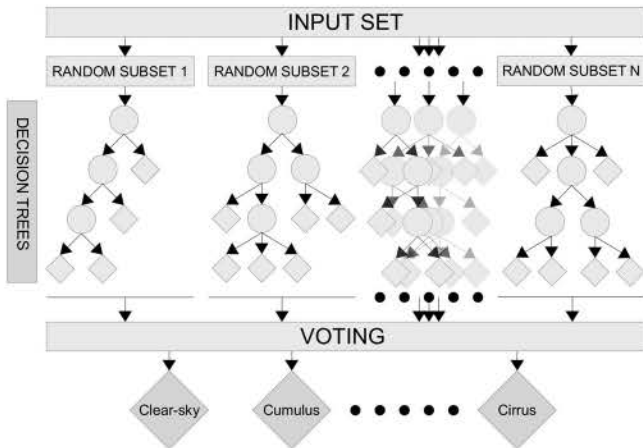
Figure 5. Example of decision tree with two input features and four possible classes. It has a maximum depth of 2, 2 decision nodes, and 4 leaf nodes.

and l the number of layers detected. In sum, a total of seven features was derived from the ceilometer to be used in the automatic classification procedure.

Machine learning algorithms require a fixed number of inputs/features. Therefore, in case the ceilometer returns just information of one or two layers, we fill the missing layers (up to 3) by replicating the information from the closest layer we have information of. In case there are no layers, values are set to an arbitrarily large number, indicating that clouds could not be detected.

3.2. Random Forests

For cloud classification, we use the random forest (RF) algorithm presented in Breiman (2001). RF has been reported to be one of the best algorithms for classification (Caruana & Niculescu-Mizil, 2006) and needs no adaptation to work in a multiclass context. This algorithm calculates N submodels (single classification trees) to form an ensemble of models that can predict the class of the given input. Every submodel is an individual decision tree. A simple example tree is shown in Figure 5. To classify an instance, the tree is navigated from the root node to a leaf node. Every nonleaf node contains a decision based on an input feature, which will determine the next node to be visited. The tree continues to be navigated through the nodes taking the path that decision nodes determine. Leaf nodes contain labels and, if a leaf node is reached, then the class is determined as the label of the given node.

The RF algorithm constructs multiple different trees from the same training data by means of a double randomization process. First, in order to build each tree, a new data set with the same size as the training data is obtained by sampling with replacement. Second, instead of considering the whole set of features, each decision node of each tree uses only a random subset of them (*mtry* is the parameter name for the size of this random subset, typically much smaller than the whole set of features). The set of decision trees in the RF ensemble classify new data by majority voting. A diagram of the whole process is represented in Figure 6.

Before building the final model, the parameter *mtry* has to be tuned for optimal classification accuracy. This parameter must be within the range $(1, (F-1))$, where F is the total number of features. The optimal F value is obtained by training and testing models with different values and selecting the best performing one. It is important to remark the tuning process uses the training partition only (the test partition is never used for training, parameter tuning being part of that training process). In this article, the RF implementation for R

Figure 6. Image depicting the construction of a random forest ensemble by random resampling and training of several decision trees. Classification is carried out by majority voting among ensemble members.

has been used (Liaw & Wiener, 2002). RF has been used together with package *caret*, which is able to deal automatically with parameter tuning (Kuhn, 2008).

3.3. Evaluation Procedure and Metrics

In order to evaluate the performance of the RF classifier for automatic cloud classification, a cross-validation procedure is carried out. Standard cross validation divides the available data in P equally sized folds or subsets. Then, for every fold n, a model is trained using all folds but n, and tested with fold n (i.e., a performance measure, such as accuracy, is computed for the trained model on fold n). The final cross-validation estimate is the average of the 10 accuracy values. The standard deviation can also be computed. In this work, we follow the common cross-validation practice and set $P = 10$.

However, applying standard cross validation to cloud image data sets can be potentially problematic if the data set contains sequences of images (cloud images, in this work) taken within short time periods, because some of the images in the sequence might be very similar. This phenomenon is called twinning, and it can lead to optimistically biased cross-validation estimates if very similar images fall into both the training and test partitions. To mitigate this problem, before splitting the data into folds, cloud images are sorted chronologically. Consequently, cloud images that are close in time will most likely fall together either into the training partition or the test partition. This evaluation process avoids the optimistic bias, and it will be more representative of a real situation, because it evaluates de classifier with data belonging to a time period different to that of the training data. However, this stricter validation should be expected to report worse metric values than other state-of-the-art works that use other evaluation methodologies.

The metrics used for measuring the effectiveness of the models are accuracy and macroaverage accuracy. Accuracy is the standard classification success rate:

$$\text{Accuracy}: \text{Acc}_{\text{Abs}} = \frac{I}{S},$$

where I is the number of correctly classified instances and S is the total number of instances.

The problem with (standard) accuracy is that classes with more instances have more weight in the success rate. For instance, in an extreme case, if class A contains 95 images and class B contains just 5 images, accuracy is basically informing about class A. In order to measure the behavior of the model independently of the number of images in each class, macroaverage accuracy can be used. Macroaverage accuracy is defined as the average of the individual class accuracies.

$$\text{Macroaverage accuracy}: \text{Acc}_{\text{Rel}} = \frac{1}{t} \sum_{k=0}^{t} \frac{I_k}{S_k},$$

where I_k is the number of well-classified instances for class k, S_k is the number of instances for class k, and t is the total number of classes.

All experiments carried out in this work follow the same flow. First, the data set is ordered chronologically and split into 10 different folds. Then, model evaluation is carried out with a tenfold cross validation. In every cross-validation iteration, the training folds are used to select the best *mtry* parameter value (see section 3.2) and then build the RF model with that value. Then the model is tested with the test fold. Given that RF is a stochastic algorithm, tenfold cross validation has been repeated 10 times, each time with a different random seed (in other words, 10 tenfold cross validations have been carried out). The results obtained are the average of these 10 different runs.

4. Results

In this section, results of the different experiments are presented and discussed. One of the aims of this work is to determine the relative contribution of the camera and ceilometer information for cloud classification. Therefore, baseline results were computed by training RF and testing the models using only image features from the camera (spectral, texture, and coverage features). Then, RF models were trained and tested with both camera and ceilometer information. To sum up, eight different experiments were conducted by (1)
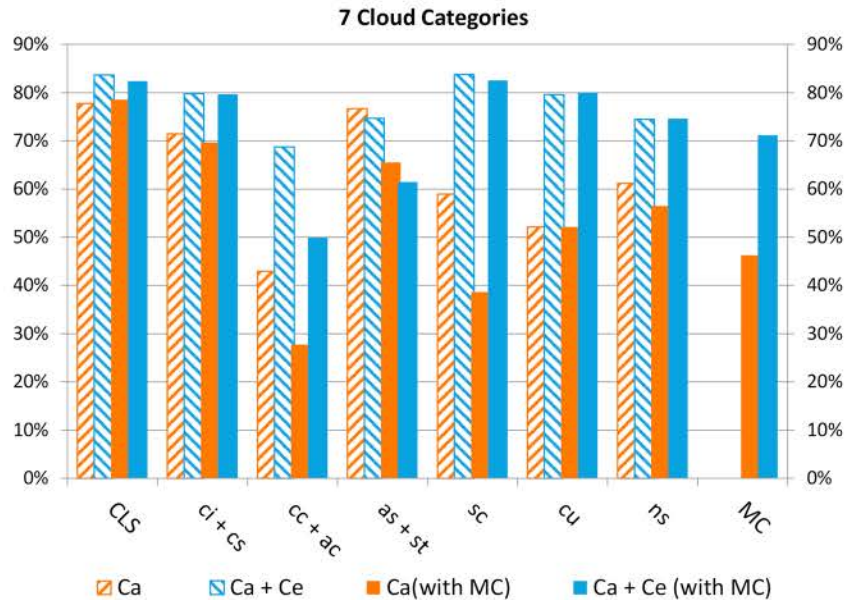
Figure 7. Relative frequencies (in percent) of correctly classified cloud classes for the seven cloud types (plus multicloud). Results are displayed separately for the four experiments: Using just the camera information (Ca) and both the camera and the ceilometer information (Ca + Ce) but not including the multicloud class, using just the camera information and including the multicloud class (Ca with MC) and using both the camera and the ceilometer information and including the multicloud class (Ca + Ce with MC).

using up to 7 or up to 10 classes, as described in section 2.3; (2) including/excluding the multicloud category; and (3) including/excluding the ceilometer information. Results of experiments have been organized in two blocks, with 7 and 10 classes (both with and without multicloud), respectively.

4.1. Seven Cloud Categories Plus Multicloud

Results of the classification procedure when considering seven cloud categories (Table 1), with or without multicloud, are given in Figure 7 and Table 3. Results are displayed for two cases: using just the camera information (Ca) and using the camera together with the ceilometer information (Ca + Ce). For the sake of comparison, results excluding and including the multicloud category are displayed separately.

Results clearly show, first, that the use of ceilometer information (Ca + Ce) improves the performance of the classifier for both cases, without multicloud (seven classes) and with multicloud (eight classes). In the former case, the use of the ceilometer improves accuracy and macroaverage accuracy by 12.91% and 15.72%, respectively (Table 3). The improvement is even larger for the multicloud case (15.97% and 17.46%, respec-

tively). Second, and as expected, including the multicloud class, results in a loss of approximately 5% accuracy when using all the features Ca + Ce, and about 8% when using only the camera (Ca). Interestingly, the ceilometer information allows the classifier to deal better with the extra (and noisy) multicloud class, compared to using only the camera information.

For the nonmulticloud experiments and breaking down results by cloud type, it can be observed (Figure 7) that the ceilometer information increases the accuracy for all cloud types except for stratus and altostratus (in this case, it gets slightly worse by 9.1%). The best improvements are observed for cirrocumulus-altocumulus (32.7%), cumulus (30.5%), and stratocumulus (31.6%). In the rest of the classes, accuracy is also improved to a lesser degree (around 8%). When the multicloud class is included, results are similar regarding the role of the ceilometer. Particularly, the use of ceilometer helps

Table 3
Overall Results for the Seven Class Experiments (Plus Multicloud)

		Seven classes	
		Features used	
	Metric	Ca	Ca + Ce
No Multicloud	Accuracy	64.4% (0.6)	77.3% (0.6)
	Macroaverage	62.3% (0.6)	78.0% (0.6)
Multicloud	Accuracy	55.7% (0.6)	71.7% (0.6)
	Macroaverage	55.1% (0.6)	72.6% (0.6)

Note. The accuracy, macroaverage accuracy (in percent), and standard deviation (within brackets) are displayed separately for experiments with camera only (Ca) and with camera and ceilometer (Ca + Ce). In addition, results are presented separately for experiments excluding (seven classes) and including (eight classes) the multicloud category.

Table 4
Contingency Matrix Results for the Seven Classes (Plus Multicloud) Experiment That Uses the Ceilometer Information

True class	Classified as								Mean success rate	
	CLS	ci + cs	cc + ac	as + st	sc	cu	ns	MC	Accuracy	Macroaverage
CLS	**82.2%**	2.3%	0%	0%	0%	9.1%	0%	0%		
ci + cs	13.9%	**79.0%**	12.5%	17.6%	0%	3.1%	0%	4.2%		
cc + ac	0.8%	2.7%	**46.2%**	1.1%	4%	4%	0%	10.6%		
as + st	0%	13.5%	3.5%	**64.2%**	2.2%	0%	8.4%	10.1%		
sc	0%	0%	8.9%	0.9%	**81%**	2.2%	6.5%	1.5%		
cu	2.6%	2.0%	1.1%	0.9%	2.2%	**81.7%**	2.3%	0.1%		
ns	0%	0%	0.1%	2.9%	2.8%	0%	**72.8%**	0.2%		
MC	0.4%	0.3%	27.8%	12.4%	7.8%	0%	10%	**73.4%**	71.7%	72.6%

Note. Rows contain the true class, and columns contain RF predictions. Bold entries represent the percentage of well-classified clouds for each cloud type.

to improve the accuracy of all classes except (again) for the stratus-altostratus class (in this case, accuracy is reduced by just 1.1%). This seems logical, as this class may contain clouds at very different altitudes. Similarly to the no multicloud case, observed improvements are for cirrocumulus-altocumulus (21.6%), for cumulus (27.6%), for stratocumulus (44%), and for multiple cloud type (24.9%). For clear-sky, cirrus-cirrostratus, and nimbostratus the improvement is smaller (around 5% and 8%). Finally, the accuracy of the multicloud class prediction is remarkable (around 73%) when using the ceilometer; otherwise, it is just about 48%.

The comparison of the results excluding and including the multicloud class reveals some interesting features. First, when using the ceilometer information, the inclusion of multicloud reduces the accuracy of the classification of just some specific cloud types, namely, cirrocumulus-altocumulus and stratus-altostratus. For the rest of the classes, scores are similar. This result makes sense, since the multicloud category somehow includes the cirrocumulus-altocumulus and the stratus-altostratus classes, which are also composed of several cloud types and cloud layers that can be located at very different altitudes. Therefore, the multicloud type may be confused by these two cloud types. This is what is observed in Table 4 (the classification contingency matrix). Even though the ceilometer helps enormously in the classification, multicloud is missclassified in about 10% of the cases as cirrocumulus-altocumulus and as stratus-altostratus. Cirrocumulus-altocumulus is classified as multicloud in 28% of the cases. Previous works have also shown that the class cirrocumulus-altocumulus is the most difficult to classify correctly (Kazantzidis et al., 2012; Wacker et al., 2015). The case of the cirrus-cirrostratus class is different, given that these clouds present a quite similar morphology and, more importantly, are usually located at a very similar elevation. As a consequence, multicloud is misclassified as cirrus-cirrostratus just about 4% of the cases.

Regarding the multicloud category, Wacker et al. (2015) reported that the inclusion of this kind of cloud class may reduce the classification rate up to a 50%. Li et al. (2016) reported this sky category to be the most difficult to classify, nevertheless obtaining an accuracy of 79:5%. This results is similar to the here presented when using the ceilometer (73%). Nevertheless, comparison is difficult given the different sky categories used in Li et al. (2016).

To sum up, the performance of the proposed procedure is highly dependent on the ceilometer information. This dependence is particularly relevant for all the "cumuliform" clouds, whose classification accuracy reduces considerably when only the camera information is used. On the other hand, the method showed to be robust against the inclusion of the multicloud class when ceilometer information is used (only the classification accuracy for the cirrocumulus-altocumulus and the stratus-altostratus is reduced).

4.2. Ten Cloud Categories Plus Multicloud

Table 5 shows the results when considering the 10 cloud types displayed in Table 1. First, it is observed that the accuracy scores

Table 5
As in Table 3 but for the 10 Classes (Plus Multicloud) Experiments

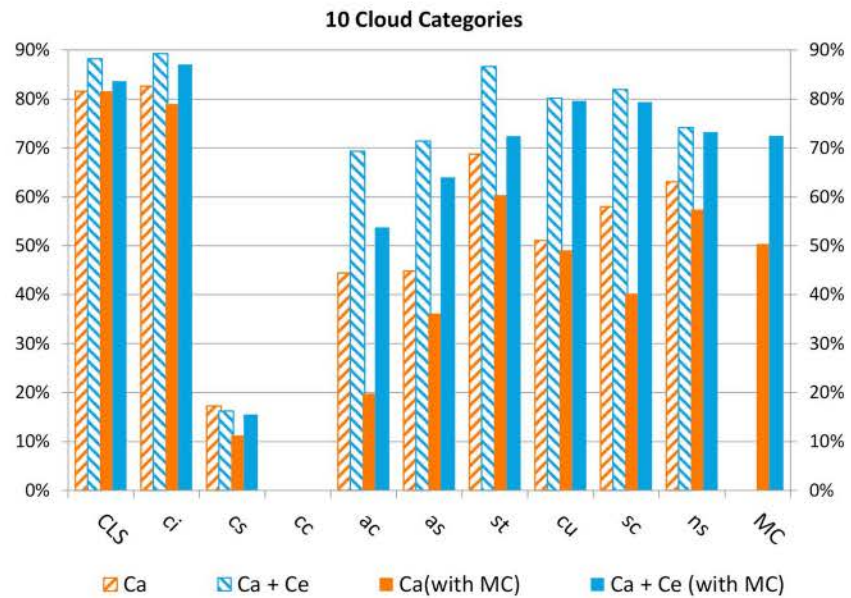|		10 classes	
		Features used	
	Metric	Ca	Ca + Ce
No Multicloud	Accuracy	58.8% (0.5)	74.8% (0.7)
	Macroaverage	51.4% (0.5)	66.4% (0.7)
Multicloud	Accuracy	50.6% (0.4)	71.1% (0.6)
	Macroaverage	44.8% (0.6)	63.5% (0.7)

Figure 8. As in Figure 7 but for the 10 cloud categories (plus multicloud).

decrease compared to the seven classes results described in section 4.1. This makes sense, given that the difficulty of classification problems tends to increase with the number of classes. Similarly to the seven class evaluation, a significant increment of the accuracy is obtained when using both camera and ceilometer (Ca + Ce). Overall, these increments are higher than in the seven class case (Table 3), indicating that ceilometer information is even more relevant when the number of classes is increased. Particularly, the accuracy and macroaverage accuracy increase by 20.5% and 18.8%, respectively, in the multicloud case.

When the multicloud class is included, accuracy and macroaverage are reduced by a 3%, approximately, if the ceilometer information is used (if only the camera is used, the reduction is larger). This result is similar to the seven class experiment. Therefore, the multicloud type does not seem to be an issue in this case. The reduction in the overall performance of the procedure seems to be related with the other categories.

Figure 8 and Table 6 break down results per class. Poor scores can be noticed for the cirrocumulus and cirrostratus classes, which show accuracies of near 0 and 20% respectively, regardless of the use of the ceilometer (Figure 8). Nevertheless, when using the ceilometer information, for some classes (clear-sky) the accuracy

Table 6
As in Table 4 but for the 10 Classes (Plus Multicloud) Experiments

True class	Classified as											Mean success rate	
	CLS	ci	cs	cc	ac	as	st	sc	cu	ns	MC	Accuracy	Macroaverage
CLS	**84.3%**	2.9%	0%	0%	0%	0%	0%	0%	9.6%	0%	0%		
ci	11.8%	**86.4%**	22.5%	35.7%	11.4%	9.5%	0%	0%	2.7%	0%	1.9%		
cs	0%	3.1%	**21.5%**	0%	0%	14%	0%	0%	0%	0%	2.2%		
cc	0%	0.6%	0%	**0%**	0.5%	0%	0%	0%	0%	0%	0.9%		
ac	0.6%	1.2%	0%	22.1%	**54.1%**	2%	0%	4%	2.2%	0%	9.0%		
as	0%	0.8%	50%	0%	2.9%	**64.8%**	0%	0%	0%	0%	5.3%		
st	0%	0%	0%	0%	0%	0%	**78.9%**	2%	0%	9.5%	4.5%		
sc	0%	0%	0%	0%	10.4%	0%	1.7%	**80.8%**	2%	7.7%	1.7%		
cu	3.3%	2.8%	0%	7.1%	0%	0%	1.8%	2%	**81.8%**	2.3%	0.2%		
ns	0%	0%	0%	0%	0%	0%	3.9%	3.8%	0%	**72.3%**	0.2%		
MC	0%	2.1%	6%	35%	20.6%	9.8%	13.7%	7.4%	1.6%	8.1%	**74.1%**	71.1%	63.5%

Note. Bold entries represent the percentage of well-classified clouds for each cloud type.

increases with respect to the seven class experiment or remain substantially the same (cumulus, stratocumulus, nimbostratus, and multicloud). Regarding the results for the formerly combined classes (cirrocumulus-altocumulus, cirrus-cirrostratus, and stratus-altostratus), now separated, some relevant outcomes were found. For instance, when using the ceilometer and including the multicloud class (Table 6), the stratus and altostratus classes show a high accuracy, 78.9% and 64.8%, respectively, higher than the combined stratus-altostratus class in Table 4 (62.2%). Note in Figure 8 the very relevant information provided by the ceilometer for these two classes. Both kinds of clouds present similar morphological features. The main difference is the location: while stratus are low-level clouds with the CBH below 2 km, the altostratus CBH are typically well above this elevation (Houze, 1993; Kokhanovsky, 2006). In our case, Table 1 data confirm these values, since the mean CBH of the stratus clouds is 833 m and the corresponding value for the altostratus is 6,701 m. Therefore, it seems that the combined information derived from the camera and especially the ceilometer is able to properly discriminate between these two classes of clouds, even when the multicloud class is included.

The separation of the class cirrus-cirrostratus is not so successful. The cirrus category is reliably classified, reaching 86.4% accuracy (Table 6). Note that the information provided by the ceilometer is not highly relevant in this case (Figure 8). Nevertheless, as commented above, cirrostratus results are poor (21.5%). They are classified as altostratus in 50% of the cases and as cirrus in 22.5% of the cases (Table 6). These results can be explained based on of the similar characteristics of the altostratus and cirrostratus clouds. Particularly, the altostratus clouds present a mean CBH of 6701 m, with 1,701 m standard deviation value (Table 1). The corresponding values of the cirrostratus clouds are 7684 m and 676 m. These experimental values are confirmed in the bibliography, which states that the range of elevation in middle latitudes is 2–7 km for the altostratus and 5–13 km for cirrostratus (Houze, 1993; Kokhanovsky, 2006). The CPD for both types of clouds is also similar: 1,858 m and 1,829 m for the altostratus and cirrostratus, respectively (Table 1). Therefore, these clouds cannot be discriminated just based on the CBH and the CPD. The main difference between these two kinds of clouds is the usual presence of the halo feature in the cirrostratus clouds but not in the altostratus. This particular feature seems to be not resolved by the image characteristics used here.

Finally, the poorest results are obtained for the cirrocumulus-altocumulus discrimination. Particularly, cirrocumulus clouds are systematically misclassified as altocumulus, cirrus, cumulus, or multicloud (Table 6). These poor results can be explained based on several reasons: first, because the mean and the standard deviation CBHs values (Table 1) of the cirrocumulus (6,833 m and 2,023 m), altocumulus (4,494 m and 2,257 m), and cirrus (9,086 m and 1,515 m) do not allow the use of the CBH to discriminate the cirrocumulus from the other two cloud types. Reference values of the CBHs in middle latitudes are 5–13 km for cirrocumulus and 7–10 km for cirrus 2–6 for altocumulus (Houze, 1993; Kokhanovsky, 2006), therefore confirming our results. Similar inferences can be derived for the role of the CPDs, which show mean and standard deviation values (Table 1) that makes the CPD inadequate to discriminate between these tree cloud types. Again, reference values in the bibliography confirm these findings; particularly, Houze (1993) and Kokhanovsky (2006) report the geometrical thickness of the cirrocumulus to be in the range 0.2–0.4 km, which overlaps the thickness of the altocumulus (0.2–0.7 km) and cirrus (0.1–3 km). Therefore, ceilometer information seems not to be relevant to distinguish the cirrocumulus clouds from many other classes. Regarding the sky camera information, from the morphological point of view, cirrocumulus and altocumulus are similar. In addition, cirrocumulus clouds often occur in small sheets located very high in the atmosphere (even 9 km values can be found in the experimental data set here used). As a consequence, and probably also because of the low resolution of the TSI images, the camera is not able to provide distinctive statistics values for this particular cloud class. Altocumulus results are more encouraging (accuracy 54%), although they are misclassified as multicloud in 20.6% of the cases.

5. Summary and Conclusions

We have presented and evaluated a methodology for automatic cloud-type classification based on the synergistic use of a sky camera and a ceilometer. The hypothesis is that given the distinctive vertical location of the different cloud types, the use of the ceilometer may improve the classification accuracy derived just from the camera images.

The methodology here evaluated aims to be fully operational, reporting an automatic classification of the cloud/sky conditions every 5 min. Because of that, among the evaluated cloud types, we have included the multicloud category, which accounts for skies covered by several cloud types and/or cloud layers. The automatic classification was conducted by random forests, a state-of-the-art machine learning classification algorithm, which used as input 19 features (12 computed from the sky camera and 7 from the ceilometer). The procedure was trained and evaluated on a set of 717 images, and up to 11 different types of clouds/skies were considered. The study is performed using a 130° field of view in the sky cameras.

A total of eight experiments were conducted by (1) excluding/including the ceilometer information in the random forest automatic classification algorithm, (2) including/excluding the multicloud category, and (3) using 7 or 10 different cloud/sky types, in addition to the multicloud category. The comparison of results allowed to evaluate the role of the ceilometer, to analyze the effect of the multicloud type in the classification accuracy and, finally, to evaluate the performance of the model when using an increased number of cloud/sky types (7 versus 10).

For the seven cloud/sky classes experiments plus multicloud (six cloud types + clear sky + multicloud) results showed an overall accuracy of the method of about 72% when using the ceilometer and about 55% when using just the camera information. Therefore, the ceilometer information showed to be crucial. The use of the ceilometer is particularly valuable for classifying cumuliform clouds, with an increment of the accuracy of about 30% compared to the use of only the camera information and, particularly, for the multicloud category, which is correctly estimated in about 73% (about 48% when just the camera is used). In addition, as may be expected, the inclusion of the multicloud class results in a loss of approximately 5% accuracy when using ceilometer and camera information and about 8% when using only the camera information. The 5% reduction was accounted for just some specific cloud types (cirrocumulus-altocumulus and stratus-altostratus). For the rest of the classes, scores were similar. To sum up, the ceilometer information allowed the classifier to deal better with the extra (and noisy) multicloud class, compared to using only the camera.

Results for the augmented 10 cloud/sky classes plus multicloud experiments (nine cloud types + clear sky + multicloud) showed lower accuracy scores. This makes sense, given that the difficulty of classification problems tends to increase with the number of classes. Particularly, for the experiment using the ceilometer information and including the multicloud class, mean macroaverage reduces from about 73% (7 cloud/sky classes plus multicloud) to about 63% (10 cloud/sky classes plus multicloud). The use of the ceilometer information resulted to be even more critical for the 11 categories than for the 8. From the analysis by categories of the experiments with 10 cloud/sky classes plus multicloud, some additional conclusions were obtained. The first one is that the reduction in the accuracy was not related with inclusion of the multicloud category but with other cloud types. Notably, the classification accuracy for cirrostratus, altocumulus, and, particularly, cirrocumulus showed to be low. Several reasons were found for this low accuracy. First, the fact that these clouds, from the morphological point of view, are similar to other cloud types: cirrostratus and cirrus, altocumulus and cirrocumulus/multicloud, and cirrocumulus and cirrus. As a consequence, the camera features were not able to distinguish between these kinds of clouds. Second, many of these clouds present similar cloud base height and geometrical thickness, making the ceilometer information not so relevant. This is the case of the cirrostratus and altostratus, and the altocumulus and cirrocumulus.

Other cloud types, as the stratus and altostratus, showed encouraging classification accuracy. Although these clouds present similar morphological characteristics, they are located at different elevation. As a consequence, ceilometer information allowed to reach better accuracy for these cloud types, even with the presence of the multicloud class.

Several applications may benefit from the here proposed automatic and operational cloud recognition system. For instance, in the field of solar energy, the here proposed method can be used to enhance the reliability of sky camera-based solar radiation estimating and forecasting procedures. Also for the characterization of the solar radiation spatial and temporal variability, that is an important issue for the solar energy grid integration. In addition, this methodology may reduce the uncertainty in the energy balance of the Earth surface, which is mainly related to the clouds. Finally, aviation weather services, which used the cloud type as a proxy of the present and for coming weather conditions, can benefit from the here proposed methodology.

Results here presented seem encouraging regarding the development of an automatic, fully operational, and highly tailored cloud classification procedure. Nevertheless, some limitations were found, and some

challenges should be addressed. First, ceilometer information was found to be highly valuable to accurately classify certain types of cloud (especially altocumulus, cumulus, stratocumulus, stratus, and altostratus) and makes the procedure robust against the inclusion of the multicloud. Camera information alone was found to be not suitable to deal with multicloud situations. Nevertheless, even the use of the ceilometer information showed some limitations. Problems are related with cloud types, which present similar morphological characteristics and, at the same time, similar elevation and geometrical thickness. In these cases, the only way to increase the classification accuracy is to develop specific features, either spectral or textural ones, able to account for the differences between cloud types. An example is the "halo" phenomena, which is present in cirrostratus but not in altostratus, and that the features used were not able to account for. Regarding this, the use of advanced sky cameras, with enhanced resolution and/or spectral responses, seems a promising tool. This will be explored in future works, as well as the role of the time window in classification accuracy.

In future works we aim to apply the here proposed methodology for different areas or patches in the image. This will allow, eventually, the classification of some of the here considered multicloud images in some specific cloud-type categories. In addition, alternative approaches to the use of the ceilometer information will be also explored, for instance, the use of stereographic methods to derived CBH (Kassianov et al., 2005; Peng et al., 2015) or the use of the cloud speed as proxy for the CBH (Peng et al., 2016; Quesada-Ruiz et al., 2014).

References

Arbizu-Barrena, C., Pozo-Vázquez, D., Ruiz-Arias, J. A., & Tovar-Pescador, J. (2015). Macroscopic cloud properties in the WRF NWP model: An assessment using sky camera and ceilometer data. *Journal of Geophysical Research: Atmospheres, 120*, 10,297–10,312. https://doi.org/10.1002/2015JD023502

Boers, R., de Haij, M. J., Wauben, W., Baltink, H. K., van Ulft, L. H., Savenije, M., & Long, C. N. (2010). Optimized fractional cloudiness determination from five ground-based remote sensing techniques. *Journal of Geophysical Research, 115*, D24116. https://doi.org/10.1029/2010JD014661

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Calbo, J., & Sabburg, J. (2008). Feature extraction from whole-sky ground-based images for cloud-type recognition. *Journal of Atmospheric and Oceanic Technology, 25*(1), 3–14.

Caruana, R., Karampatziakis, N., & Yessenalina, A. (2008). An empirical evaluation of supervised learning in high dimensions. In *Proceedings of the 25th international conference on Machine learning* (pp. 96–103).

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning* (pp. 161–168).

Cazorla, A., Olmo, F. J., & Alados-Arboledas, L. (2008). Development of a sky imager for cloud assessment. *Journal of the Optical Society of America A, 25*(1), 29–39.

Cheng, H., & Lin, C. (2016). Cloud detection in all-sky images via multi-scale neighborhood features and multiple supervised learning techniques. *Atmospheric Measurement Techniques Discussions, 10*(1), 199.

Cheng, H. Y., & Yu, C. C. (2015). Block-based cloud classification with statistical features and distribution of local texture features. *Atmospheric Measurement Techniques, 8*, 1175–1182.

Chow, C. W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., & Washom, B. (2011). Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Solar Energy, 85*(11), 2881–2893. https://doi.org/10.1016/j.solener.2011.08.025

Costa-Suros, M., Calbó, J., González, J. A., & Long, C. (2014). Comparing the cloud vertical structure derived from several methods based on radiosonde profiles and ground-based remote sensing measurements. *Atmospheric Measurement Techniques, 7*, 2757–2773.

Crammer, K., & Singer, Y. (2001). On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research, 2*, 265–292.

Haiden, T., Forbes, M., & Bozzo, A. (2015). The skill of the ECMWF cloudiness forecasts. *ECMWF Newsletter, 143*, 14–19.

Haralick, R. M., & Shanmugam, K. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, 3*(6), 610–621.

Heinle, A., Macke, A., & Srivastav, A. (2010). Automatic cloud classification of whole sky images. *Atmospheric Measurement Techniques, 3*, 557–567.

Houze, R. (1993). *Cloud dynamics*, Chap. 1. Cambridge, MA: Academic Press.

Illingworth, A. J., Hogan, R. J., O'Connor, J., Bouniol, D., Delanoë, J., Pelon, J., … Willén, U. (2007). Cloudnet—Continuous evaluation of cloud profiles in seven operational models using ground-based observations. *Bulletin of the American Meteorological Society, 88*, 883–898.

Jayadevan, V., Rodriguez, J., & Cronin, A. (2015). New contrast-enhancing feature for cloud detection in ground-based sky images. *Journal of Atmospheric and Oceanic Technology, 32*, 209–219. https://doi.org/10.1175/JTECH-D-14-00053.1

Kassianov, E., Long, C. N., & Christy, J. (2005). Cloud-base-height estimation from paired ground-based hemispherical observations. *Journal of Applied Meteorology, 44*(8), 1221–1233.

Kazantzidis, A., Tzoumanikas, P., Bais, A. F., Fotopoulos, S., Economou, G., & G. (2012). Cloud detection and classification with the use of whole-sky ground-based images. *Atmospheric Research, 113*, 80–88.

Kliangsuwan, T., & Heednacram, A. (2015). Feature extraction techniques for ground-based cloud type classification. *Expert Systems with Applications, 42*(21), 8294–8303. https://doi.org/10.1016/j.eswa.2015.05.016

Kokhanovsky, A. (2006). Cloud optics. *Atmospheric and ocenographic sciences library* (Vol. 34, Chap. 2, pp. 33–111). Berlin, Heidelberg: Springer.

Kreuter, A., Zangerl, M., Schwarzmann, M., & Blumthaler, M. (2009). All-sky imaging: A simple, system for atmospheric research. *Applied Optics, 48*, 1091–1097. https://doi.org/10.1364/AO.48.001091

Kriebel, K. T., Gesell, G., Kästner, M., & Mannstein, H. (2003). The cloud analysis tool APOLLO: Improvements and validation. *International Journal of Remote Sensing*, *24*, 2389–2408.

Kuhn, M. (2008). Caret package. *Journal of Statistical Software*, *28*(5).

Lee, J., Weger, R. C., Sengupta, S. K., & Welch, R. M. (1990). A neural network approach to cloud classification. *IEEE Transactions on Geoscience and Remote Sensing*, *28*(5), 846–855.

Li, Q., Zhang, Z., Lu, W., Yang, J., Ma, Y., & Yao, W. (2016). From pixels to patches: A cloud classification method based on a bag of micro-structures. *Atmospheric Measurement Techniques*, *9*, 753–764. https://doi.org/10.5194/amt-9-753-2016

Li, Y., Thompson, D. W., Stephens, G. L., & Bony, S. (2014). A global survey of the instantaneous linkages between cloud vertical structure and large-scale climate. *Journal of Geophysical Research: Atmospheres*, *119*, 3770–3792. https://doi.org/10.1002/2013JD020669

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, *2*(3), 18–22.

Long, C. N., & DeLuisi, J. J. (1998). Development of an automated hemispheric sky imager for cloud fraction retrievals. In *Proceedings of the 10th Symposium on Meteorological Observations and Instrumentation* (pp. 171–174). Phoenix, AZ: Am. Meteorol. Soc.

Long, C. N., Sabburg, J. M., Calbó, J., & Pagès, D. (2006). Retrieving cloud characteristics from ground-based daytime color all-sky images. *Journal of Atmospheric and Oceanic Technology*, *23*, 633–652. https://doi.org/10.1175/JTECH1875.1

Mannstein, H., Brömser, A., & Bugliaro, L. (2010). Ground-based observations for the validation of contrails and cirrus detection in satellite imagery. *Atmospheric Measurement Techniques*, *3*, 655–669. https://doi.org/10.5194/amt-3-655-2010

Marquez, R., & Coimbra, C. F. M. (2013). Intra-hour DNI forecasting based on cloud tracking image analysis. *Solar Energy*, *91*, 327–336. https://doi.org/10.1016/j.solener.2012.09.018

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. https://doi.org/10.1016/j.energy.2011.04.043

Martucci, G., Milroy, C., & O'Dowd, C. D. (2010). Detection of cloud-base height using Jenoptik CHM15K and Vaisala CL31 ceilometers. *Journal of Atmospheric and Oceanic Technology*, *27*, 305–318. https://doi.org/10.1175/2009JTECHA1326.1

Mateos, D., Antón, M., Valenzuela, A., Cazorla, A., Olmo, F. J., & Alados-Arboledas, L. (2014). Efficiency of clouds on shortwave radiation using experimental data. *Applied Energy*, *113*, 1216–1219.

Peng, Z., Yu, D., Huan, D., Heiser, J., & Kalbl, P. (2016). A hybrid approach to estimate the complex motions of clouds in sky images. *Solar Energy*, *138*, 10–25. https://doi.org/10.1016/j.solener.2016.09.002

Peng, Z., Yu, D., Huan, D., Heiser, J., Yoo, S., & Kalbl, P. (2015). 3D cloud detection and tracking system for solar forecast using multiple sky imagers. *Solar Energy*, *118*(2015), 496–519. https://doi.org/10.1016/j.solener.2015.05.037

Pincus, R., Hofmann, R. J. P., Anderson, J. L., Raeder, K., Collins, N., & Whitaker, J. S. (2011). Can fully accounting for clouds in data assimilation improve short-term forecasts by global models? *Monthly Weather Review*, *39*, 946–957. https://doi.org/10.1175/2010MWR3412.1

Quesada-Ruiz, S., Chu, Y., Tovar-Pescador, J., Pedro, H. T. C., & Coimbra, C. F. M. (2014). Cloud-tracking methodology for intra-hour DNI forecasting. *Solar Energy*, *102*, 267–275. https://doi.org/10.1016/j.solener.2014.01.030

Rumi, E., Kerr, D., Coupland, J. M., Sandford, A. P., & Brettle, M. J. (2013). Automated cloud classification using a ground based infrared camera and texture analysis techniques. In *SPIE Remote Sensing* (pp. 88900J–88900J). Germany: International Society for Optics and Photonics.

Schmidt, T., Kalisch, J., Lorenz, E., & Heinemann, D. (2015). Evaluating the spatio-temporal performance of sky imager based solar irradiance analysis and forecasts. *Atmospheric Chemistry and Physics Discussions*, *15*, 26,997–27,039.

Singh, M., & Glennen, M. (2005). Automated ground-based cloud recognition. *Pattern Analysis and Applications*, *8*(3), 258–271.

Smith, A. R. (1978). Color gamut transform pairs. *Computers and Graphics*, *12*, 12–19. https://doi.org/10.1145/965139.807361

Taravat, A., Del Frate, F., Cornaro, C., & Vergari, S. (2015). Neural networks and support vector machine algorithms for automatic cloud classification of whole-sky ground-based images. *IEEE Geoscience and Remote Sensing Letters*, *12*. https://doi.org/10.1109/LGRS.2014.2356616

Tzoumanikas, P., Nikitidou, E., Bais, A. F., & Kazantzidis, A. (2016). The effect of clouds on surface solar irradiance, based on data from an all-sky imaging system. *Renewable Energy*, *95*, 314–322. https://doi.org/10.1016/j.renene.2016.04.026

Viúdez-Mora, A., Costa-Surós, M., Calbó, J., & González, J. A. (2015). Modeling atmospheric longwave radiation at the surface during overcast skies: The role of cloud base height. *Journal of Geophysical Research: Atmospheres*, *120*, 199–214. https://doi.org/10.1002/2014JD022310

Wacker, S., Gröbner, J., Zysset, C., Diener, L., Tzoumanikas, P., Kazantzidis, A., … Kämpfer, N. (2015). Cloud observations in Switzerland using hemispherical sky cameras. *Journal of Geophysical Research: Atmospheres*, *120*, 695–707. https://doi.org/10.1002/2014JD022643

Weinberger, K. Q., & Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, *10*, 207–244.

Wess, S., Althoff, K. D., & Derwand, G. (1993). Using kd-trees to improve the retrieval step in case-based reasoning. In *European Workshop on Case-Based Reasoning* (pp. 167–181). Berlin Heidelberg: Springer.

Weya, E., & Schroedter-Homscheid, M. (2014). APOLLO cloud product statistics. *Energy Procedia*, *49*, 2414–2421. https://doi.org/10.1016/j.egypro.2014.03.256

Wiegner, M., Madonna, F., Binietoglou, I., Forkel, R., Gasteiger, J., Geiß, A., … Thomas, W. (2014). What is the benefit of ceilometers for aerosol remote sensing? An answer from EARLINET. *Atmospheric Measurement Techniques Discussions*, *7*, 2491–2543. https://doi.org/10.5194/amtd-7-2491-2014

Wild, M., Folini, D., Shär, C., Loeb, N., Dutton, E. G., & König-Langlo, G. (2013). The global energy balance from a surface perspective. *Climate Dynamics*, *40*(11–12), 3107–3134. https://doi.org/10.1007/s00382-012-1569-8

World Meteorological Organization (WMO) (2012). World Meteorological Organization/World Weather Research Programme (WMO/WWRP). Recommended methods for evaluating cloud and related parameters World Weather Research Programme (WWRP)/Working Group on Numerical Experimentation (WGNE) Joint Working Group on Forecast Verification Research (JWGFVR), Document WWRP 2012–1, March. Retrieved from http://www.wmo.int/pages/ prog/arep/wwrp/new/documents/WWRP_2012_1_web.pdf

World Meteorological Organization (WMO) (2017). International Cloud Atlas, Preface to the 2017 edition. Retrieved from https://cloudatlas.wmo.int/preface-to-the-2017-edition.html

Zhen, Z., Wang, F., Sun, Y., Mi, Z., Liu, C., Wang, B., & Lu, J. (2015). SVM based cloud classification model using total sky images for PV power forecasting. In *Innovative Smart Grid Technologies Conference (ISGT), 2015 IEEE Power & Energy Society* (pp. 1–5). New York: IEEE.