# Development of a Random-Forest Cloud-Regime Classification Model Based on Surface Radiation and Cloud Products

JOSEPH SEDLAR,[a,b] LAURA D. RIIHIMAKI,[a,b] KATHLEEN LANTZ,[b] AND DAVID D. TURNER[c]

[a] *Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado*
[b] *NOAA/Global Monitoring Laboratory, Boulder, Colorado*
[c] *NOAA/Global Systems Laboratory, Boulder, Colorado*

(Manuscript received 9 July 2020, in final form 13 January 2021)

ABSTRACT: Various methods have been developed to characterize cloud type, otherwise referred to as cloud regime. These include manual sky observations, combining radiative and cloud vertical properties observed from satellite, surface-based remote sensing, and digital processing of sky imagers. While each method has inherent advantages and disadvantages, none of these cloud-typing methods actually includes measurements of surface shortwave or longwave radiative fluxes. Here, a method that relies upon detailed, surface-based radiation and cloud measurements and derived data products to train a random-forest machine-learning cloud classification model is introduced. Measurements from five years of data from the ARM Southern Great Plains site were compiled to train and independently evaluate the model classification performance. A cloud-type accuracy of approximately 80% using the random-forest classifier reveals that the model is well suited to predict climatological cloud properties. Furthermore, an analysis of the cloud-type misclassifications is performed. While physical cloud types may be misreported, the shortwave radiative signatures are similar between misclassified cloud types. From this, we assert that the cloud-regime model has the capacity to successfully differentiate clouds with comparable cloud–radiative interactions. Therefore, we conclude that the model can provide useful cloud-property information for fundamental cloud studies, inform renewable energy studies, and be a tool for numerical model evaluation and parameterization improvement, among many other applications.

KEYWORDS: Cloud forcing; Cloud radiative effects; Clouds; Radiative fluxes; Radiative forcing

## 1. Introduction

Clouds are a crucial aspect of the climate system via their direct connection to the hydrological cycle and their influence on Earth's energy balance. Because clouds significantly interact with solar (shortwave) and infrared (longwave) fluxes, their influence on the surface and top-of-atmosphere radiation budgets drive weather and climate across a wide range of temporal and spatial scales (e.g., Peixoto and Oort 1992; Trenberth et al. 2009; Ahrens 2012). Cloud–radiative interactions depend upon a range of physical characteristics of the cloud layer(s), including the vertical distribution of the layers, and each layer's cloud phase (ice, liquid, mixed), cloud particle size, and integrated water content. These properties determine scattering and absorption characteristics of the cloud, which also depend on wavelength. Clouds of a similar regime, or type, typically have comparable microphysical and macrophysical properties. As such, classifying clouds into a limited set of regimes can be useful in understanding and characterizing their role in the weather and climate system.

The development of cloud classification regimes provides important observationally based information that serves as a metric for fundamental understanding of cloud processes and the evaluation and improvement of cloud representation in weather prediction and global circulation models (GCMs; e.g., Chen and Del Genio 2009; Jin et al. 2017; Van Weverberg et al. 2018). In general, three primary methods that have been developed to create cloud-regime datasets based on observations: 1) employing multivariable distributions of radiative and cloud-top properties retrieved from satellite (e.g., Rossow and Schiffer 1991, 1999); 2) digital processing of camera imagery viewing the sky to classify cloud texture (coverage) and structure from the surface (e.g., Long et al. 2006; Kazantzidis et al. 2012); and 3) combining of zenith-viewing active and passive remote sensing instrumentation capable of resolving cloud hydrometeors in the vertical (Kollias et al. 2007a; Tselioudis and Kollias 2007).

These methods have their inherent advantages and disadvantages. Cloud types based on joint histogram distributions of cloud optical thickness and cloud-top pressure from passive satellite observations can provide cloud-regime datasets across a wide temporal and spatial range. For example, the ISCCP (Rossow and Schiffer 1999) and Cloud, Albedo, and Surface Radiation dataset (CLARA), version A2 (CLARA-A2; Karlsson et al. 2017), joint distributions provide cloud-regime observations globally for 30+ years starting from the early 1980s. Often, these joint distributions are subjected to clustering techniques in order to categorize their dominant organization patterns, either regionally (e.g., Jakob and Tselioudis 2003) or globally (e.g., Pincus et al. 2012; Tselioudis et al. 2013; McDonald et al. 2016). The advantage of satellite-based cloud regimes is the wide spatial and temporal coverage, permitting study of the dominant cloud structures and how these have evolved in a changing climate. Clustering of global cloud

Denotes content that is immediately available upon publication as open access.

*Corresponding author*: Joseph Sedlar, joseph.sedlar@colorado.edu

DOI: 10.1175/JAMC-D-20-0153.1

regimes could also be analyzed jointly with the background weather state, indicating a linkage between synoptic weather forcing and cloud type (Tselioudis et al. 2013; McDonald and Parsons 2018). However, 2D radiative–cloud-top distributions are created by combining satellite footprints, causing a coarser spatial resolution than the actual pixel resolution; such a ''coarse graining'' process is necessary to get robust, meaningful cloud-regime statistics (Rossow and Schiffer 1991; Karlsson et al. 2017). Moreover, infrequent temporal sampling resulting from the polar orbit of the satellites presents another disadvantage of this method, especially when high-frequency changes in cloud type and radiative fluxes are an application of interest. Global coverage on long, 30+-yr data records makes this method useful for studying climate time scales and for evaluating GCMs. However, cloud regimes from satellite joint histograms potentially wash out the small-scale variability into the regional distribution/characteristics of the clouds (Leinonen et al. 2016) and these regimes may actually not be unique clouds but rather integrative features from multiple cloud types simultaneously (Mace and Wrenn 2013).

Directed camera images of the sky and from whole-sky-viewing hemispheric domes have been processed to develop automated estimates of total cloud cover and cloud structure (e.g., Martins et al. 2003; Long et al. 2006; Kazantzidis et al. 2012; Zhuo et al. 2014), and further use clustering algorithms to distinguish cloud types through digital image processing (e.g., Heinle et al. 2010; Wang et al. 2018). A primary advantage of using cloud typing of this method is the relatively high-frequency temporal sampling and the ability to capture the high-frequency variability common with cloud evolution and life cycle. In comparison to satellite cloud regimes, spatial information is limited to the viewing geometry of the camera or the hemispheric dome reflecting the sky conditions. Generally, the classification of clouds using digital image processing are reported with relatively high accuracy, although discrepancies are reported when a single, unique cloud class is not observed, such as during multilayer cloud scenes (Heinle et al. 2010). An additional disadvantage is the use of manually observed ''truth'' cloud types in order to train the classification clustering algorithms; these truth observations are often classified subjectively by human detection, potentially incorporating biased decisions (Heinle et al. 2010).

Active sensing of cloud layers from surface or spaceborne cloud radar and lidar provide the most robust vertically resolved description of cloud properties. Backscattered energy from cloud particles in the atmospheric column within the instrument field of view is used to identify cloud layers by distinguishing and partitioning individual cloud layers from precipitation signals (e.g., Moran et al. 1998; Kollias et al. 2007b). Physical geometric thresholds are applied to the cloud-base and cloud-top heights and layer thicknesses are then calculated, and the cloud layers are classified into regimes based on their physical characteristics (Kollias et al. 2007a; Tselioudis and Kollias 2007; Lim et al. 2019). The high temporal and vertical resolution of the active sensors are advantageous for capturing high-frequency cloud variability; however, these instruments are costly and require significant infrastructure and personnel to maintain operation. Their deployments are generally limited spatially to long-term observatories such as the Atmospheric Radiation Measurement (ARM) program (Mather and Voyles 2013; Turner and Ellingson 2016) or short-term field experiments.

Their zenith-viewing orientation limits the effective field of view to the sky conditions as they advect over the instruments. The high level of vertical detail in cloud properties derived from these instruments has been exploited to understand biases in numerical modeling of cloud fields. Tselioudis and Kollias (2007) used their cloud regimes to identify biases in the European Centre for Medium-Range Weather Forecasts (ECMWF) model forecasts, reporting a major gap in the model's ability to represent multilayered clouds over the southern United States. Ahlgrimm and Forbes (2012) applied the cloud regimes developed by Kollias et al. (2007a) and found critical errors in the ECMWF forecasts of shortwave irradiance were connected to biases in both shallow cumulus and overcast stratiform clouds. These vertically resolved cloud regimes have also been used to evaluate suites of model runs to identify which cloud types led to radiative and temperature biases over the southern Great Plains (Van Weverberg et al. 2018).

Characterizing different cloud regimes and exploring how surface radiative fluxes vary by these regimes provides an effective and critical tool to improve our understanding of cloud–radiative interactions and translate fundamental understanding into improvements in numerical modeling. In this regard, classifying cloud regimes based on actual surface radiation measurements seems a logical choice. While the three cloud-regime methods described above have demonstrated success in classifying cloud types, none of the three were developed based on actual surface radiation measurements. The sky cameras make use of visible imagery but not broadband radiative fluxes. Satellite histograms incorporate the retrievals of cloud optical thickness, but this cloud radiative property is valid for top of atmosphere radiance and does not equate to radiative flux measurements actually penetrating the cloud and reaching the surface.

This paper describes the development of a machine-learning cloud-regime classification model designed using surface broadband radiation measurements and cloud products. The observations used as independent input variables are based on measurements and derived data products originating from the SURFRAD surface radiation network across the United States (Augustine et al. 2005). However, the cloud-regime classification model can be applied to any observatory where similar broadband radiation and cloud measurements are made. In fact, observations identical to those made at SURFRAD are available from the ARM Southern Great Plains (SGP) Central Facility in Oklahoma; these observations at SGP, together with a fully independent cloud classification product, serve as the basis for the training and evaluation processes of the machine-learning classification model. Careful attention to the choice of input measurements has been made such that the classifier can also be run on readily available numerical model output.

In this study, the development of a novel cloud-regime method is presented using a machine-learning framework.

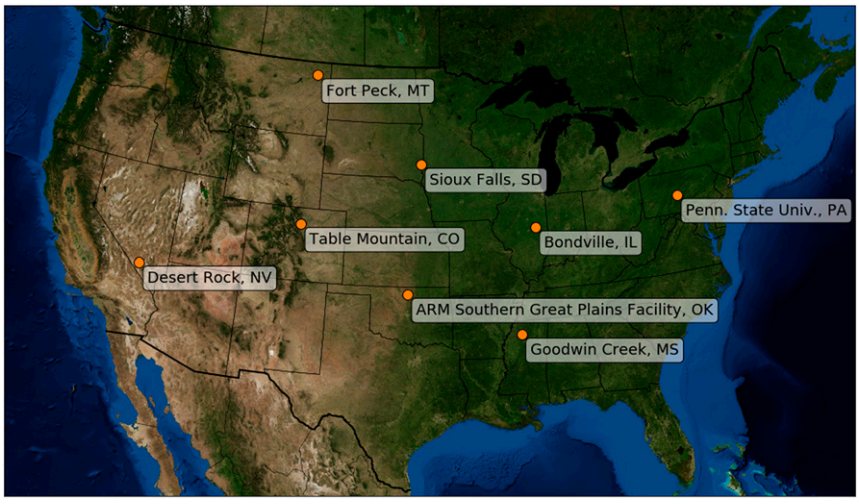


FIG. 1. Geographic distribution of the seven SURFRAD observatories. Included is the location of the ARM SGP Central Facility observatory.

Various machine-learning methods have a long history in meteor. McGovern et al. (2019) documents a number of machine-learning frameworks and their applications in the literature designed to improve weather forecasting, such as precipitation type forecasts and severe storm probability. Here, detailed yet generally disparate surface-based measurements are combined within a random-forest classifier framework to predict the overlying cloud regime. Following the input measurements, a description is given of the independent observational cloud-type data product needed for training the machine-learning model and how this cloud type was extended for our needs. Results from multiple evaluations of the cloud-regime classifier are presented, focusing on the prediction accuracy and the advantage of using surface radiation and cloud measurements as classifier input variables. The results are discussed with an emphasis on the advantages of our regime classifier method and its relatively high accuracy.

## 2. Methods and data

This study describes a machine-learning technique to classify cloud type based on measurements of surface radiation and cloud properties. The random-forest classification machine-learning algorithm requires training and optimization of input values (features) against independent truth (labels) (e.g., Breiman 2001). For our application, the input features are surface radiation, cloud-property observations, and enhanced cloud data products. The labels are the true cloud types observed that have influenced the surface radiation measurements. In the following sections, a description of the input measurements and the observed (truth) cloud regimes are presented, followed by a description of the random-forest classifier training process.

### *a. Input feature measurements*

Collocated measurements of surface radiation and cloud properties are generally abundant and have the ability to provide a wealth of information about cloud–radiative interactions. But these measurements are often disparate and as such make it difficult to determine how different cloud regimes influence the radiative fluxes at the surface. Combining these separate but interconnected measurements into data vectors for input into a machine-learning model is one way of providing a unified linkage between them. The major constraint is that measurements and subsequent data products must be consistent. The U.S. Department of Energy's (DOE) ARM User Facility sites satisfy the need for consistent, high-quality radiation and cloud measurements.

The ARM SGP observatory (see Fig. 1) in Oklahoma has been measuring high-quality surface radiation and producing radiation and cloud data products since 1992 (Sisterson et al. 2016). Strict quality assurance and quality control measures on datasets and data products have been the ethos of ARM since its inception. In addition to broadband downwelling and upwelling shortwave and longwave irradiances, ARM facilities measure direct normal and diffuse solar irradiance components using pyrheliometers and shaded radiometers mounted on automated solar trackers. Beyond the high-quality surface radiation measurements, clear-sky irradiances and cloud properties are estimated from broadband irradiance measurements through the Radiative Flux Analysis (RadFlux) processing and analysis algorithms (Long and Ackerman 2000; Long et al. 2006; Long and Turner 2008). Such products include clear-sky downwelling shortwave and longwave irradiances, cloud fractional sky coverage (hemispherical) based on shortwave and longwave radiation, respectively, and effective shortwave cloud transmissivity. Active remote sensors for measuring cloud vertical distributions and properties were installed at SGP, including zenith-viewing millimeter wavelength cloud radar (MMCR; Moran et al. 1998; Kollias et al. 2016) and lidar systems (Campbell et al. 2002).

Only a subset of all available radiation and cloud products are used as input features for our cloud classifier. Table 1 lists the specific measurements and/or data products used as input

TABLE 1. Random-forest input feature (observation or derived product) name, a description of the observation or derived data product, and the instrument(s) used to measure/derive the feature.

| Feature | Description (units) | Instrument(s) |
|---|---|---|
| swd direct | Direct downwelling shortwave irradiance ($W m^{-2}$), normalized by cosine of solar zenith angle (SZA) | Eppley Normal Incidence Pyrheliometer installed on a solar tracker |
| swd diffuse | Diffuse downwelling shortwave irradiance ($W m^{-2}$), normalized by cosine of SZA | Shaded Eppley Black and White 8–48 Pyranometer installed on a solar tracker |
| lwd | Downwelling longwave radiation ($W m^{-2}$) | Shaded Eppley Precision Infrared Pyrgeometer installed on a solar tracker |
| lwd clear | Clear-sky downwelling longwave radiation ($W m^{-2}$) | Derived quantity based on lwd measurements (produced by RadFlux) |
| diffuse/total | Ratio of downwelling shortwave diffuse irradiance to total (direct + diffuse) downwelling shortwave irradiance (unitless) | Shaded Eppley Black and White 8–48 Pyranometer and Eppley Normal Incidence Pyrheliometer |
| diffuse – diffuse clear | Downwelling shortwave diffuse irradiance minus clear-sky downwelling shortwave diffuse irradiance ($W m^{-2}$), normalized by cosine of SZA | Shaded Eppley Black and White 8–48 Pyranometer (for diffuse measurement); diffuse clear is a derived quantity that is based on all-sky diffuse measurements (produced by RadFlux) |
| cre sw | Shortwave cloud radiative effect, computed as total downwelling shortwave irradiance minus clear-sky total downwelling shortwave irradiance ($W m^{-2}$), normalized by cosine of SZA | Eppley Normal Incidence Pyrheliometer and Shaded Eppley Precision Spectral Pyranometer, combined with clear-sky downwelling shortwave flux (produced by RadFlux) |
| cre lw | Longwave cloud radiative effect; computed as downwelling longwave radiation minus clear-sky downwelling longwave radiation ($W m^{-2}$) | Shaded Eppley Precision Infrared Pyrgeometer; clear-sky downwelling longwave is a derived quantity based on lwd measurements (produced by RadFlux) |
| shortwave cfrac | Fractional sky covered by clouds; computed by combining all-sky downwelling shortwave radiation with clear-sky downwelling shortwave radiation (%) | Derived quantity produced by RadFlux; this feature combines measurements from Eppley Normal Incidence Pyrheliometer with shaded Eppley Black and White 8–48 Pyranometer and compares the shortwave components with their clear-sky values (from RadFlux) |
| 15-min std(swtran) | The 15-min running std dev of shortwave transmissivity (swtrans); swtrans is the ratio of downwelling shortwave radiation to clear-sky downwelling shortwave radiation (unitless) | Derived quantity produced by RadFlux; this feature combines measurements from Eppley Normal Incidence Pyrheliometer with shaded Eppley Black and White 8–48 Pyranometer and compares the shortwave components with their clear-sky values (from RadFlux) |
| num ceilo layers | No. of ceilometer-detected cloud layers, up to a max of three (integer) | Vaisala CL51 |
| cloud base height | Height above ground level to the ceilometer first-detected cloud-base height (m) | Vaisala CL51 |

features for our cloud-regime classification model. These features vary in complexity from straightforward broadband and solar component measurements, lowest cloud-base height and cloud presence from ceilometer profile measurements, to more advanced derived features such as shortwave/longwave cloud radiative effect (all-sky minus clear-sky downwelling radiation), fractional sky coverage from shortwave measurements, and 15-min temporal variability of shortwave transmissivity. These features have been chosen on the basis of relatively high data recovery and their ability to provide useful cloud-property information, which leads to distinguishing between different cloud regimes. The variables were specifically chosen because the same high-quality instruments, measurements, and data products are routinely observed by the NOAA Global Monitoring Laboratory's surface radiation budget SURFRAD (Augustine et al. 2005) network across the United States (see Fig. 1). SURFRAD observatories have been operational since 1995, measuring surface radiation and producing derived radiative and cloud products through the RadFlux processing and analysis algorithms. Beginning in 2018, Vaisala CL51 ceilometers (https://www.vaisala.com/sites/default/files/documents/CL51-Datasheet-B210861EN.pdf) were added to each SURFRAD observatory to provide active sensor measurements of clouds and aerosol properties across the atmospheric column; ceilometers were installed and operational at each observatory by late 2019. These zenith-viewing lidars measure attenuated backscatter across the troposphere, producing time–height retrievals of aerosol and cloud backscatter. From these retrievals, cloud fraction and cloud-base height are provided and can be used as input to the random-forest classifier (Table 1). Unlike ARM SGP, no additional active sensing instruments are deployed at SURFRAD. Because of this, SGP observations and an independently derived cloud-type data product developed at SGP (Lim et al. 2019) serve as the central

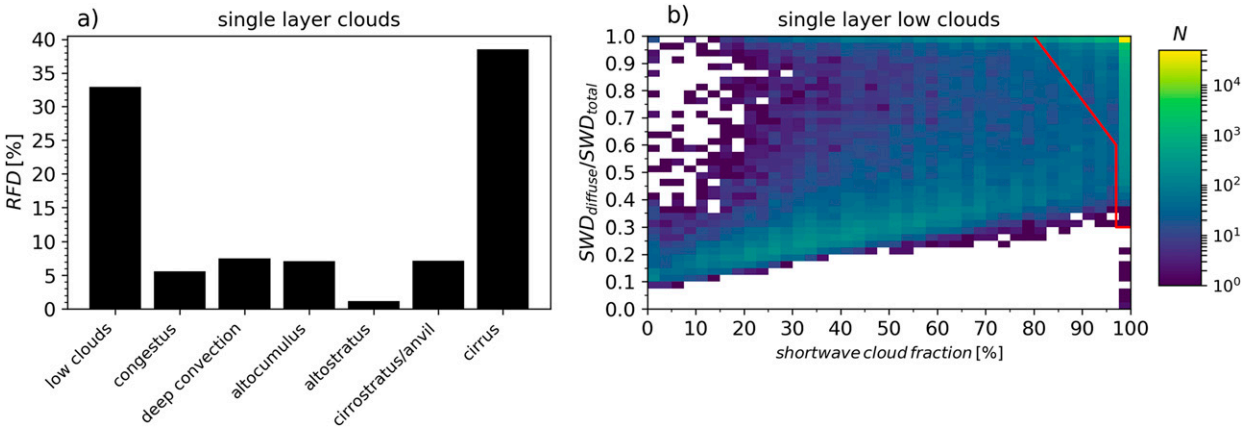


FIG. 2. (a) Relative density distribution of cloud types at ARM SGP for 2014–17 from Lim et al. (2019). (b) Frequency distribution (counts $N$; shading) of $SWD_{diffuse}/SWD_{total}$ ratio vs shortwave cloud fraction (%). The red line indicates the separability threshold between low shallow cumulus (to the left of the red line) and low stratiform (to the right of the red line) clouds in the two-parameter phase space. All observations are on a 1-min time scale.

measurements for the training and evaluation process of the random-forest classifier.

### b. *Independent cloud-type observations*

The cloud-type product produced by Lim et al. (2019) is based on zenith-viewing lidar and radar retrievals of cloud properties. The Active Remote Sensing of Cloud Layers (ARSCL) retrievals algorithm (e.g., Kollias et al. 2016) provides 2D time–height data products of cloud-base height, cloud-top height, and cloud thickness. Lim et al. (2019) developed thresholds based on the 2D cloud boundaries that type the cloud layer(s) into categories based on physical characteristics of seven different cloud types: 1) low clouds, 2) congestus, 3) deep convection, 4) altocumulus, 5) altostratus, 6) cirrostratus/anvil, and 7) cirrus. Criteria used to generate these cloud types, including a description on the thresholding of cloud boundaries and cloud-layer thickness, is found in Lim et al. (2019) and Flynn et al. (2017).

The cloud types are relatively comprehensive, covering the majority of cloud regimes present at any geographic location any time of the year. However, the ''low cloud'' category proposed by Lim et al. (2019) is too broad and categorizes low stratiform (stratus, stratocumulus) cloud types together with shallow cumulus. While a broad cloud type may be sufficient for specific applications, such as estimating ceiling height, stratiform and shallow cumulus clouds exert very different signatures in surface shortwave radiative fluxes, both in absolute magnitude of the flux and its temporal variability (L. D. Riihimaki et al. 2021, unpublished manuscript). A frequency density of single-layer cloud types from Lim et al. (2019) from 2014 through 2017 at ARM SGP shows that low clouds and cirrus were the dominant daytime cloud regimes (Fig. 2a); since our cloud-type classifier development is dependent upon solar radiation, only observations when the solar zenith angle was less than 80° (more than 10° above the horizon) are examined.

As a consequence of the distinct solar transmission properties associated with these different cloud types, it is imperative to separate the low clouds into two distinct cloud types: 1) low stratiform and 2) low cumulus. The relationship between cloud fraction and the ratio of diffuse to total solar irradiance is used to distinguish these cloud types. Stratiform clouds are generally associated with large fractional sky coverage (cloud fraction). Because of their overcast nature, the direct-beam solar radiance undergoes significant scattering and therefore diffuse radiation is the dominant contributor (often 100% diffuse) to the total surface broadband irradiance (Fig. 2b). Analogously, shallow cumulus are broken clouds, and in a spatial context, each cloud structure is small compared to stratiform clouds. These small cloud elements advect through the hemispheric footprint of a surface observatory, causing a cloud fraction frequently ranging from 10% to 70% sky coverage (Fig. 2b). Because of this, the ratio of diffuse to total shortwave radiation is notably reduced, typically ranging between 10% and 40% under shallow cumulus conditions. The separation of these two modes in the probability distribution frequency is used to differentiate low clouds into low stratiform and shallow cumulus cloud types. Cases to the left of the red line in the cloud fraction—solar diffuse/total phase space are classified as shallow cumulus and those to the right as low stratiform. To confirm a successful separation of low stratiform from low cumulus clouds, Fig. 3 illustrates the different behavior in shortwave transmissivity associated with the two separated cloud types, where shortwave transmissivity is the ratio of observed downwelling shortwave irradiance to the clear-sky downwelling irradiance. Shortwave transmissivity observations for consecutive (at least 2 h of persistent cloud type) low stratiform and low cumulus periods at SGP during 2014–18 were compiled. The density distributions for 5-min standard deviation of shortwave transmissivity capture the different solar variability characteristics associated with the two cloud types (Fig. 3a). While the low stratiform clouds are relatively invariant in shortwave transmissivity, more variability in the shortwave irradiance is found when low cumulus periods are observed. The variability occurs due to the broken sky coverage associated with shallow cumulus, resulting in periods

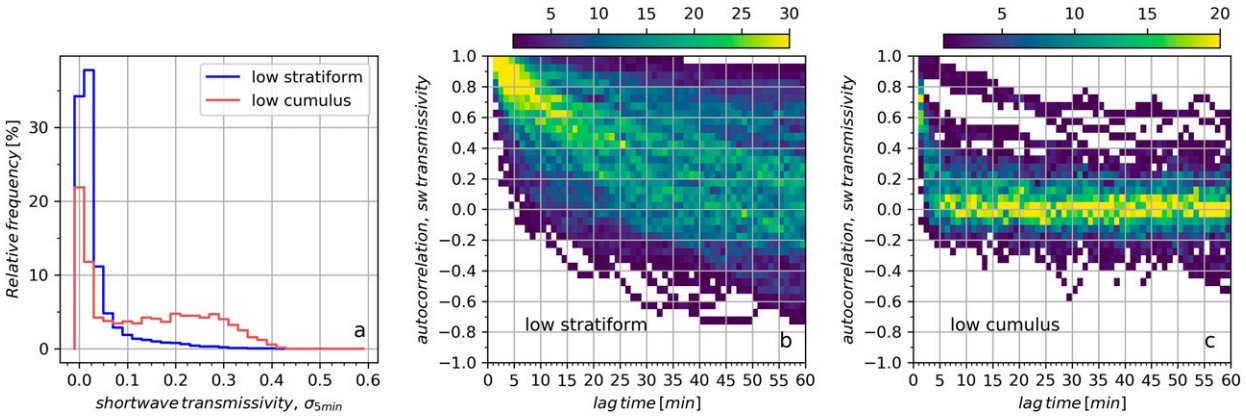


FIG. 3. Statistical properties of shortwave transmissivity during observed, consecutive low stratiform or low cumulus cloud events. (a) Relative frequency distribution of 5-min standard deviation of shortwave transmissivity for low stratiform (blue) and low cumulus (red) events. Also shown are frequency density (counts; shading) of the autocorrelation for shortwave transmissivity per lag time (in minutes) for (b) low stratiform and (c) low cumulus periods. At least 120 min of a consecutive cloud event must have been observed to be included in the analysis. For low stratiform, there were 311 events; for low cumulus, there were 124. Data are from 2014 to 2018 at ARM SGP.

that oscillate between unimpeded direct solar radiation and more fully obscured direct radiation (Fig. 2b). 2D histograms of the autocorrelation of shortwave transmissivity for different lag periods ranging from 1 to 60 min highlight the radiative separation of these cloud regimes. Low stratiform clouds have an autocorrelation that remains positive on lag periods well above 30 min (Fig. 3b), consistent with the small variability in transmissivity resulting from more homogeneous overcast sky conditions. Oppositely, autocorrelations for the low cumulus drop off rapidly and after 5 min there is effectively no correlation in the solar transmissivity (Fig. 3c) as a result of the broken nature of these clouds. This time scale is consistent with the order of the decorrelation time scale reported in Kassianov et al. (2005).

The other cloud regimes in Fig. 2a are much less frequent than low clouds and cirrus. To improve separability of cloud regimes during the training phase of the machine-learning classifier, congestus clouds are combined with deep convection, and altocumulus clouds are combined with altostratus cases; this is motivated by these cloud regimes having similar geometric thresholds in their original classification (Flynn et al. 2017; Lim et al. 2019).

The measurement principles of the lidar/radar sensors and the ARSCL retrievals of cloud boundaries effectively screen the atmosphere for separation between vertically displaced cloud layers (e.g., Kollias et al. 2007b). Exploiting the strengths of the active remote sensors, Lim et al. (2019) classified up to 10 cloud layers into the seven regimes described in Fig. 2a. However, for multiple cloud layers, this level of detail is far more advanced than is anticipated that broadband radiation measurements may provide. The cloud classification regimes developed here include multiple cloud layers, but instead of multiple vertical cloud types, multilayered cloud scenes are classified as 1) low–high clouds, 2) low–midclouds, or 3) mid–high clouds. It is anticipated the radiative signatures of multiple cloud layers will introduce ambiguity in the separation process during the training phase of the random-forest classifier. The level of separability of multilayer clouds in terms of solar radiation impacts is discussed in section 3a.

A total of eight cloud regimes, plus an additional clear-sky regime, have been created from the observations at ARM SGP. These classifications are 0) clear sky, 1) low stratiform, 2) low cumulus, 3) congestus/deep convection, 4) midlevel altocumulus/altostratus, 5) high-level cirrostratus/anvil, 6) high cirrus, 7) multilayer low–high, 8) multilayer low–mid, and 9) multilayer mid–high.

### c. *Training the random-forest classifier*

This study utilizes the machine-learning random-forest open-source software kit "Scikit-learn" (Pedregosa et al. 2011) for Python. For its application here, the random-forest classifier involves developing statistical relationships between the extended cloud classifications, surface radiation, and cloud observations (features) through a supervised learning process (Breiman 2001). The supervision portion comes from training the model against truth labels, which are the independent cloud-type labels for this particular random-forest classifier. The different input features are statistically tested and optimized following a true–false decision tree logic to determine the feature's importance in splitting branches of a decision tree such that the combination of features converges to an observed cloud classification (label). The 12 features (Table 1) and cloud labels are randomly sampled from the population using a bootstrap with replacement sampling strategy for training each node, or branch, within the decision tree. Random sampling from the population reduces the risk of overfitting the classification as each node within tree is built upon a different collection of inputs and labels (Breiman 2001).

Our forest consists of 100 decision trees, where each node of every tree is presented the 12 input features from which a random sampling of a subset of these features occurs in order to develop the decision tree logic. Sensitivity tests were made with as few as 5 trees up to as many as 500 trees; increasing the number of trees up to 100 nominally increased the accuracy.

However, continual increases above 100 trees increased the model's data storage (the full decision tree logic for each tree in the forest must be saved) with very modest to insignificant increases in classifier accuracy (not shown). It was therefore determined that 100 trees making up the forest was appropriate for the model. The classifier thus produces 100 equally plausible classifications based on the same input features vector. Using a majority rule, the classification that occurs most frequently of these 100 estimates is deemed the modeled classification for a particular series of measurement inputs.

Other parameter options and settings of Scikit-learn's random-forest classifier software were specified to control the physical setup of the training process. Of the 12 input features available at each node, 3 (approximating the square root of 12) were randomly chosen for determining the true–false logic; this is a recommended default setting of the Scikit-learn random-forest classifier software. A requirement was placed on the minimum number of samples necessary to determine a split at a node to 10 samples to avoid overfitting toward undersampled events; limiting the depth of the tree branches was detrimental to the accuracy and therefore a limit was not specified. Ultimately, the model parameters were determined through cross validation of model accuracy and minimization of computational resources and model data size/storage.

Development of the classifier was based on daytime-only (solar zenith angle $< 80°$) input features and classification labels from ARM SGP over a 4-yr period, 2014–17. Observations are made on a 1-min frequency and therefore the random-forest model was trained to classify clouds using features at the same temporal frequency. In total nearly 620 000 min (10 300 h) of daytime features and classifications were utilized. These data were split randomly at 70%–30%, for classifier training and validation, respectively. Observational data from 2018 was reserved for independent algorithm testing.

## 3. Results

### *a. Validation and independent evaluation of cloud classifier*

Following the completion of training the forest, a feature importance metric can be determined based on the 70% training data subset. Feature importance quantifies how important a specific feature is in determining a decision tree's logic structure and ultimately the feature's impact on improving the prediction. Calculation of this "impurity," or reduction in mean standard error, identifies the effectiveness of an input predictor to cause a relatively large decrease in the impurity (McGovern et al. 2019). The Scikit-learn random-forest classifier software computes the Gini impurity score for each predictor feature for the training data subset. Of the 12 features included in the model, those most critical for improving prediction accuracy are ceilometer cloud-base height and shortwave cloud fractional coverage (Fig. 4). The remaining features have smaller importance scores and can be divided into two groups with nearly equivalent scores. The group with the lowest importance scores includes features based only on radiation measurements and

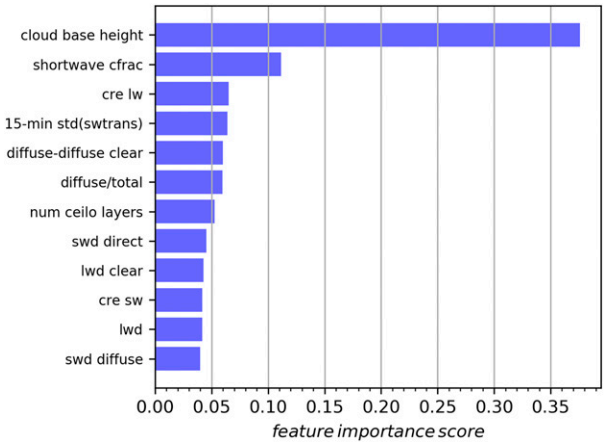


FIG. 4. Feature importance scores determined from the subset of training data at ARM SGP during 2014–17.

not derived cloud products [with the exception of shortwave cloud radiative effect (cre sw)], indicating the RadFlux products, as well as direct cloud measurements from the ceilometer, contribute valuable signals specific to different cloud regimes.

Relative frequency distributions (RFDs) of the observed cloud-regime classifications for the validation subset of observations at SGP during 2014–17 are shown in Fig. 5 (black). Clear-sky and high cirrus dominate the daytime sky conditions, followed by low stratiform clouds; these distributions are qualitatively similar to the single-layer distributions of Lim et al. (2019) in Fig. 2a. When multiple cloud layers were observed they were most frequently of the low–high cloud variety; the other multilayered cloud types were distinctly less common. Cloud-type distributions predicted by the random-forest classifier for the same period are shown in blue. The classification model successfully captures the climatological regime distributions at SGP. The model is well suited to predict the clear-sky and high cirrus clouds. Slight overestimation of low stratiform and low/shallow cumulus is present, generally compensated by an underestimation of multilayered clouds. In light of these small differences, the accuracy of the predicted cloud classifier on the 2014–17 validation subset is a remarkable 90%.

The full observational period of 2018 at SGP was withheld from the training-evaluation phase to have a completely independent year to test the robustness of the training-evaluation efforts developed from observations of the four previous years. The motivation for a fully independent year was to test whether the classifier could predict the annual cycle of daytime cloud regimes without any influence from quasi-persistent sky conditions and surface measurements that could potentially bias the predictions. As a hypothetical example, consider two days: one with homogeneous low stratiform clouds and the other experiencing high cirrus. It is plausible that the surface radiation and cloud-property measurements would be relatively consistent (homogeneous) during the respective cloud regimes. Therefore, subsampling 70% of these days to train the classifier and evaluating the other 30%

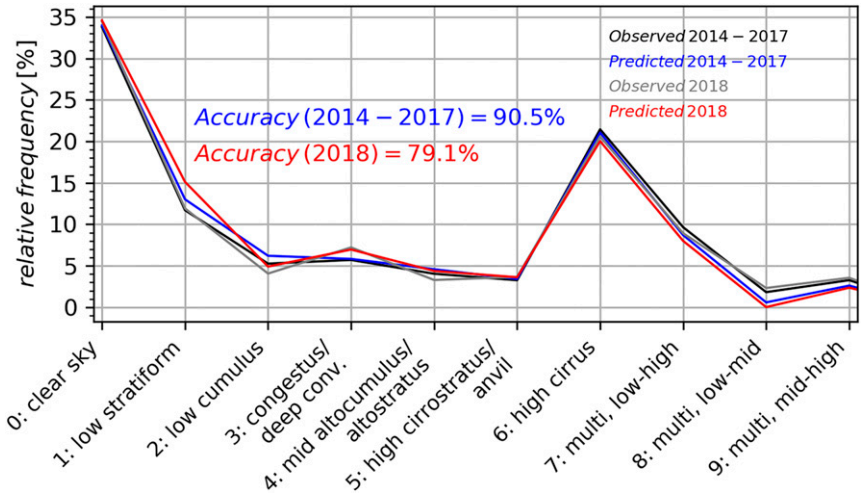


FIG. 5. RFDs (%) of observed and predicted cloud regimes (0–9) for the evaluation subset of 2014–17 (black: observed; blue: predicted) and the independent full year of 2018 at ARM SGP (gray: observed; red: predicted). Model-predicted accuracies for the periods are included.

that consider relatively consistent radiative and cloud-property signatures as the training subsample, the accuracy of the model would be expected to be high.

A reduction in accuracy of the cloud predictions by approximately 10% during 2018 (Fig. 5, red) supports the original concerns that evaluating the classifier using a subsample of the same period that was used to train the model artificially inflates the accuracy estimate, that is, slight overfitting to homogeneous cloud scenes. Despite the accuracy reduction, this model predicts the correct cloud type 8 of 10 times based entirely on surface-based measurements and is likely well within any uncertainty propagation due to the radar and lidar measurements, ARSCL retrievals, and cloud boundary threshold categories used to create the observed cloud type (Lim et al. 2019). This suggests that 80% (68% without the clear-sky regime) is the maximum achievable, yet respectable, accuracy for the cloud-regime classifier. It is encouraging that the 2018 annual climatology of cloud regimes remains well predicted by the cloud classifier. The most obvious biases include a nearly 4% overprediction in the amount of low stratiform clouds, while the multilayer cloud categories are always underpredicted.

To better determine how predicted cloud regimes are misclassified, a confusion matrix relating observed and predicted regimes during the fully independent year 2018 is created. The matrix in Fig. 6 shows the RFD (%) of the observed cloud classifications (ordinate) against the predicted classifications (abscissa); distributions sum to 100% along the abscissa and therefore the confusion matrix is an ideal tool to visualize misclassification behavior. The distributions are highest along the diagonal of the matrix, highlighting the classifier is well suited for classifying clear skies and single-layer cloud regimes (regimes 0–6); the dominant sky regimes from Fig. 5, namely clear skies, low stratiform and low cumulus, and high cirrus clouds reveal an accuracy greater than 80%. The remaining single-layer regimes 3–5 have RFDs that are more dispersed around the diagonal, a sign of reduced accuracy for these types.

While the classifier is largely able to separate single-layer cloud types, especially for those most frequently occurring, its ability to distinguish among multiple cloud layers (regimes 7–9) is less accurate. The accuracy of these cloud regimes along the diagonal is at best 35% for the low–high multilayer clouds; low–midmultilayered clouds (regime 8) are completely absent, while predictions of mid–high multilayers (regime 9) are frequently misclassified as a single-layer regime. While the model does include active-sensor ceilometer profiling measurements of the number of cloud layers identified, these measurements are zenith viewing and therefore only observe multiple layers when they are overlapping above the laser and before the ceilometer beam is fully attenuated; the ceilometer will miss higher cloud layers if an optically thick cloud layer is sampled below. The other radiative measurements are primarily flux quantities that are integrated across the full hemispheric view, and as such provide limited information to characterize multiple cloud layers.

In addition to observing limitations of the instrumentation, multilayer cloud regimes are relatively infrequent (Fig. 5); as such, the training process of the random-forest classifier is imbalanced. However, multilayered low–high (type 7) clouds do occur relatively frequently, approximately 10% (Fig. 5), and it is apparent that the predictions of multilayered low–high clouds are frequently mistyped as single-layer low stratiform clouds (Fig. 6).

In Fig. 7, statistical daytime shortwave cloud transmissivity ranges at SGP for 2014–17 are presented for each observed (truth) cloud regime. Overlap in the percentile ranges of shortwave transmissivity is found among the cloud regimes, indicating that transmissivity is not an exclusively separable parameter between the nine different cloud regimes. However, the distributions do show intriguing physical properties related to cloud type and associated cloud–shortwave radiation

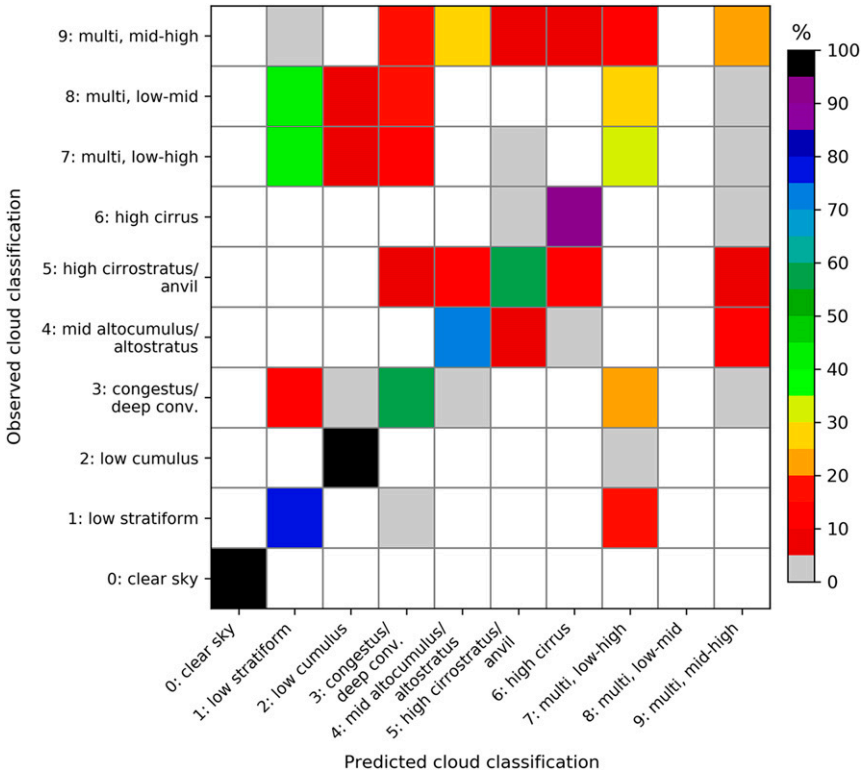


FIG. 6. Confusion matrix for observed cloud classification (ordinate) vs predicted cloud classification (abscissa) from ARM SGP 2018. Colors represent the frequency of occurrence of each predicted cloud regime (frequencies sum to 100% along the abscissa).

interactions. The two low cloud types (low stratiform and low cumulus) exert a different influence on surface shortwave radiation, with solar transmission frequently as large or larger than clear skies (shortwave transmissivity at or above unity) during the shallow cumulus regime; nearly 50% of shortwave transmissivity was larger than unity and this positive shortwave cloud forcing is a result of partially cloudy skies causing significant 3D radiative transfer effects combined with

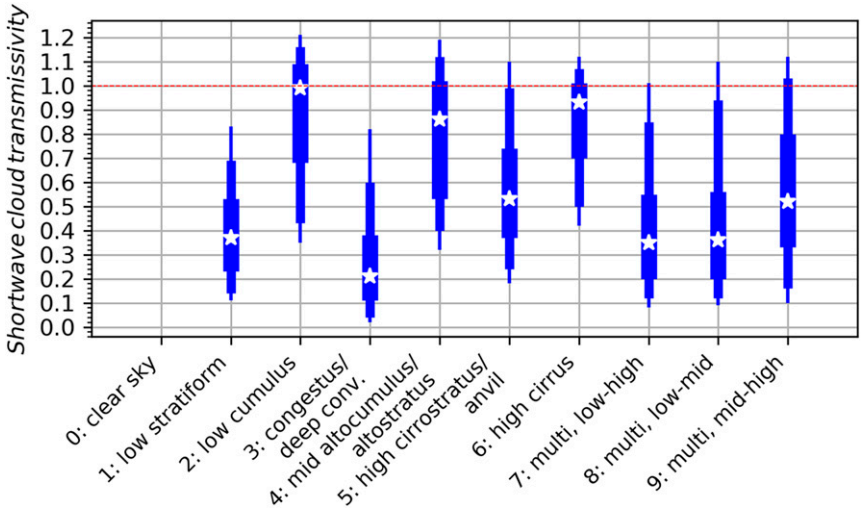


FIG. 7. The median (stars) and 25th–75th, 10th–90th, and 5th–95th percentile ranges of observed shortwave cloud transmissivity per cloud type at ARM SGP for 2014–17.

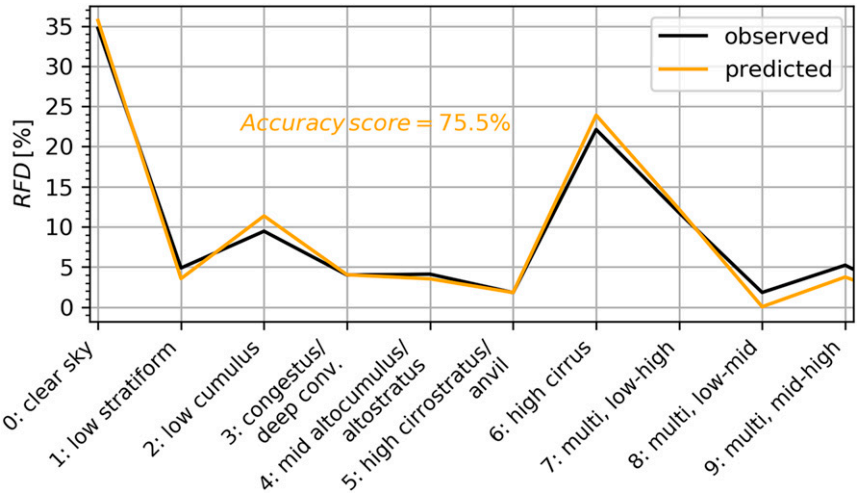


FIG. 8. Observed (black) and predicted (orange) cloud-regime climatology for 2008–13 at ARM TWP in Darwin.

frequent, unabated direct solar irradiance (L. D. Riihimaki et al. 2021, unpublished manuscript). Midlevel altocumulus and altostratus (type 4) and high cirrus (type 6) regimes have similar transmissivity ranges as the low cumulus but are still separable from low stratiform. However, the more spatially homogeneous nature of these cloud types reduces the impact of 3D radiative transfer, leading to a reduction in the positive shortwave radiative forcing of these cloud types relative to low cumulus. Shortwave transmissivity distributions further reveal that multilayered cloud regimes (type 7–9) show a very similar distribution to that of low stratiform clouds; this is especially true for the low–high and low–midcloud regimes, while the transmissivity range was shifted marginally higher for the mid–high multilayered cloud regime.

The similarities in shortwave transmissivity between low stratiform and multilayered clouds with one low-level cloud layer indicate the lowest cloud layer is largely responsible for modulating cloud transmissivity. This is an important result since the cloud classifier frequently misclassified low–high multilayered clouds as low stratiform clouds (Fig. 6).

In terms of shortwave radiation, these results indicate there is little information lost in misclassifying a low–high multilayered cloud scene as a low stratiform scene; the impact on shortwave fluxes is approximately similar. The skewness toward higher transmissivity of the 90th and 95th percentiles for the low–high and low–midmultilayered regimes compared to the low stratiform is likely related to the type of low cloud present. If the lowest cloud type is shallow cumulus, it would be expected that shortwave transmissivity should be larger than for an overcast stratiform cloud in a similar manner as shown in Fig. 7 for regimes 1 and 2. To test the multilayer cloud predictions, observations were limited to only single-layer cases and an increase in accuracy of the cloud predictions from approximately 80% (Fig. 5) to 88% (not shown) was found. Based on this increased accuracy, we assert that the model classifier is well poised to separate cloud regimes based on their surface radiative impact, even if the physical cloud type may be mislabeled.

### b. *Evaluation of cloud classifier at Darwin, Australia*

Even though the cloud classifier was trained on observations from the continental United States, the multiyear model training process is expected to sample a widespread array of cloud types. In this regard, it is anticipated that the cloud classification model can be applied to observatories within climate regimes that differ from the southern Great Plains. To test the validity of the classifier, an independent examination was performed using radiation and ceilometer measurements, together with collocated zenith-viewing radar and lidar instruments from DOE's ARM Tropical Western Pacific (TWP) site at Darwin (Long et al. 2016). Measurements from 2008 to 2013 at TWP were processed and an observed, extended cloud-type dataset (as in section 2b) was produced for evaluation purposes. When compared to SGP (Fig. 5), marginal differences are found in the observed cloud-regime climatologies at TWP (Fig. 8); the main difference is a reversal in the frequencies of low stratiform and low cumulus between the two observatories. The classifier successfully predicts this shift in cloud regimes. Furthermore, the accuracy is just over 75% (63% without clear-sky regime), indicating less than a 4% decrease in accuracy than was reported at SGP for 2018. This relatively high accuracy is encouraging since the classifier was trained on measurements from the southern Great Plains in the continental United States. Additionally, the evaluation at TWP occurred for a 6-yr period, and as such interdecadal trends and forcings should also manifest during this time period.

Like at SGP, the confusion matrix for TWP cloud predictions (Fig. 9) reveals the best agreement to the single-layer observed clouds. Based on the increased dispersion from the diagonal, the single-layer cloud predictions at TWP are slightly more variable than at SGP. This is likely a caveat of the training process that relied upon continental measurements at

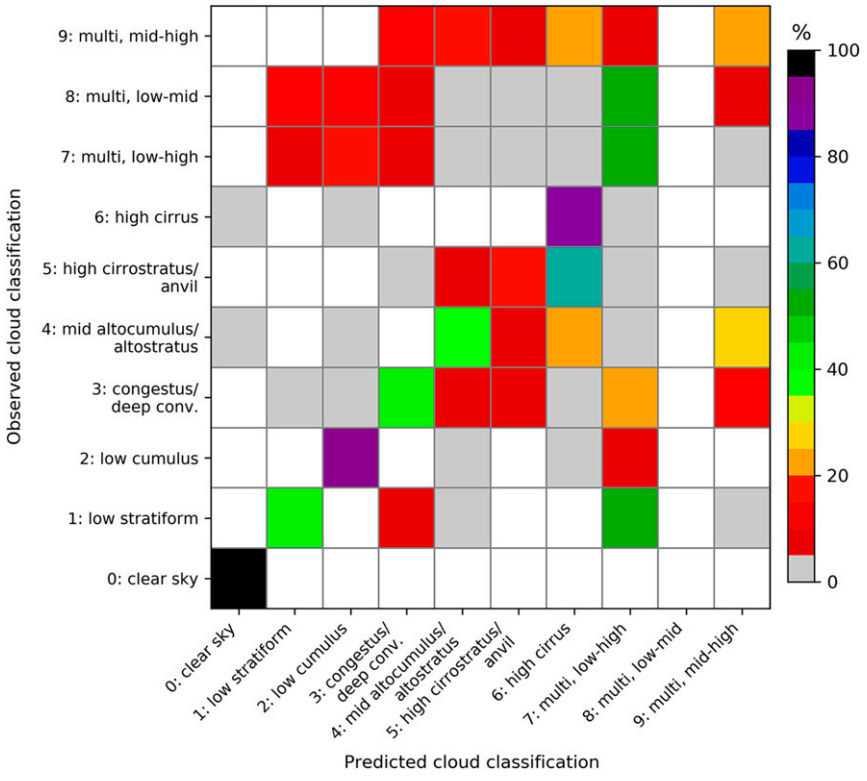


FIG. 9. Confusion matrix from 2008 to 2013 at ARM TWP at Darwin.

SGP, where systematic differences in cloud microphysical and macrophysical properties at the subtropical TWP observatory may cause cloud–radiative interaction differences between locations. For example, the high cirrostratus/anvil cloud type (type 5) is most frequently classed as a high cirrus (type 5), a misclassification not found at SGP (Fig. 6). Nevertheless, the frequencies of these single-layer cloud misclassifications are primarily below 5% and when misclassified, they often become predicted as a cloud regime having a similar shortwave transmissivity range (Fig. 7) as observed.

Multiple cloud layers continue to prove a challenge for the random-forest classifier. Recall the accuracy of the low–midmultilayered clouds, the most frequently observed multilayer cloud type, for SGP was ~35% (Fig. 6, cloud type 7). Figure 9 shows an unexpected improvement in the accuracy of this classification at Darwin. The improvement is connected to a reduction in misclassification as single-layer low stratiform (type 1) at TWP compared to SGP. In fact, whereas at SGP the low–high clouds were most frequently misclassified as low stratiform, at TWP these low–high clouds are more frequently misclassified as low cumulus (type 2). The shift is reflective of the differences in cloud-type climatologies observed between the continental SGP and subtropical TWP observatories. Multilayer low–midclouds (type 8) fail to be predicted by our classifier, just like at SGP (Fig. 6). The observed occurrence of these clouds was below 3% at SGP and TWP. Based on the similarity in observed shortwave transmissivity (Fig. 7), the low–midmultilayered regime could be removed and instead low–midclouds could be reclassified as the low–high cloud regime, which was the regime that low–midclouds were predominantly being misclassified (Figs. 6 and 9).

### c. *Example application of cloud-regime predictions at SURFRAD observatories*

Observations from the SURFRAD observatories were used as input to predict cloud classification climatologies at selected stations. Since there are no available independent cloud-type measurements at the SURFRAD observatories, cloud types are explored in terms of the relative occurrence at each location. Frequency distributions for six of the SURFRAD observatories are shown in Fig. 10. Beginning in spring 2018 and continuing through autumn 2019, a Vaisala CL51 ceilometer was installed at each SURFRAD station. The cloud climatologies in Fig. 10 are only representative of the past 6 months to nearly 2 yr; (Fig. 10a–c) include at least one full annual cycle (from spring/summer 2018 to early winter 2020) of observations, and (Fig. 10d–f) are valid for less than one full year. Daytime sky conditions at stations with a full year of observations are dominated by clear skies, followed by high cirrus. Table Mountain, Colorado, experiences more clear-sky conditions than Fort Peck, Montana, and Goodwin Creek, Mississippi, which both have a greater presence of lower-level and multilevel clouds. These distinctions in cloud regime may be connected with the topographic influence of the Front Range of the Rocky Mountains on cloud formation and cloud life cycle.

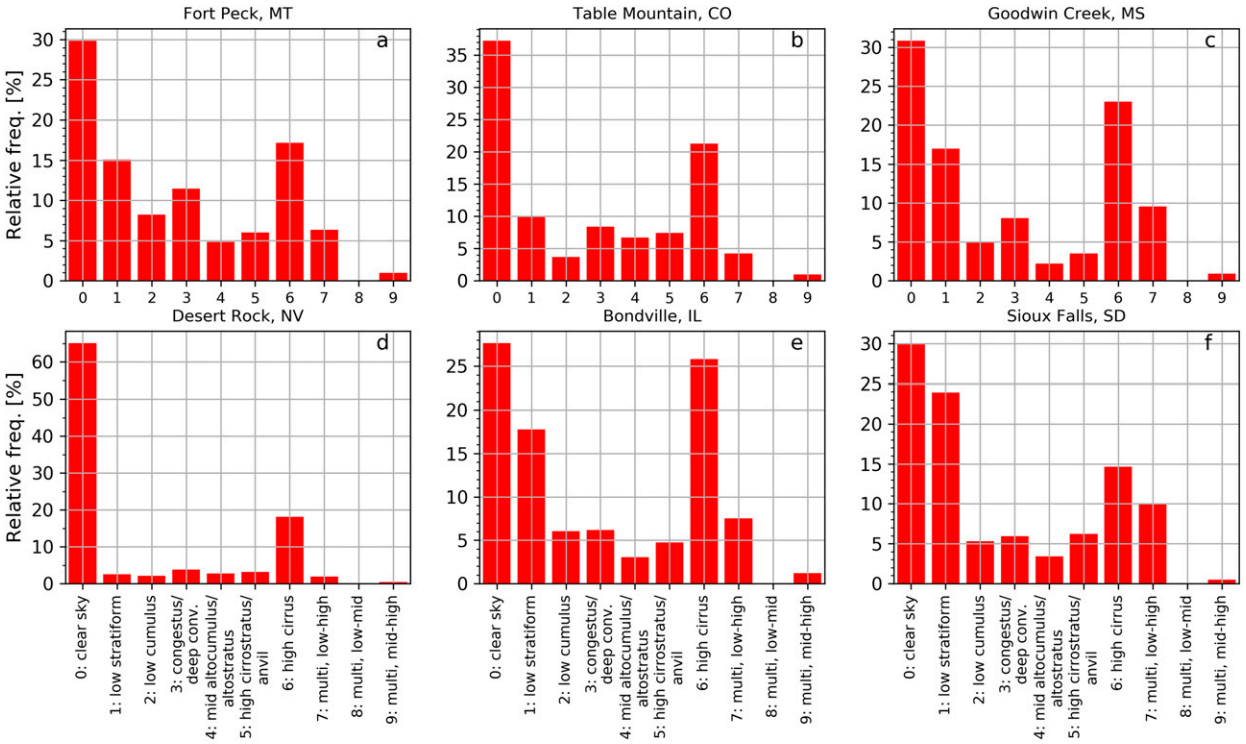


FIG. 10. Predicted cloud-regime climatologies from the following SURFRAD observatories: (a) Fort Peck (March 2018–January 2020), (b) Table Mountain (March 2018–January 2020), (c) Goodwin Creek (June 2018–January 2020), (d) Desert Rock (March 2019–February 2020), (e) Bondville (May 2019–January 2020), and (f) Sioux Falls (August 2019–January 2020). See Fig. 1 for the spatial distribution of the SURFRAD observatories. Note the differing scales of the distributions on the *y* axis.

The classification climatologies for Desert Rock, Nevada (Fig. 10d), and Bondville, Illinois (Fig. 10e), represent less than one full year, from early and late spring 2019, respectively, through January 2020. The distributions clearly indicate the behavior of different climate zones on the cloud regimes influencing the stations. At Desert Rock, clear skies dominate as the warm and arid desert climate inhibits low level, surface driven cloud formation (Fig. 10d). The cloud regimes at Bondville (Fig. 10e) are similar to those at Goodwin Creek, with the primary difference being a shift toward more cloudiness at Bondville. The increase in cloudiness is driven by an enhancement in the relative occurrence of high cirrus. The ceilometer at Sioux Falls, South Dakota, was not installed until August 2019, and therefore the cloud climatology is representative of only two seasons: autumn and winter. These two seasons show a marked shift in cloud regimes, where low stratiform cloud frequency surpasses the cirrus frequency. Because of the lack of at least one annual cycle, it is not possible to conclude whether these regime occurrences are systematic of the annual cycle or simply a caveat of the limited seasons represented.

## 4. Discussion and summary

This study reports on the method of a new cloud-regime classification tool that relies upon surface-based radiation and cloud measurements to predict cloud type. Independent evaluation efforts from different climate regions and time scales reveal our random-forest cloud classifier predicts cloud type with accuracy of 75%–80%.

Independent evaluations of the random-forest cloud-regime classification model prove the model is capable of accurately predicting cloud type for a wide variety of applications. Independent evaluations of the cloud classification accuracy were as high as 80%. Such accuracy is well within the stated accuracies of cloud classification methods based on clustering of visible sky images (Heinle et al. 2010; Zhuo et al. 2014). However, the accuracy testing in those studies was based on orders of magnitude fewer independent observations, which may result in biased accuracies. Here, seven years of 1-min frequency observational data were used across full annual cycles from two very different geographic regions to evaluate our classification model, which resulted in climatological accuracies greater than 75%. We anticipate the accuracy of future cloud-type predictions using input features from NOAA's SURFRAD observatory network will be closer to the 80% level considering the model was trained on measurements of radiation and clouds from the continental United States. The largest errors in predicted regimes emerge when multiple cloud layers are observed. An inability to sufficiently sample the hemispherical vertical structure of cloud layers, together with ambiguity in radiative signatures between single-layer and multilayered cloud structures, contributed to the prediction errors.

To quantify the error introduced by misclassification of multilayered clouds, an additional classifier training process was performed, where observations of multilayered clouds were ignored and thus only single-layer clouds regimes 0 through 6 were used for training the model (not shown). From this test, the climatological accuracy for the independent evaluation for 2018 at SGP increased to 88%, an increase of 9%. From this accuracy increase, it is concluded that misclassification of the multilayered clouds account for nearly 45% of the accuracy error. Depending upon the application, it may be sufficient to exclude the multilayer cloud-regime predictions altogether or otherwise combine all the three multilayered classes into a single multilayered clouds regime.

Of the multilayered clouds, the low–high regime 7 was the most frequently occurring type. When they were misclassified, they were typically designated as single-layer low stratiform cloud regime 1 by the classifier. From Fig. 7, the shortwave transmissivity has a similar range for both of these cloud types. Therefore, while the model may be penalized in accuracy for the common misclassification, the radiative signatures it was trained upon to determine the classification are, in general, consistent among the two cloud regimes. And because of the methods applied to develop a separation between low stratiform and low cumulus clouds (Figs. 2 and 3), the radiative signatures of low–high multilayered and low stratiform single-layer clouds are physically consistent and are not being falsely identified as shallow cumulus. Depending on the application, the misclassification of the low–high regime as low stratiform cloud may give reasonable results, for example in determining the radiative impacts of different cloud types. With respect to cloud–radiative interactions, results reveal the cloud classifier has success in distinguishing cloud regimes based on their separability in shortwave radiation characteristics. In this respect, one can justify a model classification tool whose inaccuracy in physical cloud type is ~20%, knowing that the misclassifications are frequently within a similar surface shortwave radiation flux regime.

The input measurements were also chosen such that any measurement station recording similar, high-quality measurements could run this cloud classification model on those data; a direct example of this is that measurements from ARM SGP were used to develop the model framework but applied the same framework to SURFRAD observatories where similar measurements are routinely made around the continental United States. It is critical to measure the direct and diffuse irradiance components of solar radiation as these components are needed for the predictions. Essentially, the radiative measurements should be of utmost quality and contain the necessary components needed in order to run the Radiative Flux Analysis processing software (Long and Ackerman 2000; Long et al. 2006; Long and Turner 2008). Among a wide array of products, Radiative Flux Analysis produces the shortwave cloud fractional coverage, shortwave transmissivity, and clear-sky radiative flux products needed as inputs.

Using surface radiation measurements as input features for the classification model was a strategic development choice because these measurements are either the same or similar to the types of measurements required for solar energy applications. Fluxes of direct and diffuse shortwave radiation, critical parameters necessary for optimization of plane of array photovoltaic solar panels, are dependent upon the cloud type and cloud properties overhead. Having an accurate understanding of how the cloud-type variability translates to variability in the surface solar measurements is paramount for a number of reasons. As an example, climatologies of cloud regimes at different geographical locations can be created in order to assist in optimizing where solar energy farms should be deployed to maximize solar energy production. Knowing how different cloud regimes impact the variability in surface solar fluxes can provide a level of understanding that even if a certain cloudy regime is frequently observed in a specific region, the shortwave transmissivity associated with that regime may still result in a large solar energy production (e.g., L. D. Riihimaki et al. 2021, unpublished manuscript). For example, at Table Mountain (Fig. 10b), the relatively high occurrence of cirrus clouds that typically exert shortwave transmissivity greater than 70% (Fig. 7), and the relatively low frequency of low stratiform clouds presents this location as a well-suited candidate for solar energy operations.

Characterizing variability in surface radiative fluxes as a function of cloud-regime classification provides a tool to improve fundamental understanding of cloud–radiative interactions. This is especially true in terms of numerical model evaluation and development. The input features show in Table 1 are either standard model output parameters, or they can be extracted from numerical weather models. Therefore, the random-forest classifier can be used to predict cloud type using numerical model output, which can serve as a one-to-one comparison of predicted cloud regimes from surface radiation measurements. Traditionally, surface radiation from numerical models is compared to detailed observations using statistical measures such as mean bias error and root mean squared error. The statistics provide a baseline of model performance in a bulk sense. However, having knowledge of the type of cloud and potentially how the biases may change under different cloud types would be extremely useful for model developers. In this respect, the cloud-regime classifier could facilitate these model evaluation efforts. Furthermore, the observed cloud regimes can be used to test and evaluate the performance of the life cycle of individual events. For example, the cloud regimes could be used to identify periods of persistent stratiform clouds, and observations such as the timing of cloud breakup could be used to evaluate whether numerical weather models capture the same cloud regime and how well they represent the life cycle and model processes leading to stratiform breakup.

*Acknowledgments.* This work was partially supported by the National Oceanic and Atmospheric Administration (NOAA) Atmospheric Science for Renewable Energy (ASRE) program. The authors thank Christina Kummler for her insightful comments on the paper. We are also indebted to the three anonymous reviewers who provided feedback on the paper and greatly improved the scope of this work.

*Data availability statement.* SURFRAD radiation and cloud data products are freely available for the community from the following FTP server: ftp://aftp.cmdl.noaa.gov/data/radiation/surfrad/. Data from the ARM Southern Great Plains (SGP) and Tropical Western Pacific (TWP) sites, including the cloud-type product produced by Lim et al. (2019), are freely available from the ARM Data Center archive (https://adc.arm.gov/discovery/#/).

## REFERENCES


Ahlgrimm, M., and R. Forbes, 2012: The impact of low clouds on surface shortwave radiation in the ECMWF model. *Mon. Wea. Rev.*, **140**, 3783–3794, https://doi.org/10.1175/MWR-D-11-00316.1.

Ahrens, C. D., 2012: *Meteorology Today: An Introduction to Weather, Climate, and the Environment*. Cengage Learning, 621 pp.

Augustine, J. A., G. B. Hodges, C. R. Cornwall, J. J. Michalsky, and C. I. Medina, 2005: An update on SURFRAD—The GCOS surface radiation budget network for the continental United States. *J. Atmos. Oceanic Technol.*, **22**, 1460–1472, https://doi.org/10.1175/JTECH1806.1.

Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Campbell, J. R., D. L. Hlavak, E. J. Welton, C. J. Flynn, D. D. Turner, J. D. Spinhirne, V. S. Scott III, and I. H. Hwang, 2002: Full-time, eye-safe cloud and aerosol lidar observation at atmospheric radiation measurement program sites: Instruments and data processing. *J. Atmos. Oceanic Technol.*, **19**, 431–442, https://doi.org/10.1175/1520-0426(2002)019<0431:FTESCA>2.0.CO;2.

Chen, Y., and A. D. Del Genio, 2009: Evaluation of tropical cloud regimes in observations and a general circulation model. *Climate Dyn.*, **32**, 355–369, https://doi.org/10.1007/s00382-008-0386-6.

Flynn, D., Y. Shi, K.-S. Lim, and L. Riihimaki, 2017: Cloud Type Classification (cldtype) Value Added Product. DOE/SC-ARM-TR-200 Tech. Rep., 10 pp., https://www.arm.gov/publications/tech_reports/doe-sc-arm-tr-200.pdf.

Heinle, A., A. Macke, and A. Srivastav, 2010: Automatic cloud classification of whole sky images. *Atmos. Meas. Tech.*, **3**, 557–567, https://doi.org/10.5194/amt-3-557-2010.

Jakob, C., and G. Tselioudis, 2003: Objective identification of cloud regimes in the tropical western Pacific. *Geophys. Res. Lett.*, **30**, 2082, https://doi.org/10.1029/2003GL018367.

Jin, D., L. Oreopoulos, and D. Lee, 2017: Regime-based evaluation of cloudiness in CMIP5 models. *Climate Dyn.*, **48**, 89–112, https://doi.org/10.1007/s00382-016-3064-0.

Karlsson, K.-G., and Coauthors, 2017: CLARA-A2: The second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data. *Atmos. Chem. Phys.*, **17**, 5809–5928, https://doi.org/10.5194/acp-17-5809-2017.

Kassianov, E., C. N. Long, and M. Ovtchinnikov, 2005: Cloud sky cover versus cloud fraction: Whole-sky simulations and observations. *J. Appl. Meteor.*, **44**, 86–98, https://doi.org/10.1175/JAM-2184.1.

Kazantzidis, A., P. Tzoumanikas, A. F. Bais, S. Fotopoulos, and G. Economou, 2012: Cloud detection and classification with the use of whole-sky ground-based images. *Atmos. Res.*, **113**, 80–88, https://doi.org/10.1016/j.atmosres.2012.05.005.

Kollias, P., G. Tselioudis, and B. A. Albrecht, 2007a: Cloud climatology at the southern Great Plains and the layer structure, drizzle, and atmospheric modes of continental stratus. *J. Geophys. Res.*, **112**, D09116, https://doi.org/10.1029/2006JD007307.

——, E. E. Clothiaux, M. A. Miller, B. A. Albrecht, G. L. Stephens, and T. P. Ackerman, 2007b: Millimeter-wavelength radars. New frontier in atmospheric cloud and precipitation research. *Bull. Amer. Meteor. Soc.*, **88**, 1608–1624, https://doi.org/10.1175/BAMS-88-10-1608.

——, and Coauthors, 2016: Development and applications of ARM millimeter-wavelength cloud radars. *The Atmospheric Radiation Measurement (ARM) Program: The First 20 Years, Meteor. Monogr.*, No. 57, Amer. Meteor. Soc., https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0037.1.

Leinonen, J., M. D. Lebsock, L. Oreopoulos, and N. Cho, 2016: Interregional differences in MODIS-derived cloud regimes. *J. Geophys. Res. Atmos.*, **121**, 11 648–11 665, https://doi.org/10.1002/2016JD025193.

Lim, K.-S. S., and Coauthors, 2019: Long-term retrievals of cloud type and fair-weather shallow cumulus events at the ARM SGP site. *J. Atmos. Oceanic Technol.*, **36**, 2031–2043, https://doi.org/10.1175/JTECH-D-18-0215.1.

Long, C. N., and T. P. Ackerman, 2000: Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. *J. Geophys. Res.*, **105**, 15 609–15 626, https://doi.org/10.1029/2000JD900077.

——, and D. D. Turner, 2008: A method for continuous estimation of clear-sky downwelling longwave radiative flux developed using ARM surface measurements. *J. Geophys. Res.*, **113**, D18206, https://doi.org/10.1029/2008JD009936.

——, J. M. Sabburg, J. Calbò, and D. Pagès, 2006: Retrieving cloud characteristics from ground-based daytime color all-sky images. *J. Atmos. Oceanic Technol.*, **23**, 633–652, https://doi.org/10.1175/JTECH1875.1.

——, J. H. Mather, and T. P. Ackerman, 2016: The ARM Tropical Western Pacific (TWP) sites. *The Atmospheric Radiation Measurement (ARM) Program: The First 20 Years, Meteor. Monogr.*, No. 57, Amer. Meteor. Soc., https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0024.1.

Mace, G. G., and F. J. Wrenn, 2013: Evaluation of the hydrometeor layers in the east and west Pacific within ISCCP cloud-top pressure-optical depth bins using merged *CloudSat* and *CALIPSO* data. *J. Climate*, **26**, 9429–9444, https://doi.org/10.1175/JCLI-D-12-00207.1.

Martins, F. R., M. P. Souza, and E. B. Pereira, 2003: Comparative study of satellite and ground techniques for cloud cover determination. *Adv. Space Res.*, **32**, 2275–2280, https://doi.org/10.1016/S0273-1177(03)90554-0.

Mather, J. H., and J. W. Voyles, 2013: The ARM Climate Research Facility: A review of structure and capabilities. *Bull. Amer. Meteor. Soc.*, **94**, 377–392, https://doi.org/10.1175/BAMS-D-11-00218.1.

McDonald, A. J., and S. Parsons, 2018: A comparison of cloud classification methodologies: Differences between cloud and dynamical regimes. *J. Geophys. Res. Atmos.*, **123**, 11 173–11 193, https://doi.org/10.1029/2018JD028595.

——, J. J. Cassano, B. Jolly, S. Parsons, and A. Schuddeboom, 2016: An automated satellite cloud classification scheme using self-organizing maps: Alternative ISCCP weather states. *J. Geophys. Res. Atmos.*, **121**, 13 009–13 030, https://doi.org/10.1002/2016JD025199.

McGovern, A., R. Lagerquist, D. J. Gagne II, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Amer. Meteor. Soc.*, **100**, 2175–2199, https://doi.org/10.1175/BAMS-D-18-0195.1.

Moran, K. P., B. E. Martner, M. J. Post, R. A. Kropfil, D. C. Welsh, and K. B. Widener, 1998: An unattended cloud-profiling radar for use in climate research. *Bull. Amer. Meteor. Soc.*, **79**, 443–455, https://doi.org/10.1175/1520-0477(1998)079<0443:AUCPRF>2.0.CO;2.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

Peixoto, J. P., and A. H. Oort, 1992: *Physics of Climate*. American Institute of Physics, 520 pp.

Pincus, R., S. Platnick, S. A. Ackerman, R. S. Hemler, and R. J. P. Hofmann, 2012: Reconciling simulated and observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators. *J. Climate*, **25**, 4699–4720, https://doi.org/10.1175/JCLI-D-11-00267.1.

Rossow, W. B., and R. A. Schiffer, 1991: ISCCP cloud data products. *Bull. Amer. Meteor. Soc.*, **72**, 2–20, https://doi.org/10.1175/1520-0477(1991)072<0002:ICDP>2.0.CO;2.

——, and ——, 1999: Advances in understanding clouds from ISCCP. *Bull. Amer. Meteor. Soc.*, **80**, 2261–2287, https://doi.org/10.1175/1520-0477(1999)080<2261:AIUCFI>2.0.CO;2.

Sisterson, D. L., R. A. Peppler, T. S. Cress, P. J. Lamb, and D. D. Turner, 2016: The ARM Southern Great Plains (SGP) site. *The Atmospheric Radiation Measurement (ARM) Program: The First 20 Years, Meteor. Monogr.*, No. 57, Amer. Meteor. Soc., https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0004.1.

Trenberth, K. E., J. T. Fasullo, and J. Kiehl, 2009: Earth's global energy budget. *Bull. Amer. Meteor. Soc.*, **90**, 311–324, https://doi.org/10.1175/2008BAMS2634.1.

Tselioudis, G., and P. Kollias, 2007: Evaluation of ECMWF cloud type simulations at the ARM Southern Great Plains site using a new cloud type climatology. *Geophys. Res. Lett.*, **34**, L03803, https://doi.org/10.1029/2006GL027314.

——, W. Rossow, Y. Zhang, and D. Konsta, 2013: Global weather states and their properties from passive and active satellite cloud retrievals. *J. Climate*, **26**, 7734–7746, https://doi.org/10.1175/JCLI-D-13-00024.1.

Turner, D. D., and R. G. Ellingson, 2016: Introduction. *The Atmospheric Radiation Measurement Program: The First 20 Years, Meteor. Monogr.*, No. 57. Amer. Meteor. Soc., v–x, https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0001.1.

Van Weverberg, K., and Coauthors, 2018: CAUSES: Attribution of surface radiation biases in NWP and climate models near the U.S. southern Great Plains. *J. Geophys. Res. Atmos.*, **123**, 3612–3644, https://doi.org/10.1002/2017JD027188.

Wang, Y., C. Shi, C. Wang, and B. Xiao, 2018: Ground-based cloud classification by learning stable local binary patterns. *Atmos. Res.*, **207**, 74–89, https://doi.org/10.1016/j.atmosres.2018.02.023.

Zhuo, W., Z. Cao, and Y. Xiao, 2014: Cloud classification of ground-based images using texture-structure features. *J. Atmos. Oceanic Technol.*, **31**, 79–92, https://doi.org/10.1175/JTECH-D-13-00048.1.