



# A cloud screening algorithm for ground-based sun photometry using all-sky images and deep transfer learning

Eric A. Wendt<sup>1</sup>, Bonne Ford<sup>2</sup>, and John Volckens<sup>1</sup>

<sup>1</sup>Department of Mechanical Engineering, Colorado State University, Fort Collins, Colorado, USA, 80523

<sup>2</sup>Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado, USA, 80523

Correspondence to: Eric A. Wendt (eawendtjr@gmail.com)

## Abstract.

Aerosol optical depth (AOD) is used to characterize aerosol loadings within Earth's atmosphere. Sun photometers measure AOD from the Earth's surface based on direct-sunlight intensity readings by spectrally narrow light detectors. However, when the solar disk is partially obscured by cloud cover, sun photometer measurements can be biased due to the interaction of sunlight with cloud constituents. We present a novel deep transfer learning model on all-sky images to support more accurate AOD retrievals. We used three independent image datasets for training and testing: the novel Northern Colorado All-Sky Image (NCASI), the Whole Sky Image SEGmentation (WSISEG), and the METCRAX-II datasets from the National Center for Atmospheric Research (NCAR). We visually partitioned all-sky images into three categories: 1) clear sky around the solar disk, 2) thin cirrus obstructing the solar disk, and 3) thick, non-cirrus clouds obstructing the solar disk. Two-thirds of the images were allocated for training and one-third were allocated for testing. We trained models based on all possible combinations of the training sets. The best-performing model successfully classified 95.5%, 96.9%, and 89.1% of testing images from NCASI, METCRAX-II and WSISEG datasets, respectively. Our results demonstrate that all-sky imaging with deep transfer learning can be applied toward cloud screening, which would aid ground-based AOD measurements.

## 1 Introduction

The abundance of aerosols in the atmosphere can be quantified optically from surface-based instruments called sun photometers, which measure aerosol optical depth (AOD), a dimensionless metric of light extinction by particles. Accurate AOD measurement requires a clear view of the solar disk. If the solar disk is partially obscured by clouds, measured AOD will be biased high. Thus, reliable AOD measurements from sun photometers require robust cloud screening.

Prior work has implemented quality control algorithms to reduce errors from cloud cover. The Aerosol Robotics Network (AERONET) provides automatic, multi-wavelength AOD measurements at hundreds of locations (Holben et al., 1998). Smirnov et al. (2000) leveraged the functionality of AERONET sun photometers to develop an automated cloud screening protocol for AOD measurements. Each AOD measurement is the average of a triplet of measurements, with 30 seconds between each measurement (Smirnov et al., 2000). The triplet is classified as cloud-



contaminated if the maximum difference within the triplet exceeds an empirically derived threshold, based on the assumption that AOD variability within short time periods is more likely due to clouds than rapid changes in aerosol (Smirnov et al., 2000). Additional screening steps incorporated all AOD measurements across a day (Smirnov et al., 2000).

40 Alexandrov et al. (2004) proposed an algorithm based on the change in AOD as a function of the Azimuthal position of the sun for Multi-Filter Rotating Shadowband Radiometers (MFRSRs). The algorithm had a 3.4% false negative rate, wherein the data point was identified as clear when actually cloud-contaminated, and a 4.0% false positive rate (Alexandrov, 2004). This approach is less conservative than the Smirnov et al., (2000) algorithm. (i.e. it will classify fewer points as cloud-contaminated) (Alexandrov, 2004). A subsequent study (Giles  
45 et al., 2019), used LIDAR to detect cirrus clouds and derive empirical thresholds, improving the removal of cirrus contamination by the AERONET algorithm.

A limitation of sensor-based cloud screening algorithms is their instrument-specific nature (Smirnov et al., 2000). Algorithms and empirical thresholds for AERONET instruments are likely not applicable to different instruments, which have different hardware (Wendt et al., 2021). When porting algorithms, unique empirical  
50 thresholds must be determined to remain effective, assuming the necessary sensors are present at all (Wendt et al., 2021). Previous studies using handheld sun photometers relied on operator observations of the sky to assess clouds (Boersma and de Vroom, 2006; Wendt et al., 2019; Smirnov et al., 2009; Ford et al., 2019). In Smirnov et al., (2009) operators were instructed to take a series of five to six consecutive measurements using a manually-operated device, adding an additional layer of quality control. However, the decision to initiate a measurement was still up to  
55 operator judgment based on visual assessment (Smirnov et al., 2009).

To date, image-based analysis of clouds has not been used in support of ground-based AOD measurement quality control. However, prior work in the area of cloud classification of sky images can be leveraged toward this aim. Long et al. (2006) used whole-sky images (collected using fish-eye lenses to achieve at least 160 degrees field of view) to evaluate cloud brokenness, uniformity, and solar obstruction (Long et al., 2006). Their algorithm, which  
60 separated images into their red, green, and blue (RGB) components, and classified pixels based on their the ratio of red to blue pixel values (R/B), performed well under uniform sky conditions (>95% accuracy for solar disk obstruction), but was less accurate for images with more irregular cloud coverage (<85% accuracy for solar disk obstruction) (Long et al., 2006).

Calbo & Sabburg (2008) mathematically defined sky conditions based on whole sky images, determined  
65 using six image features: mean, standard deviation, smoothness, third moment, uniformity, and entropy determined using the R/B values and an intensity parameter. The Calbo & Sabburg (2008) algorithm was least effective at discriminating between cases with thin clouds present and covering or not covering the sun (Calbó and Sabburg, 2008). For cloud classification, this distinction may not be significant; however, for AOD cloud screening applications, the presence of thin cirrus clouds can substantially bias a measurement (Smirnov et al., 2000;  
70 Alexandrov, 2004; Giles et al., 2019); thus, determining if thin clouds are present (and covering the sun) is critical.

Other cloud classification schemes aim to better distinguish cirrus from clear sky. Heinle et al. (2010) use R – B rather than R/B along with additional image features (difference, energy, contrast, and homogeneity); and a non-



parametric k-Nearest-Neighbors classifier. Li et al. (2011) found better results for cirrus clouds when they first  
classified sky cases as unimodal (only clouds or only sky) or bimodal (mixture of cloud and sky) before applying  
75 their algorithm. Liu et al. (2013) proposed the use of multiple images taken over the course of several minutes to  
form a Tensor Ensemble of images (Shuang Liu et al., 2013), an approach which leverages the dynamic nature of  
cloud movement to help distinguish it from forward scattering around the solar disk.

Recent research has applied machine learning techniques to evaluate sky condition and cloud coverage  
from images (Gu et al., 2018). Taravat et al. (2015) applied both a multilayer perceptron neural network and a  
80 support vector machine classification to whole-sky image classification, which both showed improved classification  
accuracy relative to previous thresholding approaches. Xia et al. (2015) proposed the use of a hybrid method using  
an extreme learning machine and k-nearest neighbors (kNN). Prior to classification, textural, color, and shape  
features were extracted. The classification accuracy increased with increasing numbers of features (Xia et al., 2015).  
Deep convolutional neural network (CNN) models are particularly effective for feature extraction and classification  
85 on image data (Gu et al., 2018). Shi et al. (2017) presented a CNN model for cloud identification based on common  
image classification architectures and demonstrated that their model outperformed prior feature-based models on the  
same dataset. Liu et al. (2018) incorporated temperature, humidity, pressure, wind speed, and maximum wind speed  
parameters from a local weather station with visual image data to further refine cloud type classification on whole  
sky images. Zhang et al. (2018) presented CloudNet, a deep CNN model tailored to extract cloud features and  
90 classify images based on cloud type, achieving better accuracy on both all-sky and partial-sky images compared  
with prior approaches. The success of prior work in image-based cloud identification supports the use of similar  
algorithms for solar obstruction screening for AOD quality control.

In previous machine learning cloud classification models, training and testing data were drawn from the  
same domain, namely sky images (Shi et al., 2017; Liu et al., 2018; Zhang et al., 2018). Recent research into deep  
95 learning has explored transfer learning, in which data from outside the application domain are incorporated into  
model development (Pan and Yang, 2010; Zhuang et al., 2021 and references therein). Transfer learning for image  
classification leverages the most effective pre-trained CNN models to compute features useful for building  
classifiers for a wide variety of image classes and then applies the models to specific domains (Zhuang et al., 2021).  
Effective deep CNN models can have over 10 million trainable parameters (Simonyan & Zisserman, 2015);  
100 however, in transfer learning, these parameters are trained in advance, outside of the application domain (Pan &  
Yang, 2010; Zhuang et al., 2021), thus reducing the time and resource requirements. In practice, pre-trained  
parameters may be frozen or left trainable. If frozen, application-specific training is reduced to the classification  
layer parameters, which are a small fraction of the total parameters. If left trainable, application-specific training is  
accelerated as parameters will likely be initialized closer to their optimal values.

The Visual Geometry Group (VGG) at Oxford University developed the VGG-16 model, a deep CNN  
105 model designed for image classification (Simonyan and Zisserman, 2015). VGG-16 was trained on ImageNet, an  
image database consisting of over 14 million images from 20,000+ unique image classes (Deng et al., 2009;  
Simonyan and Zisserman, 2015). VGG-16 has been used effectively for transfer learning in image classification  
applications (e.g. Tammina, 2019; Guan et al., 2019; Kaur and Gandhi, 2019).



110 Here, we develop a transfer learning model based on VGG-16 for cloud screening on all sky images. We  
first present a new all-sky-image data set collected using a low-cost prototype imaging module. We then describe  
the image pre-processing algorithm used to prepare images for input into our classification model and the design and  
training of our classification model. Finally, we evaluate the performance of our algorithm on three independent all-  
sky-image sets.

## 115 **2 Materials and Methods**

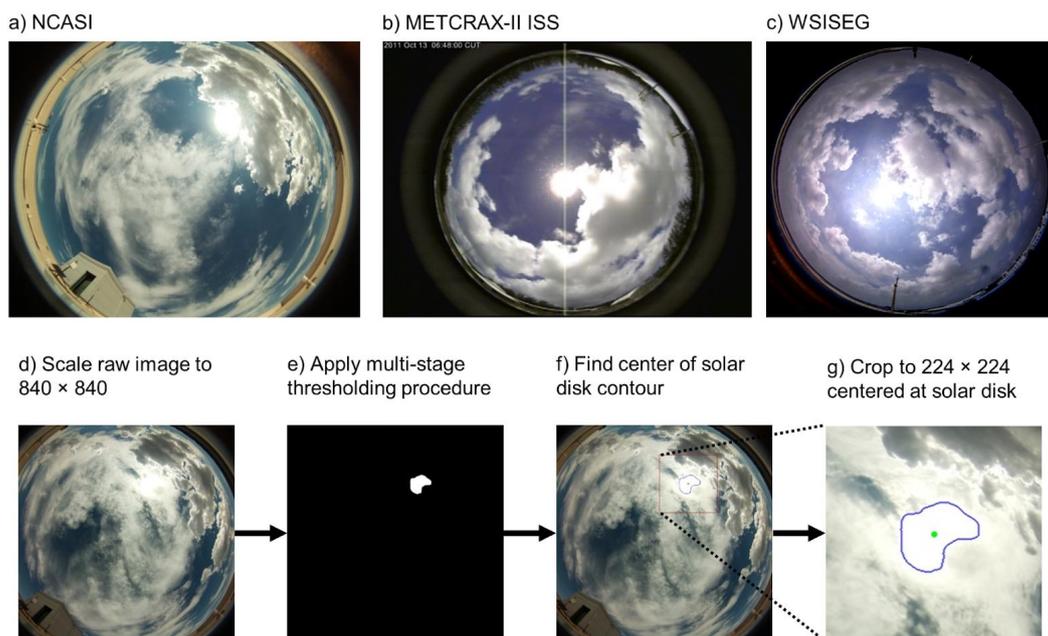
### **2.1 All-Sky Images**

Image-based cloud screening for AOD measurements requires the co-location or hardware integration of an  
all-sky camera with a sun photometer. All-sky cameras suitable for AOD cloud screening must image the solar disk  
and the sky surrounding the solar disk while preserving the edge detail of the solar disk and nearby clouds. Images  
120 cannot be used for AOD screening if the image area surrounding the solar disk is saturated with sunlight. For this  
reason, all-sky cameras used in previous cloud identification studies were designed to block or attenuate incident  
sunlight (e.g Calbó & Sabburg, 2008; Fa et al., 2019; Xie et al., 2020).

Accordingly, we located two pre-existing sets of all-sky image data with which to train and test our model,  
in addition to a third data set that was collected specifically for this project (see the top row of Fig. 1 for example  
125 images). The Whole Sky Image SEGmentation (WSISEG) data set includes 400 all-sky images captured on the 25  
rooftop of Anhui Air Traffic Management Bureau, Civil Aviation Administration of China in July 2018 (Xie et al.,  
2020). The resolution of images in WSISEG is 2,000 pixels  $\times$  1,944 pixels. The National Center for Atmospheric  
Research (NCAR) published METCRAX-II ISS All Sky Camera Imagery (UCAR/NCAR - Earth Observing  
Laboratory, 2016). The images in METCRAX-II were collected on the island of Diego Garcia between September  
130 2011 and February 2012. The resolution of these images is 640 pixels  $\times$  480 pixels. Due to the large data storage  
requirements (6,128 MB), we did not save the entire data set. We manually selected all day-time images that were  
later evaluated for inclusion into training and testing data.

Finally, we created a collection of all sky images called the Northern Colorado All-Sky Image (NCASI)  
set, using a custom-designed imaging module (see Text S1, Fig. S1, and Table S1 in the supporting information for  
135 details). We collected a total of 3,544 images between 1-21 September 2021. Images from the 1st, 20th, and 21st of  
September 2021 were collected near a private residence in Boulder, Colorado, USA. Images from all other days  
were collected at the Powerhouse Energy Campus at Colorado State University (430 N. College Avenue, Fort  
Collins, Colorado, USA). The image resolution is 1,920 pixels  $\times$  1,200 pixels. The module was configured such that  
when started, it captured an all-sky image every 30 seconds.

140



**Figure 1: Top: example images from the a) NCASI, b) METCRAX-II ISS, and c) WSISEG datasets. Bottom: Transformations performed in image preparation algorithm. d) Raw image files are scaled and cropped to a uniform size 840 pixels  $\times$  840 pixels). e) A multi-stage thresholding algorithm isolates sunlit pixels. f) From the sunlit contour, we calculate the center of the sunlit pixels in the image. g) Using the center, we crop the image (224 pixels  $\times$  224 pixels).**

145

150

155

We built our algorithm to classify three types of sky conditions: 1) sun not obscured by clouds; 2) sun obscured by thin cirrus clouds; and 3) sun obscured by thick, non-cirrus clouds. We manually labeled images according to these designations. We discriminated between cirrus and thick, non-cirrus clouds based on how the cloud cover impacted the shape of the solar disk. Images where the solar disk presented as circular, despite the presence of thin cloud cover, were designated as cirrus. Images where the shape of the solar disk deviated from circular due to the presence of cloud cover were designated as non-cirrus. We built and selected our model training and testing datasets to include samples from all three enumerated sky conditions. We provide example images from each data set under each sky condition in Fig. S2.

## 2.2 Image Preparation

Initially, input images of potentially varying sizes are scaled to a common size of 840 pixels  $\times$  840 pixels using an area pixel model as illustrated in Fig. 1d (Chun-Ho Kim et al., 2003). This operation normalizes important features such as the size of the solar disk in square pixels. From the scaled image, we isolate the blue color channel



160 and perform a binary threshold operation, where 8-bit pixel values greater than or equal to 252 are set to the  
maximum value of 255 (white) and all other pixels are set to zero (black). We then apply smoothing to the  
thresholded image using  $15 \times 15$  Gaussian kernel with the standard deviations in the horizontal and vertical axes set  
to zero (Burt, 1981). We found that for images with sunlit clouds near, but not obstructing, the solar disk, a Gaussian  
filter alone was insufficient to smooth the edges of sunlit contours. Therefore, we applied an additional bilateral  
165 filter with a pixel-neighborhood diameter of 15 pixels with color and space standard deviation values set to 25  
(Tomasi and Manduchi, 1998). The smoothing steps blend black pixels with white pixels, leaving pixels near the  
contour edge with values between 0 and 255. To restore the image to a binary image, we apply an additional binary  
thresholding operation with the threshold set to a pixel value of 50. The result of these thresholding operations is  
illustrated in Fig. 1e. We then apply a contour detection algorithm to the binary image to derive contour arrays for  
170 each area of contiguous white pixels (Suzuki and be, 1985). For images with a high number of optically saturated  
clouds (common in the METCRAX-II data set), there can be multiple sunlit contours that are not the solar disk. To  
isolate the solar disk contours from surrounding sunlit clouds, we apply pixel area and circularity criteria. Across all  
three datasets, the solar disk for scaled images was greater than 2750 square pixels. Contours greater than that  
threshold are evaluated based on their circularity, defined as follows:

$$175 \quad C = \frac{4\pi A}{L^2} \quad (1)$$

where  $C$  is the circularity,  $A$  is the area, and  $L$  is the arc length of the contour. Among contours within the  
acceptable area range, the contour with the highest circularity is considered the contour of the solar disk. We then  
calculate the centroid of the solar disk contour and crop the image to a region sized 224 pixels  $\times$  224 pixels centered  
at the centroid of the solar disk, as depicted in Fig. 1f and Fig. 1g. For images where the solar disk is fully obscured  
180 and no pixels pass the binary thresholding tests (most often due to heavy cloud cover obscuring the solar disk), the  
center of the cropped image is placed at the center of the original scaled image and the resulting image is given the  
“cloud” label. In other cases, as in Fig. 1, the solar disk is obscured by clouds such that its true center cannot be  
determined. For these images, the approximate center is used and the image is given the cloud label. For images  
where the calculated solar disk center is within 112 pixels of an edge, the image cannot be cropped to 224 pixels  $\times$   
185 224 pixels. In these cases, the edge of the scaled image (Fig. 1d) is used as the edge of the cropped image, and the  
cropped image will be smaller than 224 pixels  $\times$  224 pixels. We automated these pre-processing steps using  
functions from the OpenCV library.

### 2.3 Model Design

For our feature extraction layers, we used the VGG-16 deep CNN model with parameters pre-trained on  
190 ImageNet (Deng et al., 2009; Simonyan and Zisserman, 2015). We used the Tensorflow implementation of VGG-16  
(Abadi et al., 2015). VGG-16 expects input tensors sized  $224 \times 224 \times 3$ , with the third dimension representing  
RGB color channels present in colored images (Simonyan and Zisserman, 2015). After image preparation, most  
images were suitable for input into the VGG-16 model without further resizing. Images that were cropped to smaller  
proportions (ie., center of the solar disk was close to the image edge) were scaled to the proper input size and then



195 passed to a data generator implemented in Tensorflow’s VGG-16 model (Abadi et al., 2015). The output of the pre-  
trained VGG-16 model is a  $7 \times 7 \times 512$  tensor representing the features learned from the ImageNet database (Deng  
et al., 2009; Simonyan and Zisserman, 2015). To interface with VGG-16, the output tensor from the feature  
extraction layers is flattened to a one-dimensional vector, which is interfaced with a three-node dense classification  
layer with a softmax activation function (Bridle, 1989). For a particular input sample, the output layer gives a  
200 probability estimate for each of the three possible sky condition classes. The class associated with the highest  
probability value is the classification of the model.

## 2.4 Model Training and Evaluation

We created training and testing subdatasets for NCASI, METCRAX-II, and WSISEG using a random split  
of approximately two-thirds training and one-third testing data. Training-testing data partitions for each set and class  
205 designation are provided in Table 1. The images in the training and testing sets were pre-processed using our image  
preparation algorithm. We built seven datasets, each with training and testing subsets: one for each data set, one for  
each of the possible combinations, and a single combination of all three.

**Table 1: Training and testing data partitions by data set and class designation.**

Image data set	Class designation	Number of training samples	Number of testing samples
NCASI	Clear	59	27
NCASI	Cirrus	70	24
NCASI	Cloud	52	38
METCRAX-II	Clear	204	118
METCRAX-II	Cirrus	124	73
METCRAX-II	Cloud	239	133
WSISEG	Clear	35	18
WSISEG	Cirrus	26	22
WSISEG	Cloud	198	98

210 We trained all models using the same training parameters. We used the categorical cross-entropy loss  
function and the Adam optimizer with a learning rate of 0.00012 (Kingma and Ba, 2017). We trained for 100 epochs  
with a batch size of 32. We did not modify pre-trained weights of VGG-16. The weight and bias parameters of the  
output layer were the only trainable parameters, which comprised 0.51% of the total model parameters. To limit  
215 overfitting, we also applied data augmentation, which supplements training data by producing batches of randomly  
modified images created via transformation operations. We implemented data augmentation in Tensorflow using the  
ImageDataGenerator module (Abadi et al., 2015) allowing rotation of 20 degrees, width/height shifting of 10%, a



zooming range of 20%, and random horizontal and vertical flips. We trained our model using a GPU (NVIDIA,  
 Tesla K80, Santa Clara, California, USA) on the Google Colaboratory platform (Google, Mountain View,  
 220 California, USA).

### 3 Results and Discussion

#### 3.1 Model Evaluation

Holding model and training parameters constant, we trained seven different models using the seven  
 possible combinations of the three training datasets. We evaluated each of the seven models according to  
 225 classification accuracy on the three testing datasets. Separating cirrus and non-cirrus cloud classifications is useful  
 for interpreting classification results, though it is not strictly necessary for sun photometer quality control. We  
 assessed the model's performance on the binary classification problem typically addressed by sun photometer  
 quality control algorithms by combining cirrus and non-cirrus cloud designations. The results of our analysis are  
 presented in Table 2.

230

**Table 2. Three-class (two-class) classification accuracy of models trained on seven different training datasets. Three-class results are from cirrus, clear, and cloud categories. Two-class results are from clear and cloud categories. Accuracy metrics for each model were calculated using the testing data from NCASI, METCRAX-II, and WSISEG individually. The best performing model on each test data set is given in bold text.**

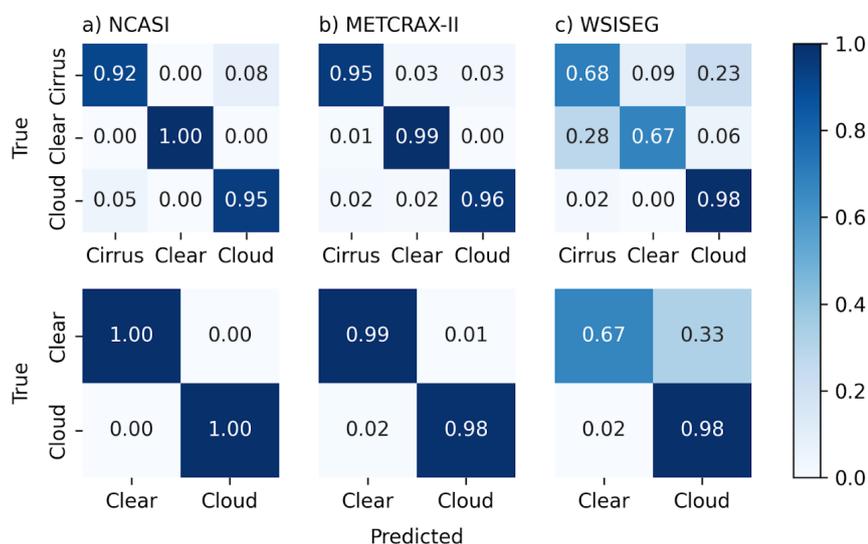
Training data set(s)	Model Number	Accuracy on NCASI (%)	Accuracy on METCRAX-II (%)	Accuracy on WSISEG (%)
NCASI	1	<b>97.8 (100.0)</b>	76.2 (81.2)	75.4 (87.0)
METCRAX-II	2	84.3 (84.2)	95.7 (95.7)	79.0 (79.0)
WSISEG	3	64.0 (78.7)	62.0 (68.5)	89.1 (94.2)
NCASI and METCRAX-II	4	95.4 ( <b>100.0</b> )	95.5 (97.5)	82.6 (87.0)
NCASI and WSISEG	5	92.1 ( <b>100.0</b> )	70.0 (72.5)	<b>90.0 (95.7)</b>
METCRAX-II and WSISEG	6	84.3 (96.6)	95.7 (97.8)	89.1 (94.9)
All	7	95.5 ( <b>100.0</b> )	<b>96.9 (98.4)</b>	89.1 (94.2)

235

The model trained on all training data (model 7) generalized the best to the testing data. Model 7 correctly classified 95.5%, 96.9% and 89.1% of testing samples from NCASI, METCRAX-II, and WSISEG images, respectively (Table 2). The three-class confusion matrix for model 7 is given in the top row of Fig. 2 (three-class confusion matrices for the remaining six models are provided in the top rows of Figures S4-S9). The accuracies of



240 model 7 for the binary classification were 100.0%, 98.4%, and 94.2% for NCASI, METCRAX-II and WSISEG, respectively (Table 2). We present the binary confusion matrix for model 7 in the bottom row of Fig. 2 (see bottom rows of Figures S4-S9 for the remaining six models).



245 **Figure 2. Top row: three-class (cirrus, clear, and cloud) confusion matrices for transfer learning model trained on NCASI, METCRAX-II, and WSISEG training datasets (model 7). Bottom row: Two-class (clear and cloud) confusion matrices for transfer learning model trained on NCASI, METCRAX-II, and WSISEG training datasets (model 7). a) Results on NCASI testing data set. b) Results on METCRAX-II testing data set. c) Results on WSISEG testing data set. “Predicted” refers to the model output and “True” refers to the observed class designation.**

250 For both the three-class and binary classification problems, model 7 performed best on the NCASI images, followed by METCRAX-II, and WSISEG, respectively (Fig. 2). Three of the four models at least partially trained on the NCASI training set achieved over 95.0% accuracy on the NCASI testing set (Table 2). On both the NCASI and METCRAX-II testing images, model 7 exhibited the highest error rate on images manually labeled as cirrus. On the NCASI data, 8.0% of cirrus images were classified as cloud (not cirrus) images and 0.0% were classified as clear images. For METCRAX-II, 3.0% of the cirrus images were classified as cloud (not cirrus) images, and 3.0% were classified as clear images. However, in the two-class problem, both cirrus and non-cirrus clouds are classified identically, rendering errors between cirrus and cloud categories immaterial for the inherently binary problem of AOD cloud screening. When the two cloud types are combined, model 7 correctly classifies 100.0% of samples from the NCASI set. For METCRAX-II data, model 7 incorrectly classifies 1.0% of clear images as cloud, and incorrectly classifies 2.0% of clear images as cloud. For WSISEG data, high rates of confusion between clear and cirrus images led to relatively poor results in the two-class problem, with 33.0% of clear images being classified as cloud images, but only 2.0% of cloud images being classified as clear.

255

260



Model 7 performed well relative to prior AOD screening algorithms, but these results partially depended on the testing dataset. The algorithm proposed in Alexandrov et al. (2004) had a false negative (i.e. cloud classified as clear) rate of 3.4% and a false positive (i.e. clear classified as cloud) rate of 4.0% for the binary cloud screening  
265 problem. Our model performed better than the prior algorithm on NCASI and METCRAX-II images, and worse on WSISEG images (Fig. 2). The false negative rates on NCASI, METCRAX-II, and WSISEG testing data were 0.0%, 2.0%, and 2.0%, respectively (Fig. 2). The false positive rates on NCASI, METCRAX-II, and WSISEG testing data were 0.0%, 1.0%, and 33.0%, respectively (Fig. 2). In Alexandrov et al. (2004), the authors do not specify the  
270 relative proportions of cirrus and non-cirrus cloud samples present in the 575 cloud-contaminated samples they analyzed, precluding a direct comparison of performance on cirrus cases. The generalizability of our models to images similar to those in NCASI is supported by the performance of model 6 (trained only on METCRAX-II and WSISEG) on NCASI. Despite not seeing any NCASI images during training, model 6 correctly classified 96.6% of NCASI testing images in the two-class problem (Table S3).

275 Model 7 performed well on NCASI images because, in part, thin and thick clouds not directly in front of the sun were less likely to be saturated (i.e. maximum RGB values) with light (e.g. Fig. 1a). In the METCRAX-II and WSISEG datasets, clouds outside of direct sunlight, yet still in the 224 pixel  $\times$  224 pixel frame, were more likely to be fully saturated with light (e.g. Figs. 1b and 1c). In the WSISEG data set, there were apparent camera glare spots around the solar disk (Figs. S10-S12 shows all images misclassified by model 7). We suspect  
280 these glare spots were a result of sunlight reflection off of plastic or glass protective coverings over the imaging systems. The METCRAX-II data set has less severe glare spots, and the NCASI data set lacks glare spots entirely.

### 3.2 Limitations

A limitation of this work was the subjectivity of the class label designations, particularly for the METCRAX-II and WSISEG datasets (Figs. S10-S12). For the NCASI data that was collected specifically for this  
285 project, we had the advantage of visually observing the actual condition of the sky as the images were collected. However, with the METCRAX-II and WSISEG datasets, we could only assign class designations using images. This complicated class designations, particularly between cirrus and clear images in the WSISEG data set, where it was difficult to distinguish between glare and cirrus clouds (e.g. Fig. S12). Glare in WSISEG images may have contributed to relatively poor results distinguishing between clear and cirrus designations for all models (Fig. 2 and  
290 Figs. S4-S9). However, issues with the WSISEG data did not impact the performance of models 5-7, which partially used WSISEG data. Models 5-7 generalized well to NCASI and METCRAX-II despite potential mislabeling of WSISEG samples. Clarity issues in WSISEG images emphasize the importance of high image fidelity around the solar disk for AOD screening application. In future work, we will test the generalization ability of model 7 on additional independent data without additional training.

295 Our model generally performed well on a variety of different cloud types from three independent datasets. However, we did not test our model on images with high aerosol loading. Wildfire smoke, volcanic ash, heavy industrial emissions and other sources of high atmospheric aerosol concentrations could be erroneously classified as



clouds. Additional images with the sun obscured by high levels of aerosol are needed to test the model's sensitivity to aerosol concentrations.

300 Our datasets were also limited to images from three unique camera configurations. Different imaging systems produce images with different hues, resolutions, and saturation levels, among other visual differences. As demonstrated on the NCASI and METCRAX-II datasets (Tables S2 and S3), our model performs best on images with limited lens glare spots. To improve performance on images from different imaging systems, we recommend further training of our output layer parameters using additional labeled images. Models 1-7 models and their  
305 respective weights are publicly available (see data availability section).

#### 4 Conclusions

In this work, we presented a novel approach for cloud screening that could be applied to AOD measurements and that builds on the literature surrounding cloud classification from whole sky images. We designed an imaging module to capture all-sky images with high-fidelity, particularly around the solar disk. Using  
310 this imaging module, we produced the NCASI data set, a novel collection of all-sky images from Northern Colorado that includes images of a variety of cloud and sky conditions. Combining our NCASI data set with two other independent all-sky image datasets and the pre-trained VGG-16 model, we applied transfer learning to develop an effective cloud screening model. Our model performed well classifying the solar disk as unobstructed (clear), obstructed by thin cirrus clouds, or obstructed by optically thick, non-cirrus clouds.

315 Image-based cloud screening is agnostic of sun photometer hardware, making it especially applicable for lower-cost sun photometers (Wendt et al., 2019, 2021), which may lack the hardware and personnel required to fully implement state-of-the-art sensor-based cloud screening. Leveraging modern imaging technology, all-sky imagers can be incorporated with AOD measurement systems at relatively low cost (Table S1). When paired with AOD measurements, all-sky images may be used for relatively simple quality control and descriptive purposes. Our work  
320 suggests that cameras configured to produce images like the NCASI datasets will perform well in quality control applications. The performance of our model on cirrus cloud cover represents a promising advancement in AOD quality control for cloud cover most difficult to distinguish from elevated aerosol.

#### *Data availability.*

Original WSISEG images are available for download in the following GitHub repository:  
325 <https://github.com/CV-Application/WSISEG-Database>. Original METCRAX-II images are available for download on the UCAR Earth Observing Laboratory (EOL) website (<https://data.eol.ucar.edu/dataset/386.016>). The complete set of raw NCASI images, selected and partitioned images from all three datasets, trained deep learning models, and example code related to this work are available in the following GitHub repository:  
<https://github.com/eawendt/aodqc>.



330 *Supplement.*

The supplement related to this article is available online at:

*Author contribution.*

EAW designed the imaging module and collected the NCASI dataset. EAW designed the pre-processing algorithm and the transfer learning model. EAW, BF, and JV evaluated model performance. EAW and BF acquired  
335 independent validation datasets. EAW, BF, and JV prepared the manuscript.

*Competing Interests.*

The authors declare that they have no conflict of interest.

*Acknowledgment.*

The authors wish to thank Casey Quinn, Christian L'Orange, Jeffrey R. Peirce, Shantanu Jathar, Sudeep Pasricha,  
340 Ben Strauss, and Todd Hochwitz (Zebulon Solutions LLC, Longmont, CO, USA) for their contributions to this work.

*Financial support.*

This research was supported by NASA grant 80NSSC18M0120.

**References**

- 345 Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: A system for large-scale machine learning, 21, 2015.
- Alexandrov, M. D.: Automated cloud screening algorithm for MFRSR data, *Geophys. Res. Lett.*, 31, L04118, <https://doi.org/10.1029/2003GL019105>, 2004.
- 350 Boersma, K. F. and de Vroom, J. P.: Validation of MODIS aerosol observations over the Netherlands with GLOBE student measurements, *J. Geophys. Res.*, 111, D20311, <https://doi.org/10.1029/2006JD007172>, 2006.
- Bridle, J. S.: Training Stochastic Model Recognition Algorithms as Networks can Lead to Maximum Mutual Information Estimation of Parameters, *Proc. 2nd Int. Conf. Neural Inf. Process. Syst.*, 211–217, 1989.
- Burt, P. J.: Fast Filter Transforms for Image Processing, *Comput. Graph. Image Process.*, 16, 20–51, [https://doi.org/10.1016/0146-664X\(81\)90092-7](https://doi.org/10.1016/0146-664X(81)90092-7), 1981.
- 355 Calbó, J. and Sabburg, J.: Feature Extraction from Whole-Sky Ground-Based Images for Cloud-Type Recognition, *J. Atmospheric Ocean. Technol.*, 25, 3–14, <https://doi.org/10.1175/2007JTECHA959.1>, 2008.
- Chun-Ho Kim, Si-Mun Seong, Jin-Aeon Lee, and Lee-Sup Kim: Winscale: an image-scaling algorithm using an area pixel model, *IEEE Trans. Circuits Syst. Video Technol.*, 13, 549–553,



- 360 <https://doi.org/10.1109/TCSVT.2003.813431>, 2003.  
Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, and Li Fei-Fei: ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops), Miami, FL, 248–255, <https://doi.org/10.1109/CVPR.2009.5206848>, 2009.
- 365 Fa, T., Xie, W., Wang, Y., and Xia, Y.: Development of an all-sky imaging system for cloud cover assessment, *Appl. Opt.*, 58, 5516, <https://doi.org/10.1364/AO.58.005516>, 2019.  
Ford, B., Pierce, J. R., Wendt, E., Long, M., Jathar, S., Mehaffy, J., Tryner, J., Quinn, C., van Zyl, L., L’Orange, C., Miller-Lionberg, D., and Volckens, J.: A low-cost monitor for measurement of fine particulate matter and aerosol optical depth – Part 2: Citizen-science pilot campaign in northern Colorado, *Atmos Meas Tech*, 15, 2019.
- 370 Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I.: Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements, *Atmospheric Meas. Tech.*, 12, 169–209, <https://doi.org/10.5194/amt-12-169-2019>, 2019.
- 375 Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., and Chen, T.: Recent advances in convolutional neural networks, *Pattern Recognit.*, 77, 354–377, <https://doi.org/10.1016/j.patcog.2017.10.013>, 2018.  
Guan, Q., Wang, Y., Ping, B., Li, D., Du, J., Qin, Y., Lu, H., Wan, X., and Xiang, J.: Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study, *J. Cancer*, 10, 4876–4882, <https://doi.org/10.7150/jca.28769>, 2019.
- 380 Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization, *Remote Sens. Environ.*, 66, 1–16, [https://doi.org/10.1016/S0034-4257\(98\)00031-5](https://doi.org/10.1016/S0034-4257(98)00031-5), 1998.
- 385 Kaur, T. and Gandhi, T. K.: Automated Brain Image Classification Based on VGG-16 and Transfer Learning, in: 2019 International Conference on Information Technology (ICIT), 2019 International Conference on Information Technology (ICIT), Bhubaneswar, India, 94–98, <https://doi.org/10.1109/ICIT48102.2019.00023>, 2019.  
Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, *ArXiv14126980 Cs*, 2017.
- 390 Liu, S., Li, M., Zhang, Z., Xiao, B., and Cao, X.: Multimodal Ground-Based Cloud Classification Using Joint Fusion Convolutional Neural Network, *Remote Sens.*, 10, 822, <https://doi.org/10.3390/rs10060822>, 2018.  
Long, C. N., Sabburg, J. M., Calbó, J., and Pagès, D.: Retrieving Cloud Characteristics from Ground-Based Daytime Color All-Sky Images, *J. Atmospheric Ocean. Technol.*, 23, 633–652, <https://doi.org/10.1175/JTECH1875.1>, 2006.  
Pan, S. J. and Yang, Q.: A Survey on Transfer Learning, *IEEE Trans. Knowl. Data Eng.*, 22, 1345–1359, <https://doi.org/10.1109/TKDE.2009.191>, 2010.
- 395 Shi, C., Wang, C., Wang, Y., and Xiao, B.: Deep Convolutional Activations-Based Features for Ground-Based Cloud Classification, *IEEE Geosci. Remote Sens. Lett.*, 14, 816–820, <https://doi.org/10.1109/LGRS.2017.2681658>,



- 2017.
- Shuang Liu, Chunheng Wang, Baihua Xiao, Zhong Zhang, and Xiaozhong Cao: Tensor Ensemble of Ground-Based Cloud Sequences: Its Modeling, Classification, and Synthesis, *IEEE Geosci. Remote Sens. Lett.*, 10, 1190–1194, <https://doi.org/10.1109/LGRS.2012.2236073>, 2013.
- 400 Simonyan, K. and Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, *ArXiv14091556 Cs*, 2015.
- Smirnov, A., Holben, B. N., Eck, T. F., Dubovik, O., and Slutsker, I.: Cloud-Screening and Quality Control Algorithms for the AERONET Database, *Remote Sens. Environ.*, 73, 337–349, [https://doi.org/10.1016/S0034-4257\(00\)00109-7](https://doi.org/10.1016/S0034-4257(00)00109-7), 2000.
- 405 Smirnov, A., Holben, B. N., Slutsker, I., Giles, D. M., McClain, C. R., Eck, T. F., Sakerin, S. M., Macke, A., Croot, P., Zibordi, G., Quinn, P. K., Sciare, J., Kinne, S., Harvey, M., Smyth, T. J., Piketh, S., Zielinski, T., Proshutinsky, A., Goes, J. I., Nelson, N. B., Larouche, P., Radionov, V. F., Goloub, P., Krishna Moorthy, K., Matarrese, R., Robertson, E. J., and Jourdin, F.: Maritime Aerosol Network as a component of Aerosol Robotic Network, *J. Geophys. Res.*, 114, D06204, <https://doi.org/10.1029/2008JD011257>, 2009.
- 410 Suzuki, S. and be, K.: Topological structural analysis of digitized binary images by border following, *Comput. Vis. Graph. Image Process.*, 30, 32–46, [https://doi.org/10.1016/0734-189X\(85\)90016-7](https://doi.org/10.1016/0734-189X(85)90016-7), 1985.
- Tamma, S.: Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images, *Int. J. Sci. Res. Publ. IJSRP*, 9, p9420, <https://doi.org/10.29322/IJSRP.9.10.2019.p9420>, 2019.
- 415 Tomasi, C. and Manduchi, R.: Bilateral filtering for gray and color images, in: Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), IEEE 6th International Conference on Computer Vision, Bombay, India, 839–846, <https://doi.org/10.1109/ICCV.1998.710815>, 1998.
- UCAR/NCAR - Earth Observing Laboratory: METCRAX-II ISS All Sky Camera Imagery. 1.0, 2016.
- Wendt, E. A., Quinn, C. W., Miller-Lionberg, D. D., Tryner, J., L'Orange, C., Ford, B., Yalin, A. P., Pierce, J. R., Jathar, S., and Volckens, J.: A low-cost monitor for simultaneous measurement of fine particulate matter and aerosol optical depth – Part 1: Specifications and testing, *Atmospheric Meas. Tech.*, 12, 5431–5441, <https://doi.org/10.5194/amt-12-5431-2019>, 2019.
- 420 Wendt, E. A., Quinn, C., L'Orange, C., Miller-Lionberg, D. D., Ford, B., Pierce, J. R., Mehaffy, J., Cheeseman, M., Jathar, S. H., Hagan, D. H., Rosen, Z., Long, M., and Volckens, J.: A low-cost monitor for simultaneous measurement of fine particulate matter and aerosol optical depth – Part 3: Automation and design improvements, *Atmospheric Meas. Tech.*, 14, 6023–6038, <https://doi.org/10.5194/amt-14-6023-2021>, 2021.
- 425 Xia, M., Lu, W., Yang, J., Ma, Y., Yao, W., and Zheng, Z.: A hybrid method based on extreme learning machine and k-nearest neighbor for cloud classification of ground-based visible cloud image, *Neurocomputing*, 160, 238–249, <https://doi.org/10.1016/j.neucom.2015.02.022>, 2015.
- 430 Xie, W., Liu, D., Yang, M., Chen, S., Wang, B., Wang, Z., Xia, Y., Liu, Y., Wang, Y., and Zhang, C.: SegCloud: a novel cloud image segmentation model using a deep convolutional neural network for ground-based all-sky-view camera observation, *Atmospheric Meas. Tech.*, 13, 1953–1961, <https://doi.org/10.5194/amt-13-1953-2020>, 2020.
- Zhang, J., Liu, P., Zhang, F., and Song, Q.: CloudNet: Ground-Based Cloud Classification With Deep Convolutional



Neural Network, *Geophys. Res. Lett.*, 45, 8665–8672, <https://doi.org/10.1029/2018GL077787>, 2018.

- 435 Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., and He, Q.: A Comprehensive Survey on Transfer Learning, *Proc. IEEE*, 109, 43–76, <https://doi.org/10.1109/JPROC.2020.3004555>, 2021.