We would like to thank the reviewers for their suggestions, edits and questions which contributed to a hopefully improved revised manuscript. In the following, please find our point by point response to both reviewers' comments.

Anonymous Referee #1
Received and published: 30 April 2020

Yuan et al. present a method to identify marine low-level cloud regimes. Using MODIS reflectances, and creating a training dataset by human visual inspection, they apply a Deep convolutional neural network to objectively assign each scene to one of six pre-defined types. The method is well described and carefully evaluated. The authors aim to make their product publicly available which is potentially of great usefulness to studies of clouds. The paper is very well written and of interest to the readership of Atmos. Meas. Tech. I only have a few minor remarks which the authors should consider in a revision.

l27 "shows"
Changed from 'show' to 'shows'.

l28 "suggests"
Added 's' after suggest.

l39 "histograms"; however aren't pixel-level retrievals and joint histograms redundant? the latter is just a way to statistically retain the pixel-level information at level 3 aggregation.
Correct. It depends on how the pixel-level retrievals are used. But ultimately, they both use the pixel level retrievals. Some methods use a one-dimensional PDF and others may use joint-histograms.

l41 only since then? Or not rather since ever / since the first cloud observations (such as Howard,
The reference we used is the best example we are aware of. The Howard manuscript is quite interesting. All co-authors had a discussion about it and we decided that our scope is more focused on the mesoscale morphology and not to include it here.

l62 "a plan" or "plans"
Changed to 'plans'.

l70 The Platnick reference should be updated (actual author list is longer, and it appeared 2017 (vol 55)).
Changed to Platnick et al., 2017.

l71 Please specify the horizontal resolution for reflectances and retrieval products.
Added a sentence. "The spatial resolution of these parameters is 1km at nadir."

l77 Provide the unit here. I assume it is 128 x 128 pixels of 1x1 km2 size each?

Changed. Since the pixel size changes slightly with view angle we now simply quote scene size in terms of pixels rather than physical area.

l84 It is a nice idea to include this a bit technical detail. This illustrated well what is actually done.

l87 And this is a good idea!
Thanks!

l94 Omit "keep the task manageable" once.
Changed.

l119 Are the PDFs exactly the retrievals from the scenes provided in Fig. 3? It would be good if it was such, and should be clarified in the text.
The PDFs are mean distributions of samples belonging to a particular type. We randomly selected 1000 scenes for each cloud type. We added explanations at line 130 and caption for Figure 4.

l141 I don't understand what "flipping" means if not rotating by 180◦ . The authors should clarify this.
Now clarified: "The flipping operation is achieved by mirroring the original image across a horizontal axis."

l154 It would be useful to explain in one sentence to the non-specialized readership what the confusion matrix is.
Added : "The confusion matrix summarizes the classification prediction results. For each cloud type, or row, it shows the percentage of correct predictions on the diagonal and percentages of incorrect predictions off the diagonal."

l161 It would be interesting to know how often this occurs for the different cloud types. E.g. a fraction of disagreement for each type?
With the added explanation about the confusion matrix, readers should be able to read how often misclassifications occur for each type and how they distribute across different types.

l165 This mostly looks quite reasonable. However, some results seem rather strange to the naked eye. E.g. where the solid stratus diagnosed at 14◦S/78◦W I don't see any cloud, let alone a stratus.
The dot represents the center location of a scene. The right half of this scene is indeed occupied by stratus clouds. A scene does not have to be mostly cloudy to be classified as stratus. The algorithm examines the textual information of the clouds. We added explanations at line 188.

l186 drop "the"

Changed.

l338 "indicates". And what is the difference between the light pink and red lines?
Changed. The solid lines are running means of the light pink lines.

l365 Help from which other authors?
Clarified. It's 'all' co-authors.

We would like to thank the reviewers for their suggestions, edits and questions which contributed to a hopefully improved revised manuscript. In the following, please find our point by point response to the reviewer' comments.

Anonymous Referee #2

Received and published: 4 May 2020

The authors report an interesting work of applying a deep learning model to 16 years of satellite data to create an observational classification of marine low cloud mesoscale morphology. The deep learning technique is quite novel in this area of remote sensing measurement and analysis. The science topic is also of interest to the atmospheric and climate science community. The paper is well written. I only have a few minor comments and questions for the authors to consider for improving the presentation quality of the paper.

Specific comments:
Line 18: Considering that AMT is an international journal, the authors might want to clarify on "NASA funded project" or remove it (which I don't think is critical to mention here)
Changed to simply 'our project'.

Line 21, Line 52, Line 77: Are these (128x128 or 256x256) the number of pixels? Is the pixel size 250 m? Please clarify in the main text. I wonder how the size of each scene has been determined. I imagine that a too big or too small size might cause some ambiguity in the classification of mesoscale cloud morphology. For example, some of the disorganized MCC scenes in Figure 7 look like evolving open-cell or closed-cell MCC. Have any sensitivity tests been performed to decide on the scene size for the training data?
Great point. We indeed spent months thinking about this question before deciding on 128x128 pixels. The pixel size is close to 1km. The main consideration is that if the size gets too large, e.g. 256x256, the chance of mixed types in a scene increases. On the other hand, if the size is too small, the lack of context renders classification by even humans hard because it can become quite ambiguous. We added a sentence at L119 to reflect this.

The example you raised for Figure 7 is important in showing that the scale really matters. The difference is more apparent at the native 128x128 scale. Looking at the scene when zooming out, some of the disorganized MCCs indeed can be confused with open-cell MCC.

Line 97-98: Except for the scenes got filtered out, does each scene have to belong to one of the six types when being analyzed for the frequency distribution? Please clarify.
Yes. We added a sentence for this point. "In the current version, each low cloud scene will be assigned one of these six types." At L103.

Line 105: Is the droplet size information used for disorganized MCC in the classification algorithm? This could be useful to remove ambiguity mentioned above.

We did not include the droplet size information. It would indeed provide extra information in many circumstances. However, including it would make the algorithm less general. We opt to not include it in this trade-off.

Line 141: how does rotating or flipping scenes help to increase the open-cell MCC sample size? That makes me wonder how the orientation of each scene affects the pattern recognition of the deep learning model here.
Rotating and flipping are standard operations to enhance the sample size as well as the robustness of the algorithm. A robust algorithm should be agnostic to orientation and vantage point. We increased the sample size of open-cell scenes to reduce the imbalance between cloud types.

Line 160-162: Was each scene in the training dataset labeled by at least two people? How if there is a disagreement?
Not every scene was repeatedly labelled by two experts, but there are hundreds of scenes that are labeled by at least two experts. When there is disagreement, an accompanying discussion can be found online. In these situations, we also examine the scenes closely to determine the true label.

Line 226-227: Please clarify on the "internal mechanisms". Are you referring to the self-organizing mechanisms (e.g., Feingold et al., 2010)? Feingold G, Koren I, Wang H, Xue H, Brewer WA. (2010): Precipitation-generated oscillations in open-cellular cloud fields. Nature 466:doi:10.1038/nature09314.
Yes, this paper would definitely count as supporting the internal mechanism hypothesis. We want to focus our discussion and to avoid diverting attention, we removed this speculation and only discuss the results we have.

Line 330: More details are needed for the PDFs in Figure 4. How many scenes? What time periods and regions?
Good point. We added relevant information in the revision at L 130 as well as in the caption for Figure 4. We randomly selected 1000 scenes for each cloud type from 2006 data in the southeast Pacific region. This is clarified in the text and caption.

Line 78: No more than 10% of the scenes got filtered out? Please clarify.
Changed to clarify. We remove scenes with more than 10% land cover.

Line 117: classify -> classifying Line 146: remove the first "low"
Changed. We removed the first 'low'.

Line 311: units of LWP in Figure 2 are wrong.
Nice catch! We changed the units to kg/m^2.

Applying Deep Learning to NASA MODIS Data to Create a Community Record of Marine Low Cloud Mesoscale Morphology

Tianle Yuan[1,2], Hua Song[3], Robert Wood[4], Johannes Mohrmann[4], Kerry Meyer[1], Lazaros Oreopoulos[1], Steven Platnick[1]

[1]Earth Science Directorate, NASA Goddard Space Flight Center
[2]Joint Center for Earth Systems Technology, University of Maryland, Baltimore County
[3]Science Systems and Applications, Inc.
[4]Department of Atmospheric Sciences, University of Washington

Correspondence: tianle.yuan@nasa.gov

Abstract:

Marine low clouds display rich mesoscale morphological types, distinct spatial patterns of cloud fields. Being able to differentiate low cloud morphology offers a tool for the research community to go one step beyond bulk cloud statistics such as cloud fraction and advance the understanding of low clouds. Here we report the progress of our project that aims to create an observational record of low cloud mesoscale morphology at a near-global (60S-60N) scale. First, a training set is created by our team members manually labeling thousands of mesoscale (128x128) MODIS scenes into six different categories: stratus, closed cellular convection, disorganized convection, open cellular convection, clustered cumulus convection, and suppressed cumulus convection. Then we train a deep convolutional neural network model using this training set to classify individual MODIS scenes at 128x128 resolution, and test it on a test set. The trained model achieves a cross-type average precision of about 93%. We apply the trained model to 16 years of data over the Southeast Pacific. The resulting climatological distribution of low cloud morphology types shows both expected and unexpected features and suggests promising potential for low cloud studies as a data product.

Deleted: a

Deleted: NASA funded

1. Introduction

Marine low clouds are important for the mass, heat, and momentum transport in the planetary boundary layer (PBL) and between the PBL and free troposphere, the radiative energy balance of the climate, and the magnitude of feedback strength under climate change. Observations of marine low clouds are indispensable for advancing our understanding of these clouds for deriving new theories and insights and for model validation and constraining. Modern satellite observations have the advantage of providing global and long-term coverage. Current satellite products offer detailed pixel-level retrievals of cloud properties such as cloud optical depth, cloud droplet effective radius, and cloud phase. Most cloud classification schemes are based on either single pixel measurements or joint-histograms of two cloud properties.

However, marine low clouds are known to have various mesoscale morphology types since first satellite observations of clouds became available (Agee and Dowell, 1974). These mesoscale morphology types are created by the characteristic patterns into which clouds are organized (Figure 1). Cloud mesoscale morphology types are not only phenological classifications of satellite images, but also manifestation of complex mixture of underlying physical processes (Atkinson and Zhang, 1996; Stevens et al., 2005; Wang and Feingold, 2009; Wood, 2012; Wood and Hartmann, 2006). These physical processes are critical for fundamental understanding and better modeling of marine low clouds because of their impact on mass, heat, and momentum transport, on radiative energy balance, and their feedbacks to climate change. Wood and Hartmann (2006) trained a two-layer neural network on probability distribution functions and 2-d power spectra of liquid water path to classify cloud morphology into four categories for 256x256 scenes. The method has been successfully used to analyze morphology types and associated cloud properties (McCoy et al., 2017; Muhlbauer et al., 2014).

Here we introduce a NASA funded project to classify marine low cloud observations into six different mesoscale morphology types based directly on full images without engineering features. The goal is to produce a community data record that spans about two decades at near-global scales that will enable the research community to go beyond bulk cloud statistics and will advance our understanding of low-level mesoscale convective clouds through exploiting the rich spatial information content of observations. Section 2 describes the data and methodology; section 3 introduces preliminary results and section 4 gives discussions of future plans and outlook of the data product; section 5 concludes.

2. Data and methods

a. Data source

The primary observational data for this study are from the MODerate resolution Imaging Spectrometer (MODIS) onboard the Aqua satellite. We use reflectance from channels 1 (0.65µm) , 3 (0.47µm), and 4 (0.55µm) and cloud optical depth, cloud droplet effective radius, cloud mask, and cloud top height from the MODIS cloud product (Platnick et al., 2017) in building up the training set. The spatial resolution of these parameters is 1km at nadir. The cloud optical depth and effective radius retrievals are combined to produce cloud liquid water path (Platnick et al., 2017). Reflectance from channel 4 is used for deep neural network model training and inference, while the other MODIS observations and products are used for data quality control, filtering, and contextual information, as explained below.

We first break MODIS images into 128x128 pixels scenes. The selection of 128x128 results from a balance because larger sizes suffer from too much mixing of different types in a scene while smaller sizes contain not enough contextual information for classification. We filter out scenes that contain significant fraction of high clouds (no more than 10%), defined as pixels with cloud top height above 6km, or whose low cloud fraction is lower than 5%. We also exclude scenes whose viewing zenith angle is greater than 45 degrees. Scenes with more than 10% land coverage are also excluded. The resulting scenes are treated as dominated by marine low clouds.

Deleted: (Platnick et al., 2016)

Field Code Changed

Deleted: (Platnick et al., 2016)

Field Code Changed

Deleted:

Deleted: and

Deleted:

For training purpose, we create auxiliary images that contain the broad context of the scene of interest and distributions of the liquid water path and cloud top height for the scene (Figure 2). The scene image together with the auxiliary images are presented to a panel of human experts on the Zooniverse platform (www.zooniverse.org) for manual labeling. We intend to use the same platform in the future to crowdsource the labeling task.

Spatiotemporally collocated Modern-Era Retrospective analysis for Research and Applications, version 2 (MERRA-2) (Gelaro et al., 2017) data is used to provide meteorological variables for each scene.

b. Morphology types

Marine low cloud mesoscale morphology patterns are extremely diverse. In order to keep the task manageable, we settle on six representative types. They are stratus, closed cellular convection, disorganized cellular convection, open cellular convection, clustered cumulus, and suppressed cumulus (Figure 3). These types are by no means exhaustive given the diversity of observable patterns. However, these six types are the most common and largely representative of the data when we inspect a large collection of scenes. In the current version, each low cloud scene will be assigned one of these six types. We also believe that these types have distinct underlying physical processes. Stratus is mostly created by relatively uniform radiative cooling or driven by synoptic weather systems such as fronts while closed cellular convection is driven by radiative cooling and organized into distinctive honeycomb mesoscale patterns. Disorganized cellular convection is characterized by a combination of elements of convection and large portion of stratiform clouds that tend to have large droplet sizes and small cloud optical depths, creating their characteristic appearance. Their cellular sizes are typically larger, on the order of 100km, compared to closed cellular convection, on the order of 10km. Open cellular convection is characterized by cells that are clear in the center and exhibit vigorous shallow convection around it. These convective clouds are often precipitating based on satellite and ship-based observations, which is a likely driving force that creates and maintains this mesoscale morphology type (Wang and Feingold, 2009). Clustered cumulus convection is made up of shallow, vigorous convective elements that aggregate together, accompanied by scattered shallower and optically thinner cumulus clouds nearby. The suppressed cumulus type is dominated by individual, scattered cumulus clouds that can sometimes have patterns like lines and branches.

Deleted: to keep the task manageable

c. Method

To illustrate the difficulty of classifying morphology types using one-point statistics such as histograms, we show the mean probability density functions (PDFs) of cloud optical depth and droplet effective radius for each type in Figure 4. We randomly select 1000 scenes for each cloud type from 2006 data in the Southeast Pacific region. The significant overlap between PDFs of different types makes it quite hard to classify the scenes based on these PDFs. On the other hand, deep convolutional neural network (DCNN) models have been shown to separate complex patterns into different categories at a human level (LeCun et al., 2015). We apply a transfer learning approach to our classification task in a supervised fashion although separate efforts of unsupervised training also seem promising (Yuan, 2019).

Specifically, we use a pretrained model (Simonyan & Zisserman, 2015) as a feature extractor and fine-tune it with our training set. The pretrained model is a 16-layer DCNN that is trained on the large-scale ImageNet dataset (Deng et al., 2009). Its weights are fixed. We add three additional layers to the pretrained model, called VGG-16 and train the resulting full model on our training set, the fine-turning step. The output of the full DCNN model is a six-element vector whose elements sum up to 1 and are interpreted as the probability that the model assigns to one of the corresponding types. We assign every scene to the type that has the highest probability and therefore effectively we have a metric to measure how confident the model is for each classification, which provides useful information for users who may apply filters to the data.

To build the training set, our team together with several expert level volunteers first manually labeled thousands of scenes using the Zooniverse online tool. We retain only those scenes that are unambiguously belonging to a certain type to present the best possible training set, which includes hundreds of samples for each type. We augment the training set by rotating each scene by 90 and 180 degrees and also flipping the open cellular scenes to increase their sample size. The flipping operation is achieved by mirroring the original image across a horizontal axis.

3. Results

Here we report results for the training, show the classification at work at a granule level and for two typical low marine low cloud regimes: winter time mid-latitude region downwind of the East Coast of US and Canada and sub-tropical Southeast Pacific region.

a. Training performance

The training asymptotically converges to a plateau in terms of accuracy pretty quickly, within about 30 epochs (Figure 5). Around epoch 30, the validation accuracy reaches a maximum. The training and validation accuracies are at around 98% and 93%. We save the model configuration with the best validation accuracy. After training, the model is applied to a test set that it has never seen before. The resulting confusion matrix is shown in Figure 6. The confusion matrix summarizes the classification prediction results. For each cloud type, or row, it shows the percentage of correct predictions on the diagonal and percentages of incorrect predictions off the diagonal. The trained model achieves an average precision of about 93% across different types. Open cellular and disorganized cellular convection, are the two morphology types with the lowest accuracy mainly because they had the lowest number of training samples. With further increase in training samples in the future, we are confident that corresponding accuracies can be further improved. The biggest challenge for the model comes from separating disorganized cellular, open cellular, and clustered cumulus types. It is also worth noting that there is inherent uncertainty with the classification since even expert labelers sometimes disagree on the same scenes.

b. An example granule

An example of a classified MODIS granule is shown in Figure 7. The classification results are overlaid on the visible MODIS image as colored circles whose position represents the center of corresponding 128x128 scene. This is a low cloud dominated granule with a complex mix of different morphology types. The few missing scenes within the viewing zenith angle limits are due to subvisible high clouds overlapping the visible low clouds, which is not rare even for these low cloud dominated regions (Yuan and Oreopoulos, 2013), as well as a couple of scenes with too little low clouds. One can visually confirm that the model performs quite well in picking up morphology types and their transitions corroborating the results in Figure 5. It is worth noting that a scene does not have to be fully occupied by a cloud type to be classified into this particular type. For example, the scene centered around 14S and 78W is partially occupied by stratus and nonetheless classified as stratus.

c. Test run over the wintertime Northwest Atlantic

During the winter, there can be many cold air outbreak events over the Northwest Atlantic region. They create maritime low cloud systems with various mesoscale morphology types. We apply our model to data in winter of 2011. We first filter the raw data to include only marine low cloud scenes using the criteria discussed in section 2. The 128x128 pixel scenes are fed into the trained DCNN model for classification. For each scene, we record its morphology type, geolocation, time and save the 2-D MODIS cloud retrieval parameters such as cloud optical depth, cloud droplet effective radius, and cloud top pressure. In this run, we do not oversample the data and therefore scenes do not overlap with each other.

Figure 8 shows frequency of occurrence maps for each cloud type along with surface wind vectors. Stratus clouds dominate in the Hudson Bay and Labrador Sea. They also frequently appear over waters around Newfoundland and, to a lesser degree, along the east coast of US and Canada. There is also a local maximum in the western part of the Gulf of Mexico. Closed cellular type dominates the warm water of the Gulf Stream where cold continental air meets the warm water, which induces large flux of moisture and heat from the ocean into the boundary layer and gives rise to formation of low clouds. These low clouds mostly appear as the closed cellular type according to MODIS. The disorganized type only appears in significant quantity in the subtropics away from the coast. Open cellular clouds peak in the area south of the Greenland and in the Labrador Sea and have a local maximum that is centered around 60°W and 35°N. Both are downwind of the closed cellular cloud peaks. The clustered and suppressed cumulus clouds mostly occur in the subtropics and tropics.

Deleted: the

d. Results over the Southeast Pacific region

We obtained all relevant Aqua MODIS level-1b and level-2 files for the Southeast Pacific region (5°S-45°S, 70°W-125°W) between 2003 and 2018. The total volume of data is about 30 Tb. This region is well known for the semi-permanent stratocumulus clouds.

Figure 9 shows the 16-year climatology of sea surface temperature (SST), estimated inversion strength (EIS) (Wood and Bretherton, 2006), and frequency of occurrence maps for each morphology type in the Southeast Pacific region. The frequency is normalized by the number of total MODIS scenes, including both low cloud and non-low cloud ones.

Stratus clouds predominantly occur near coastal upwelling regions in the subtropics as well as in the mid-latitude regions south of 40 degrees. Both features agree with our expectations. Stratus can still occur in other parts of the domain, but with frequencies generally below 10%. Their frequency significantly drops away from the local maxima in the mid-latitudes and along the coast. The local maxima of stratus occurrence frequency coincide spatially with cold SST.

The closed cellular type occurs most frequently about five hundred kilometers away from the coastlines. The absolute maximum is located around 27^{o}S and 75^{o}W, which is also where EIS peaks. Indeed, the frequency of closed cellular type roughly correlates with the EIS pattern. The frequency of this type drops off from its peak location more gradually compared to that of the stratus. Its frequency is nevertheless below 10% west of 90^{o}W and the direction of the frequency of occurrence gradient is almost east to west. The location of peak frequency for the disorganized type is further away from the coast and occurs around 21^{o}S and 89^{o}W. The frequency map of this type also has an overall correlation with the EIS west of 90^{o}W.

The frequency map for the open cellular type is the most distinct. Its peak features a bullseye pattern and occurs further downwind of the peak of the disorganized type, with a peak frequency of only about 10%. This type also appears relatively frequently in the mid-latitudes associated with mid-latitude cyclones. Its spatial pattern has no direct correlation with either EIS or SST patterns, possibly implying internal mechanisms that are responsible for their appearances. Both the closed and open cellular locations agree qualitatively with the findings from Wood and Hartmann (2006), although the addition of other cloud types resulted in lower frequencies of these types in our dataset. It is also worth mentioning that the disorganized cellular type has a different geographic occurrence when compared to Wood and Hartmann (2006). This is because under that classification scheme, 'disorganized' includes the bulk of scenes which we classify as suppressed and clustered; the more narrowly-defined disorganized cellular type in our classification is geographically more closely associated with the other cellular cloud types. The clustered cumulus type occurrence appears to have a general anticorrelation with the EIS map. The suppressed cumulus type occurs most frequently in the tropics where the SST is the warmest.

4. Discussions and future work

a. Notable new insights

Open cellular clouds are less prevalent than previously thought (Atkinson and Zhang, 1996; McCoy et al., 2017; Muhlbauer et al., 2014), especially in subtropical regions. We attribute this to the combination of advanced quantitative observation techniques developed here and the delineation of clustered cumulus and open cellular types. The early studies did not have comprehensive observations to rely on. The more recent results may have included the two types together into the open cellular type, which overestimated the occurrence frequency of the open cellular type in the subtropics. However, given the relatively minor presence of clustered cumulus type in the midlatitudes, the open cellular type may indeed be quite prevalent there, which agrees with previous studies.

There is a strong spatial correlation between both EIS and SST and the frequency of stratus in two regions analyzed, especially north of 35ºN, suggesting a strong control of atmospheric stability and cold SST on this cloud type in higher latitude regions. Their control on other cloud types may not be as tight given the loose spatial correspondence between both EIS and SST and frequency of other cloud types, implying either other large-scale variables are in control or internal cloud processes are more important. We will leave such explorations for future studies.

b. Expanding the scale of test runs and further analysis

We plan to expand the test run to near-global scales for about two years. These runs will include time periods that overlap those of several field campaigns that have rich in-situ and ground and airborne remote sensing data. Together with these datasets, the satellite product will help to advance the understanding of low cloud mesoscale morphology. The global scale will also allow us to examine the general distributions of morphology types and intercompare the characteristics of low cloud morphology in different ocean basins. Further data analysis of the current test run and future runs will target questions related to the variability of low cloud morphology and its driving forces. We plan to release part or all of the test run results to beta testers for feedback and test use from the community.

c. Collocating with other satellite sensors and meteorology

We plan to collocate each classified low cloud scene with data from sensors like CloudSat cloud profiling radar, CALIOP lidar, the Advanced Microwave Scanning Radiometer for EOS (AMSR-E and AMSR-2), and Atmospheric InfraRed Sounder (AIRS) as well as the MERRA-2 reanalysis products. Such collocated set of variables will be useful to the research community for studying the behavior of low cloud morphology under different environmental conditions

d. Further improvement of the model

The current model works pretty well overall, particularly for closed cellular, suppressed cumulus and clustered cumulus types. However, there is room to improve for other types. We target two fronts for improvement: improving the model itself and increasing the quality and quantity of training data. For the former goal, we plan to test different pre-trained models and what features to keep and how to best set up the classifier on top of these extracted feature vectors. For the latter goal, we have developed analysis tools to help us understand the agreement among human experts in the training set. This helps us to target types that need the improvement. We will use the Zooniverse tool to achieve this. Further increase in training data also allows us to better characterize the uncertainty in expert labeling of each category. We are looking for expert level volunteers to join us to increase the training sample size.

e. Increasing the number of types

Some of the mesoscale types can be further divided into subtypes. For example, the frequency of suppressed cumulus type is quite high in the low latitudes and based on the manual labeling they could be further divided into multiple subtypes. We will explore the feasibility of this by assessing resource constraints and the feedback from the community.

5. Conclusions

We have developed a working deep neural network model to automatically classify cloudy scenes into six mesoscale morphology types. Initial test run results showed promising results for the Southeast Pacific and Northwest Atlantic. Using the tool, we plan to extend the dataset and create a community mesoscale morphology type product for low marine clouds observed by MODIS. We will further develop the product and actively look forward to community involvement such as beta testing, volunteering, and user feedback.

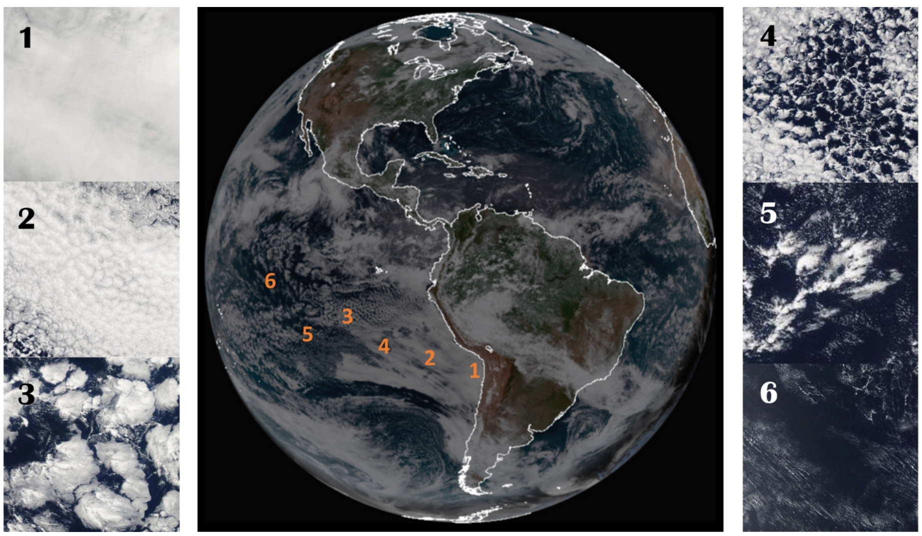


Figure 1: A full disk image of GOES-16 on Aug 6, 2018 and six scenes of MODIS images at smaller scales representing different morphology types at corresponding locations in the GOES image. Except scene 1, all scenes are from the same day. Scene 1 is from a different day because there was no representative stratus scenes on this day in the Southeast Pacific region.

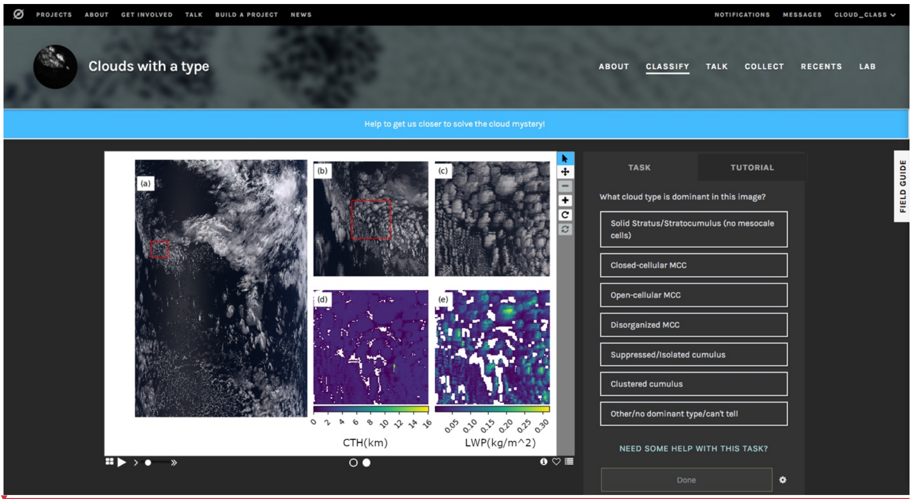
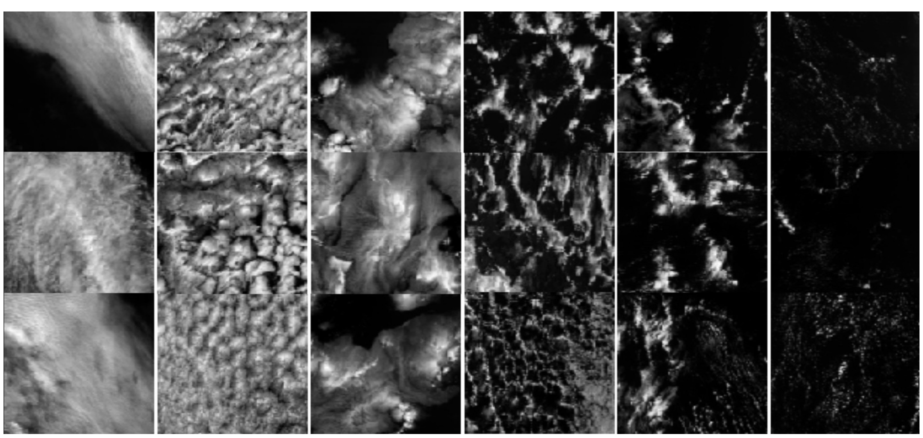


Figure 2: the Zooniverse interface for manual labelling. The center image is made up of five panels. Panel a shows the full granule (usually 2030x1350 pixels) true color image for large context. Panel b shows a portion of the granule immediately surrounding the scene to be labelled, outlined by the red square. Panel c shows the visible scene image while panels d and e show the cloud top height and LWP fields in the scene to be labelled. The panels to the right of the center image show labelling choices. The tutorial document is available by clicking on the 'FIELD GUIDE' tab on the right side. Additional options for scenes with heavily mixed types, scenes with sea ice, or scenes with other issues are found in the 'other' menu. The image is a screenshot of our Zooniverse project.



Deleted:

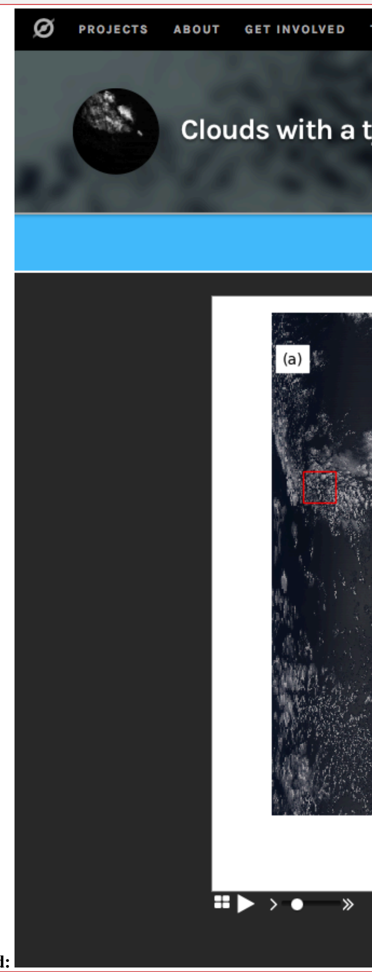


Figure 3: Example scenes of MODIS single channel images for the six different types. From left to right: stratus, closed cellular, disorganized cellular, open cellular, clustered cumulus, and suppressed cumulus types. Images taken by the NASA MODIS.

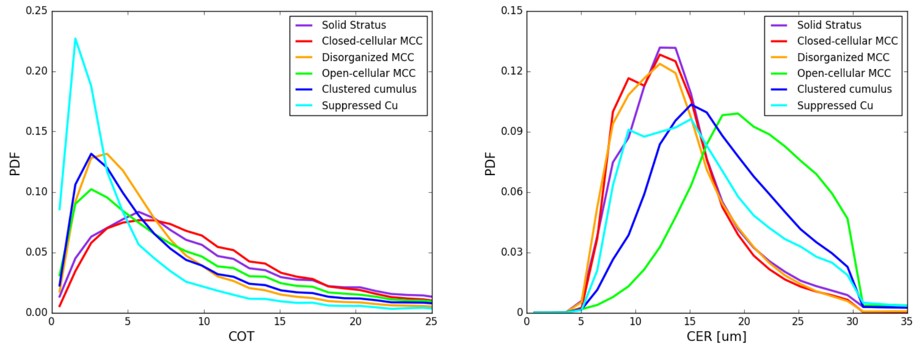


Figure 4: PDFs of cloud optical depth and cloud effective radius for six morphology types. We randomly selected 1000 samples for each type and mean distributions are shown here. Significant overlaps are observed for PDFs of both variables among different morphology types.

Deleted:

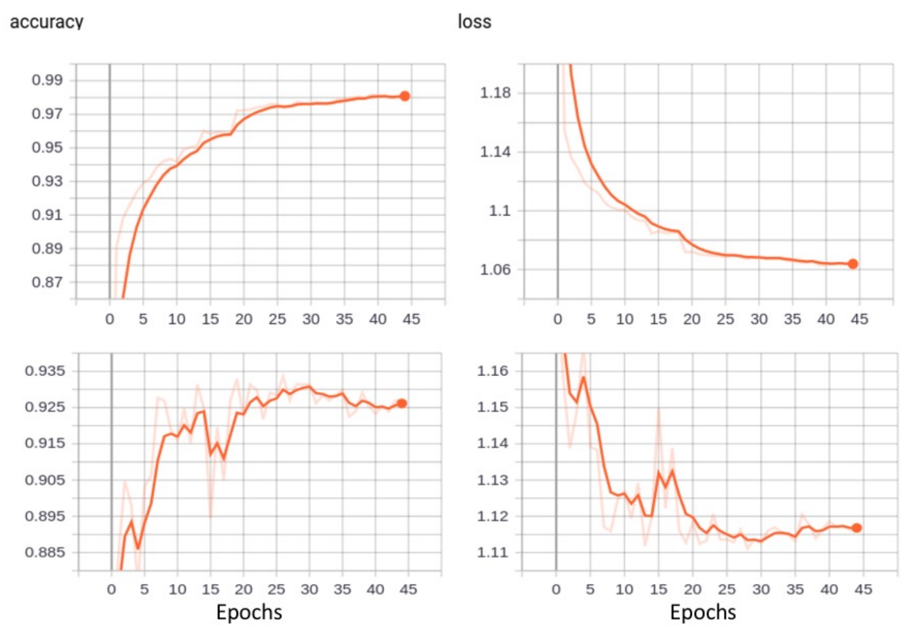


Figure 5: Training (upper two panels) and validation (lower ones) accuracy and loss trajectories. By around epoch 30, the validation accuracy peaks while validation loss bottoms out and the training loss and accuracy asymptotically reach their minimum and maximum, respectively, which indicates further training may be overfitting the model.

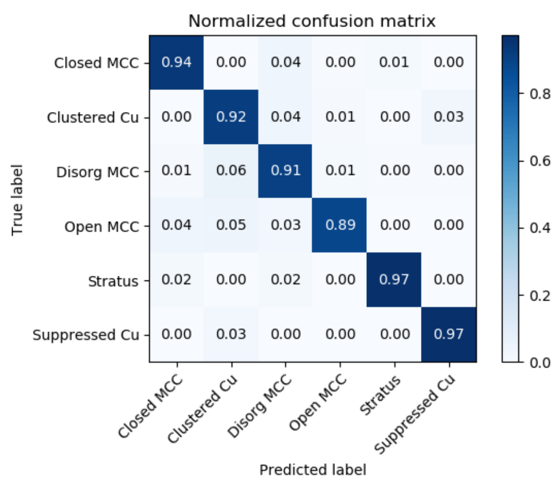


Figure 6: Confusion matrix of the model predictions on test data.

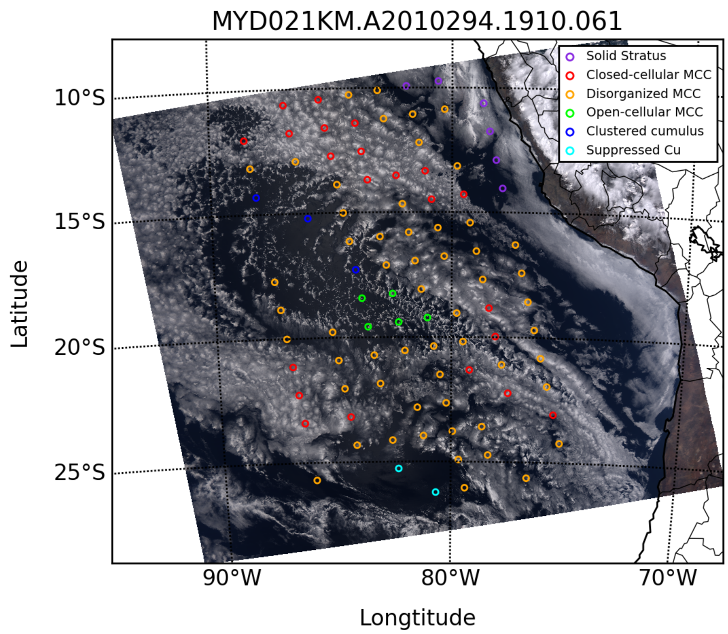


Figure 7: An example granule illustrating the results of the classification algorithm. This is quite a complex granule with different morphology types mixed together. The left and right margins are not classified because current algorithm filters out scenes whose sensor viewing zenith angles are greater than 45 degrees. The image is taken by NASA MODIS.

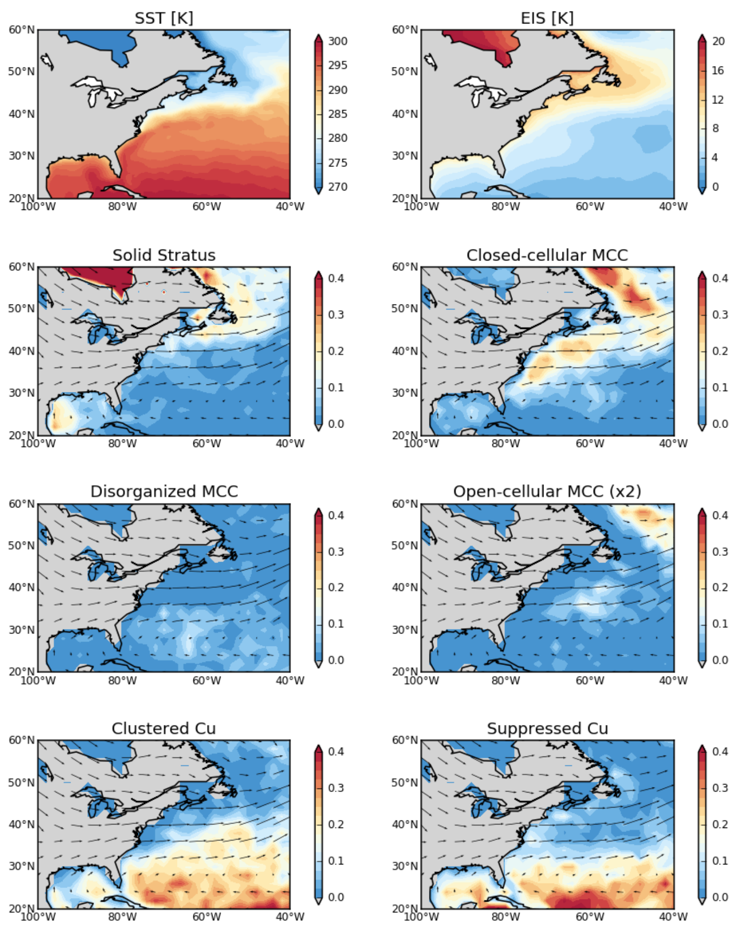


Figure 8: Frequency distributions of six morphology types obtained from the classification algorithm in the Northwest Atlantic region off the east coasts of US and Canada in the winter of 2011. The top two panels show the SST and EIS distributions using MERRA-2. Seasonal mean wind vectors at 850hPa are plotted to illustrate the flow. We double the values for frequency of the open-cellular type to make them numerically comparable with other types.

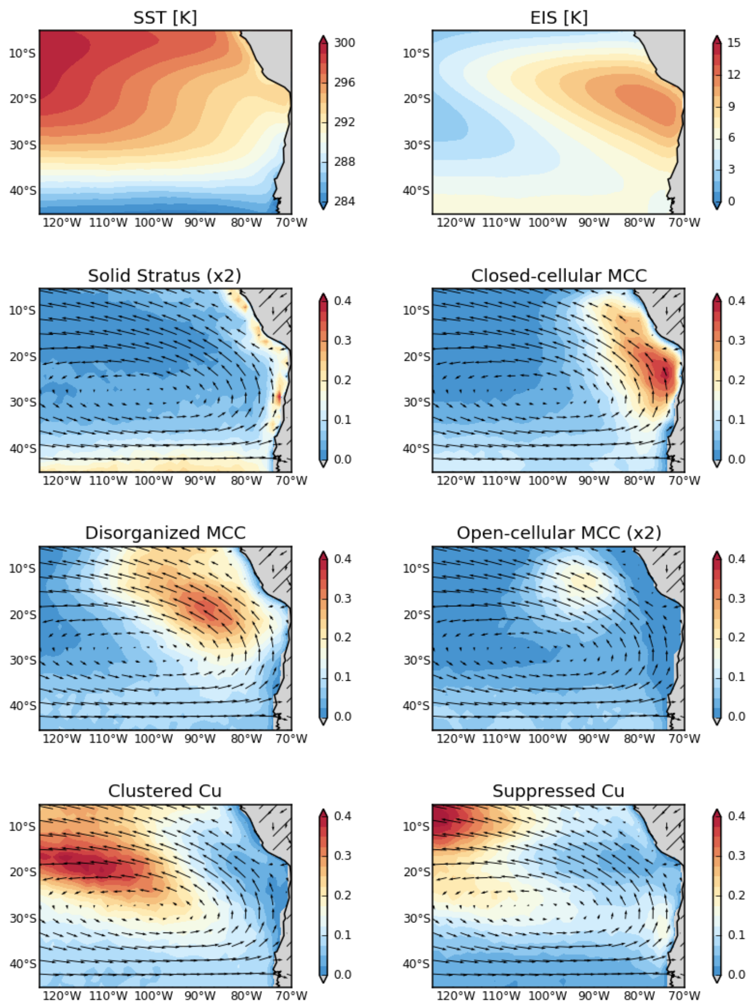


Figure 9: Frequency distributions of various morphology types obtained from the classification algorithm in the subtropical eastern Pacific off the coast of South America for the period 2003-2018. The top two panels show the SST and EIS climatology from MERRA-2 for the same period. Note the doubling of scale on the stratus and open-cellular types.

6. Author Contribution

T. Y. implemented the method to train the network model. H. S., J. M., and T.Y. prepared the training data. All co-authors contributed to compiling the training dataset. T. Y. wrote the manuscript with contributions from all co-authors.

7. Reference

Agee, E. M., & Dowell, K. E. (1974). Observational Studies of Mesoscale Cellular Convection. *Journal of Applied Meteorology*, *13*(1), 46–53. https://doi.org/10.1175/1520-0450(1974)013<0046:OSOMCC>2.0.CO;2

Atkinson, B. W., & Zhang, W. J. (1996). Mesoscale shallow convection in the atmosphere. *Reviews Of Geophysics*, *34*(4), 403–431. https://doi.org/10.1029/96RG02623

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255). https://doi.org/10.1109/CVPR.2009.5206848

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

McCoy, I. L., Wood, R., & Fletcher, J. K. (2017). Identifying Meteorological Controls on Open and Closed Mesoscale Cellular Convection Associated with Marine Cold Air Outbreaks. *Journal of Geophysical Research: Atmospheres*, *122*(21), 11,678-11,702. https://doi.org/10.1002/2017JD027031

Deleted:

Deleted: other

Formatted: Font: (Default) Calibri

Formatted: Bibliography, Automatically adjust right indent when grid is defined, Widow/Orphan control, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

Muhlbauer, A., McCoy, I. L., & Wood, R. (2014). Climatology of stratocumulus cloud morphologies: microphysical properties and radiative effects. *Atmospheric Chemistry And Physics*, *14*(13), 6695–6716. https://doi.org/10.5194/acp-14-6695-2014

Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., et al. (2017). The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(1), 502–525. https://doi.org/10.1109/TGRS.2016.2610522

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv:1409.1556 [Cs]*. Retrieved from http://arxiv.org/abs/1409.1556

Stevens, B., Vali, G., Comstock, K., Wood, R., van Zanten, M. C., Austin, P. H., et al. (2005). Pockets of open cells and drizzle in marine stratocumulus. *Bulletin Of The American Meteorological Society*, *86*(1), 51—-. https://doi.org/10.1175/BAMS-86-1-51

Wang, H., & Feingold, G. (2009). Modeling mesoscale cellular structures and drizzle in marine stratocumulus. Part I: Impact of drizzle on the formation and evolution of open cells. *Journal Of The Atmospheric Sciences*, *66*(11), 3237–3256. https://doi.org/10.1175/2009JAS3022.1

Wood, R, & Bretherton, C. S. (2006). On the relationship between stratiform low cloud cover and lower-tropospheric stability. *Journal Of Climate*. Retrieved from http://journals.ametsoc.org/doi/pdf/10.1175/JCLI3988.1

Wood, Robert. (2012). Stratocumulus Clouds. *Monthly Weather Review*, *140*(8), 2373–2423. https://doi.org/10.1175/MWR-D-11-00121.1

Wood, Robert, & Hartmann, D. L. (2006). Spatial variability of liquid water path in marine low cloud: The importance of mesoscale cellular convection. *Journal Of Climate*, *19*(9), 1748–1764.

Yuan, T. (2019). Understanding Low Cloud Mesoscale Morphology with an Information Maximizing Generative Adversarial Network. https://doi.org/10.31223/osf.io/gvebt

Yuan, T., & Oreopoulos, L. (2013). On the global character of overlap between low and high clouds. *Geophysical Research Letters*, *40*(19), 5320–5326. https://doi.org/10.1002/grl.50871

Deleted: Agee, E. M., & Dowell, K. E. (1974). Observational Studies of Mesoscale Cellular Convection. *Journal of Applied Meteorology*, *13*(1), 46–53. https://doi.org/10.1175/1520-0450(1974)013<0046:OSOMCC>2.0.CO;2¶
Atkinson, B. W., & Zhang, W. J. (1996). Mesoscale shallow convection in the atmosphere. *Reviews Of Geophysics*, *34*(4), 403–431. https://doi.org/10.1029/96RG02623¶
Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255). https://doi.org/10.1109/CVPR.2009.5206848¶
Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1¶
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539¶
McCoy, I. L., Wood, R., & Fletcher, J. K. (2017). Identifying Meteorological Controls on Open and Closed Mesoscale Cellular Convection Associated with Marine Cold Air Outbreaks. *Journal of Geophysical Research: Atmospheres*, *122*(21), 11,678-11,702. https://doi.org/10.1002/2017JD027031¶
Muhlbauer, A., McCoy, I. L., & Wood, R. (2014). Climatology of stratocumulus cloud morphologies: microphysical properties and radiative effects. *Atmospheric Chemistry And Physics*, *14*(13), 6695–6716. https://doi.org/10.5194/acp-14-6695-2014¶
Platnick, S., Meyer, K. G., King, M. D., & Wind, G. (2016). The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua. *\ldots on Geoscience and \ldots*. Retrieved from http://ieeexplore.ieee.org/abstract/document/7707459/¶
Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv:1409.1556 [Cs]*. Retrieved from http://arxiv.org/abs/1409.1556¶
Stevens, B., Vali, G., Comstock, K., Wood, R., van Zanten, M. C., Austin, P. H., et al. (2005). Pockets of open cells and drizzle in marine stratocumulus. *Bulletin Of The American Meteorological Society*, *86*(1), 51—-. https://doi.org/10.1175/BAMS-86-1-51¶
Wang, H., & Feingold, G. (2009). Modeling mesoscale cellular structures and drizzle in marine stratocumulus. Part I: Impact of drizzle on the formation and evolution of open cells. *Journal Of The Atmospheric Sciences*, *66*(11), 3237–3256. https://doi.org/10.1175/2009JAS3022.1¶
Wood, R, & Bretherton, C. S. (2006). On the relationship between stratiform low cloud cover and lower-tropospheric stability. *Journal Of Climate*. Retrieved from http://journals.ametsoc.org/doi/pdf/10.1175/JCLI3988.1¶
Wood, Robert. (2012). Stratocumulus Clouds. *Monthly Weather Review*, *140*(8), 2373–2423. https://doi.org/10.1175/MWR-D-11-00121.1¶ ... [1]

Point-by-point response to reviewers:

We would like to thank the reviewers for their suggestions, edits and questions which contributed to a hopefully improved revised manuscript. In the following, please find our point by point response to the reviewer's comments.

Anonymous Referee #1
Received and published: 30 April 2020

Yuan et al. present a method to identify marine low-level cloud regimes. Using MODIS reflectances, and creating a training dataset by human visual inspection, they apply a Deep convolutional neural network to objectively assign each scene to one of six pre-defined types. The method is well described and carefully evaluated. The authors aim to make their product publicly available which is potentially of great usefulness to studies of clouds. The paper is very well written and of interest to the readership of Atmos. Meas. Tech. I only have a few minor remarks which the authors should consider in a revision.

l27 "shows"
Changed.

l28 "suggests"
Changed.

l39 "histograms"; however aren't pixel-level retrievals and joint histograms redundant? the latter is just a way to statistically retain the pixel-level information at level 3 aggregation.
Correct. It depends on how the pixel-level retrievals are used. But ultimately, they both use the pixel level retrievals. Some methods use a one-dimensional PDF and others may use joint-histograms.

l41 only since then? Or not rather since ever / since the first cloud observations (such as Howard,
The reference we used is the best example we are aware of. Could the reviewer provide us with a complete reference?

l62 "a plan" or "plans"
Changed.

l70 The Platnick reference should be updated (actual author list is longer, and it appeared 2017 (vol 55)).
Changed.

l71 Please specify the horizontal resolution for reflectances and retrieval products.
Added.

Formatted: Font: Bold

l77 Provide the unit here. I assume it is 128 x 128 pixels of 1x1 km2 size each?
Changed. Since the pixel size changes slightly with view angle we now simply quote scene size in terms of pixels rather than physical area.

l84 It is a nice idea to include this a bit technical detail. This illustrated well what is actually done.

l87 And this is a good idea!
Thanks!

l94 Omit "keep the task manageable" once.
Changed.

l119 Are the PDFs exactly the retrievals from the scenes provided in Fig. 3? It would be good if it was such, and should be clarified in the text.
The PDFs are mean distributions of samples belonging to a particular type. We randomly selected 1000 scenes for each cloud type.

l141 I don't understand what "flipping" means if not rotating by 180◦ . The authors should clarify this.
Now clarified.

l154 It would be useful to explain in one sentence to the non-specialized readership what the confusion matrix is.
Now explained.

l161 It would be interesting to know how often this occurs for the different cloud types. E.g. a fraction of disagreement for each type?
With the added explanation about the confusion matrix, readers should be able to read how often misclassifications occur for each type and how they distribute across different types.

l165 This mostly looks quite reasonable. However, some results seem rather strange to the naked eye. E.g. where the solid stratus diagnosed at 14◦S/78◦W I don't see any cloud, let alone a stratus.
The dot represents the center location of a scene. The right half of this scene is indeed occupied by stratus clouds. A scene does not have to be mostly cloudy to be classified as stratus. The algorithm examines the textual information of the clouds.

l186 drop "the"
Changed.

l338 "indicates". And what is the difference between the light pink and red lines?
Changed. The solid lines are running means of the light pink lines.

l365 Help from which other authors?
Clarified.

We would like to thank the reviewers for their suggestions, edits and questions which contributed to a hopefully improved revised manuscript. In the following, please find our point by point response to the reviewer' comments.

Anonymous Referee #2

Received and published: 4 May 2020

The authors report an interesting work of applying a deep learning model to 16 years of satellite data to create an observational classification of marine low cloud mesoscale morphology. The deep learning technique is quite novel in this area of remote sensing measurement and analysis. The science topic is also of interest to the atmospheric and climate science community. The paper is well written. I only have a few minor comments and questions for the authors to consider for improving the presentation quality of the paper.

Specific comments:
Line 18: Considering that AMT is an international journal, the authors might want to clarify on "NASA funded project" or remove it (which I don't think is critical to mention here)
Changed.

Line 21, Line 52, Line 77: Are these (128x128 or 256x256) the number of pixels? Is the pixel size 250 m? Please clarify in the main text. I wonder how the size of each scene has been determined. I imagine that a too big or too small size might cause some ambiguity in the classification of mesoscale cloud morphology. For example, some of the disorganized MCC scenes in Figure 7 look like evolving open-cell or closed-cell MCC. Have any sensitivity tests been performed to decide on the scene size for the training data?
Great point. We indeed spent months thinking about this question before deciding on 128x128 pixels. The pixel size is close to 1km. The main consideration is that if the size gets too large, e.g. 256x256, the chance of mixed types in a scene increases. On the other hand, if the size is too small, the lack of context renders classification by even humans hard because it can become quite ambiguous.

The example you raised for Figure 7 is important in showing that the scale really matters. The difference is more apparent at the native 128x128 scale. Looking at the scene when zooming out, some of the disorganized MCCs indeed can be confused with open-cell MCC.

Line 97-98: Except for the scenes got filtered out, does each scene have to belong to one of the six types when being analyzed for the frequency distribution? Please clarify.

Yes. We added a sentence for this point.

Line 105: Is the droplet size information used for disorganized MCC in the classification algorithm? This could be useful to remove ambiguity mentioned above.
We did not include the droplet size information. It would indeed provide extra information in many circumstances. However, including it would make the algorithm less general. We opt to not include it in this trade-off.

Line 141: how does rotating or flipping scenes help to increase the open-cell MCC sample size? That makes me wonder how the orientation of each scene affects the pattern recognition of the deep learning model here.
Rotating and flipping are standard operations to enhance the sample size as well as the robustness of the algorithm. A robust algorithm should be agnostic to orientation and vantage point. We increased the sample size of open-cell scenes to reduce the imbalance between cloud types.

Line 160-162: Was each scene in the training dataset labeled by at least two people? How if there is a disagreement?
Not every scene was repeatedly labelled by two experts, but there are hundreds of scenes that are labeled by at least two experts. When there is disagreement, an accompanying discussion can be found online. In these situations, we also examine the scenes closely to determine the true label.

Line 226-227: Please clarify on the "internal mechanisms". Are you referring to the self-organizing mechanisms (e.g., Feingold et al., 2010)? Feingold G, Koren I, Wang H, Xue H, Brewer WA. (2010): Precipitation-generated oscillations in open-cellular cloud fields. Nature 466:doi:10.1038/nature09314.
Here we are making broad separations between two camps, one advocating for large-scale forcing and the other for an internal mechanism. But yes, this paper would definitely count as supporting the internal mechanism hypothesis.

Line 330: More details are needed for the PDFs in Figure 4. How many scenes? What time periods and regions?
Good point. We added relevant information in the revision. We randomly selected 1000 scenes for each cloud type from 2006 data in the southeast Pacific region.

Line 78: No more than 10% of the scenes got filtered out? Please clarify.
Changed to clarify. We remove scenes with more than 10% land cover.

Line 117: classify -> classifying Line 146: remove the first "low"
Changed.

Line 311: units of LWP in Figure 2 are wrong.
Nice catch! Changed.

Page 17: [1] Deleted Yuan, Tianle (GSFC-613.0)[UNIVERSITY OF MARYLAND BALTIMORE COUNTY] 6/16/20 2:31:00 PM