

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

Tianle Yuan^{1,2}, Hua Song³, Robert Wood⁴, Johannes Mohrmann⁴, Kerry Meyer¹, Lazaros Oreopoulos¹, Steven Platnick¹

¹Earth Science Directorate, NASA Goddard Space Flight Center

²Joint Center for Earth Systems Technology, University of Maryland, Baltimore County

³Science Systems and Applications, Inc.

⁴Department 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.

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.

41 However, marine low clouds are known to have various mesoscale morphology types since first
42 satellite observations of clouds became available (Agee and Dowell, 1974). These mesoscale
43 morphology types are created by the characteristic patterns into which clouds are organized
44 (Figure 1). Cloud mesoscale morphology types are not only phenological classifications of
45 satellite images, but also manifestation of complex mixture of underlying physical processes
46 (Atkinson and Zhang, 1996; Stevens et al., 2005; Wang and Feingold, 2009; Wood, 2012; Wood
47 and Hartmann, 2006). These physical processes are critical for fundamental understanding and
48 better modeling of marine low clouds because of their impact on mass, heat, and momentum
49 transport, on radiative energy balance, and their feedbacks to climate change. Wood and
50 Hartmann (2006) trained a two-layer neural network on probability distribution functions and
51 2-d power spectra of liquid water path to classify cloud morphology into four categories for
52 256x256 scenes. The method has been successfully used to analyze morphology types and
53 associated cloud properties (McCoy et al., 2017; Muhlbauer et al., 2014).

54
55 Here we introduce a NASA funded project to classify marine low cloud observations into six
56 different mesoscale morphology types based directly on full images without engineering
57 features. The goal is to produce a community data record that spans about two decades at
58 near-global scales that will enable the research community to go beyond bulk cloud statistics
59 and will advance our understanding of low-level mesoscale convective clouds through
60 exploiting the rich spatial information content of observations. Section 2 describes the data and
61 methodology; section 3 introduces preliminary results and section 4 gives discussions of future
62 plans and outlook of the data product; section 5 concludes.

64 **2. Data and methods**

65 a. Data source

66 The primary observational data for this study are from the MODerate resolution Imaging
67 Spectrometer (MODIS) onboard the Aqua satellite. We use reflectance from channels 1
68 ($0.65\mu\text{m}$), 3 ($0.47\mu\text{m}$), and 4 ($0.55\mu\text{m}$) and cloud optical depth, cloud droplet effective radius,
69 cloud mask, and cloud top height from the MODIS cloud product (Platnick et al., 2017) in
70 building up the training set. The spatial resolution of these parameters is 1km at nadir. The
71 cloud optical depth and effective radius retrievals are combined to produce cloud liquid water
72 path (Platnick et al., 2017). Reflectance from channel 4 is used for deep neural network model
73 training and inference, while the other MODIS observations and products are used for data
74 quality control, filtering, and contextual information, as explained below.

75
76
77 We first break MODIS images into 128x128 pixels scenes. The selection of 128x128 results from
78 a balance because larger sizes suffer from too much mixing of different types in a scene while
79 smaller sizes contain not enough contextual information for classification. We filter out scenes
80 that contain significant fraction of high clouds (no more than 10%), defined as pixels with cloud
81 top height above 6km, or whose low cloud fraction is lower than 5%. We also exclude scenes
82 whose viewing zenith angle is greater than 45 degrees. Scenes with more than 10% land
83 coverage are also excluded. The resulting scenes are treated as dominated by marine low
84 clouds.

85
86 For training purpose, we create auxiliary images that contain the broad context of the scene of
87 interest and distributions of the liquid water path and cloud top height for the scene (Figure 2).
88 The scene image together with the auxiliary images are presented to a panel of human experts
89 on the Zooniverse platform (www.zooniverse.org) for manual labeling. We intend to use the
90 same platform in the future to crowdsource the labeling task.

91
92 Spatiotemporally collocated Modern-Era Retrospective analysis for Research and Applications,
93 version 2 (MERRA-2) (Gelaro et al., 2017) data is used to provide meteorological variables for
94 each scene.

95
96 b. Morphology types
97 Marine low cloud mesoscale morphology patterns are extremely diverse. In order to keep the
98 task manageable, we settle on six representative types. They are stratus, closed cellular
99 convection, disorganized cellular convection, open cellular convection, clustered cumulus, and
100 suppressed cumulus (Figure 3). These types are by no means exhaustive given the diversity of
101 observable patterns. However, these six types are the most common and largely representative
102 of the data when we inspect a large collection of scenes. In the current version, each low cloud
103 scene will be assigned one of these six types. We also believe that these types have distinct
104 underlying physical processes. Stratus is mostly created by relatively uniform radiative cooling
105 or driven by synoptic weather systems such as fronts while closed cellular convection is driven
106 by radiative cooling and organized into distinctive honeycomb mesoscale patterns.
107 Disorganized cellular convection is characterized by a combination of elements of convection
108 and large portion of stratiform clouds that tend to have large droplet sizes and small cloud
109 optical depths, creating their characteristic appearance. Their cellular sizes are typically larger,
110 on the order of 100km, compared to closed cellular convection, on the order of 10km. Open
111 cellular convection is characterized by cells that are clear in the center and exhibit vigorous
112 shallow convection around it. These convective clouds are often precipitating based on satellite
113 and ship-based observations, which is a likely driving force that creates and maintains this
114 mesoscale morphology type (Wang and Feingold, 2009). Clustered cumulus convection is made
115 up of shallow, vigorous convective elements that aggregate together, accompanied by
116 scattered shallower and optically thinner cumulus clouds nearby. The suppressed cumulus type
117 is dominated by individual, scattered cumulus clouds that can sometimes have patterns like
118 lines and branches.

119
120 c. Method
121 To illustrate the difficulty of classifying morphology types using one-point statistics such as
122 histograms, we show the mean probability density functions (PDFs) of cloud optical depth and
123 droplet effective radius for each type in Figure 4. We randomly select 1000 scenes for each
124 cloud type from 2006 data in the Southeast Pacific region. The significant overlap between PDFs
125 of different types makes it quite hard to classify the scenes based on these PDFs. On the other
126 hand, deep convolutional neural network (DCNN) models have been shown to separate
127 complex patterns into different categories at a human level (LeCun et al., 2015). We apply a

128 transfer learning approach to our classification task in a supervised fashion although separate
129 efforts of unsupervised training also seem promising (Yuan, 2019).

130

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

141

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

148

149 **3. Results**

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

153

154 a. Training performance

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

170

171 b. An example granule

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

183

184 c. Test run over the wintertime Northwest Atlantic

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

193

194 Figure 8 shows frequency of occurrence maps for each cloud type along with surface wind
195 vectors. Stratus clouds dominate in the Hudson Bay and Labrador Sea. They also frequently
196 appear over waters around Newfoundland and, to a lesser degree, along the east coast of US
197 and Canada. There is also a local maximum in the western part of the Gulf of Mexico. Closed
198 cellular type dominates the warm water of the Gulf Stream where cold continental air meets
199 the warm water, which induces large flux of moisture and heat from the ocean into the
200 boundary layer and gives rise to formation of low clouds. These low clouds mostly appear as the
201 closed cellular type according to MODIS. The disorganized type only appears in significant
202 quantity in the subtropics away from the coast. Open cellular clouds peak in the area south of
203 the Greenland and in the Labrador Sea and have a local maximum that is centered around
204 60°W and 35°N. Both are downwind of the closed cellular cloud peaks. The clustered and
205 suppressed cumulus clouds mostly occur in the subtropics and tropics.

206

207 d. Results over the Southeast Pacific region

208 We obtained all relevant Aqua MODIS level-1b and level-2 files for the Southeast Pacific region
209 (5°S-45°S, 70°W-125°W) between 2003 and 2018. The total volume of data is about 30 Tb. This
210 region is well known for the semi-permanent stratocumulus clouds.

211

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

216
217 Stratus clouds predominantly occur near coastal upwelling regions in the subtropics as well as
218 in the mid-latitude regions south of 40 degrees. Both features agree with our expectations.
219 Stratus can still occur in other parts of the domain, but with frequencies generally below 10%.
220 Their frequency significantly drops away from the local maxima in the mid-latitudes and along
221 the coast. The local maxima of stratus occurrence frequency coincide spatially with cold SST.
222
223 The closed cellular type occurs most frequently about five hundred kilometers away from the
224 coastlines. The absolute maximum is located around 27°S and 75°W, which is also where EIS
225 peaks. Indeed, the frequency of closed cellular type roughly correlates with the EIS pattern. The
226 frequency of this type drops off from its peak location more gradually compared to that of the
227 stratus. Its frequency is nevertheless below 10% west of 90°W and the direction of the
228 frequency of occurrence gradient is almost east to west. The location of peak frequency for the
229 disorganized type is further away from the coast and occurs around 21°S and 89°W. The
230 frequency map of this type also has an overall correlation with the EIS west of 90°W.
231
232 The frequency map for the open cellular type is the most distinct. Its peak features a bullseye
233 pattern and occurs further downwind of the peak of the disorganized type, with a peak
234 frequency of only about 10%. This type also appears relatively frequently in the mid-latitudes
235 associated with mid-latitude cyclones. Its spatial pattern has no direct correlation with either
236 EIS or SST patterns, possibly implying internal mechanisms that are responsible for their
237 appearances. Both the closed and open cellular locations agree qualitatively with the findings
238 from Wood and Hartmann (2006), although the addition of other cloud types resulted in lower
239 frequencies of these types in our dataset. It is also worth mentioning that the disorganized
240 cellular type has a different geographic occurrence when compared to Wood and Hartmann
241 (2006). This is because under that classification scheme, 'disorganized' includes the bulk of
242 scenes which we classify as suppressed and clustered; the more narrowly-defined disorganized
243 cellular type in our classification is geographically more closely associated with the other
244 cellular cloud types. The clustered cumulus type occurrence appears to have a general
245 anticorrelation with the EIS map. The suppressed cumulus type occurs most frequently in the
246 tropics where the SST is the warmest.

247

248 **4. Discussions and future work**

249 a. Notable new insights

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

259

260 There is a strong spatial correlation between both EIS and SST and the frequency of stratus in
261 two regions analyzed, especially north of 35°N, suggesting a strong control of atmospheric
262 stability and cold SST on this cloud type in higher latitude regions. Their control on other cloud
263 types may not be as tight given the loose spatial correspondence between both EIS and SST and
264 frequency of other cloud types, implying either other large-scale variables are in control or
265 internal cloud processes are more important. We will leave such explorations for future studies.
266

267 b. Expanding the scale of test runs and further analysis

268 We plan to expand the test run to near-global scales for about two years. These runs will
269 include time periods that overlap those of several field campaigns that have rich in-situ and
270 ground and airborne remote sensing data. Together with these datasets, the satellite product
271 will help to advance the understanding of low cloud mesoscale morphology. The global scale
272 will also allow us to examine the general distributions of morphology types and intercompare
273 the characteristics of low cloud morphology in different ocean basins. Further data analysis of
274 the current test run and future runs will target questions related to the variability of low cloud
275 morphology and its driving forces. We plan to release part or all of the test run results to beta
276 testers for feedback and test use from the community.
277

278 c. Collocating with other satellite sensors and meteorology

279 We plan to collocate each classified low cloud scene with data from sensors like CloudSat cloud
280 profiling radar, CALIOP lidar, the Advanced Microwave Scanning Radiometer for EOS (AMSR-E
281 and AMSR-2), and Atmospheric InfraRed Sounder (AIRS) as well as the MERRA-2 reanalysis
282 products. Such collocated set of variables will be useful to the research community for studying
283 the behavior of low cloud morphology under different environmental conditions
284

285 d. Further improvement of the model

286 The current model works pretty well overall, particularly for closed cellular, suppressed
287 cumulus and clustered cumulus types. However, there is room to improve for other types. We
288 target two fronts for improvement: improving the model itself and increasing the quality and
289 quantity of training data. For the former goal, we plan to test different pre-trained models and
290 what features to keep and how to best set up the classifier on top of these extracted feature
291 vectors. For the latter goal, we have developed analysis tools to help us understand the
292 agreement among human experts in the training set. This helps us to target types that need the
293 improvement. We will use the Zooniverse tool to achieve this. Further increase in training data
294 also allows us to better characterize the uncertainty in expert labeling of each category. We are
295 looking for expert level volunteers to join us to increase the training sample size.
296

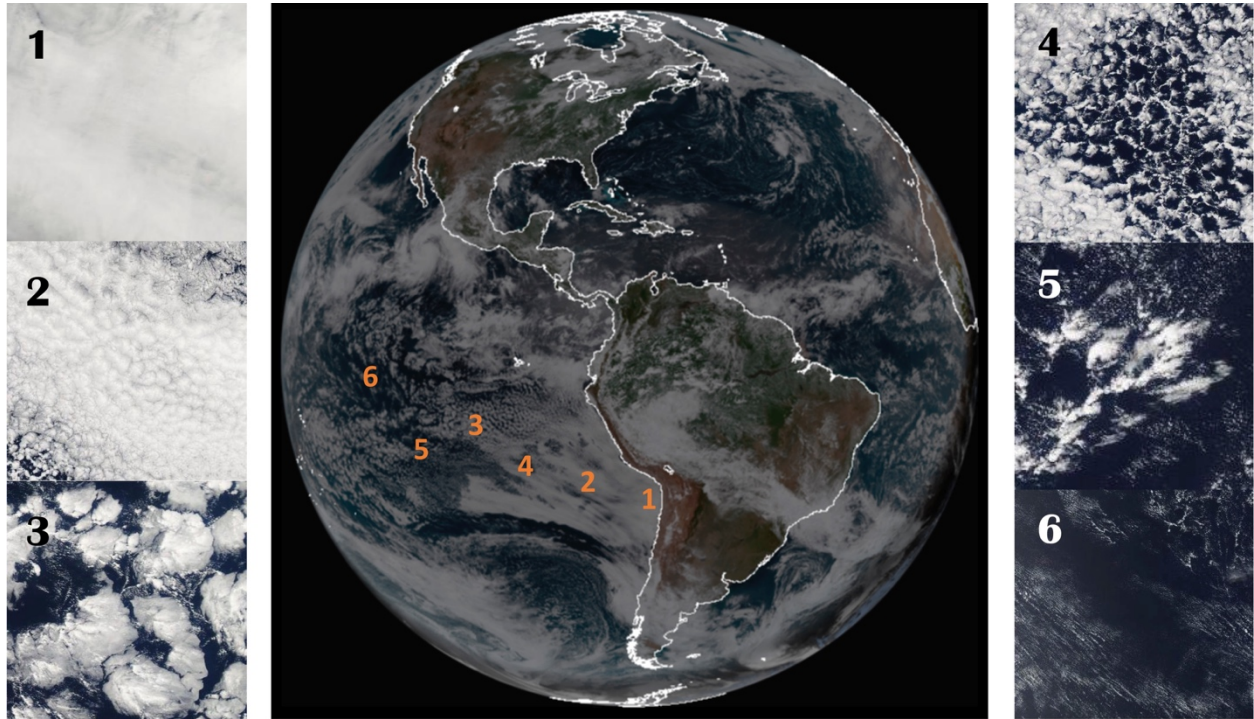
297 e. Increasing the number of types

298 Some of the mesoscale types can be further divided into subtypes. For example, the frequency
299 of suppressed cumulus type is quite high in the low latitudes and based on the manual labeling
300 they could be further divided into multiple subtypes. We will explore the feasibility of this by
301 assessing resource constraints and the feedback from the community.
302
303

304 **5. Conclusions**

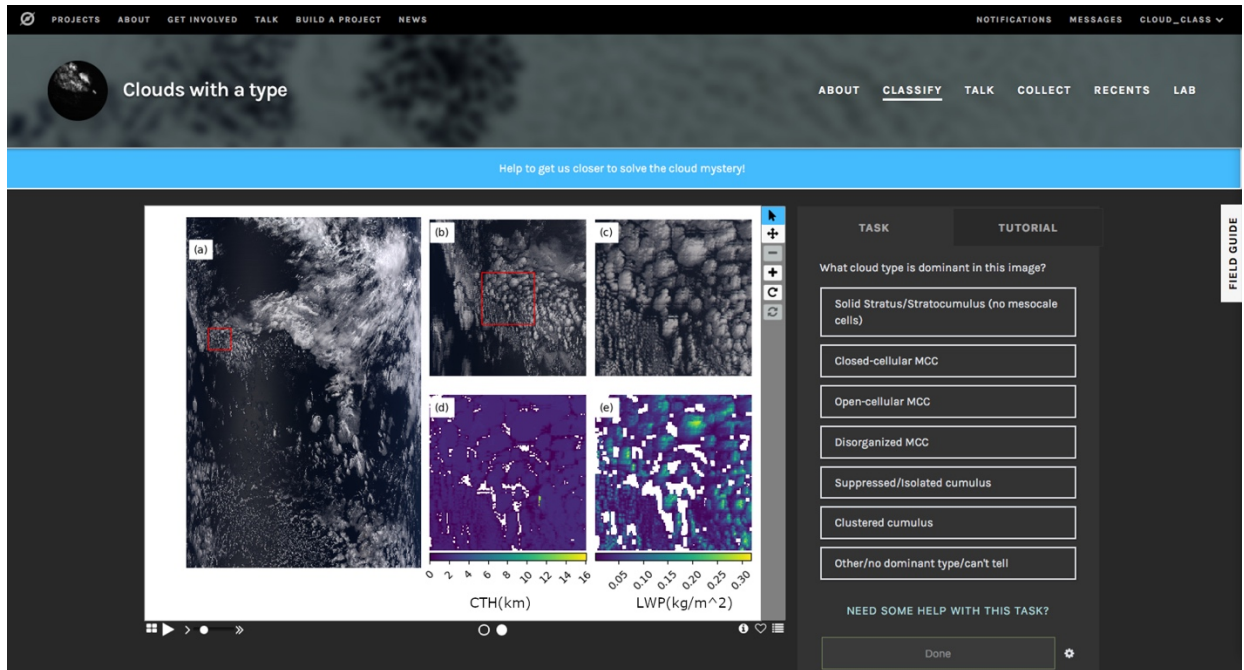
305 We have developed a working deep neural network model to automatically classify cloudy
306 scenes into six mesoscale morphology types. Initial test run results showed promising results
307 for the Southeast Pacific and Northwest Atlantic. Using the tool, we plan to extend the dataset
308 and create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

311
312



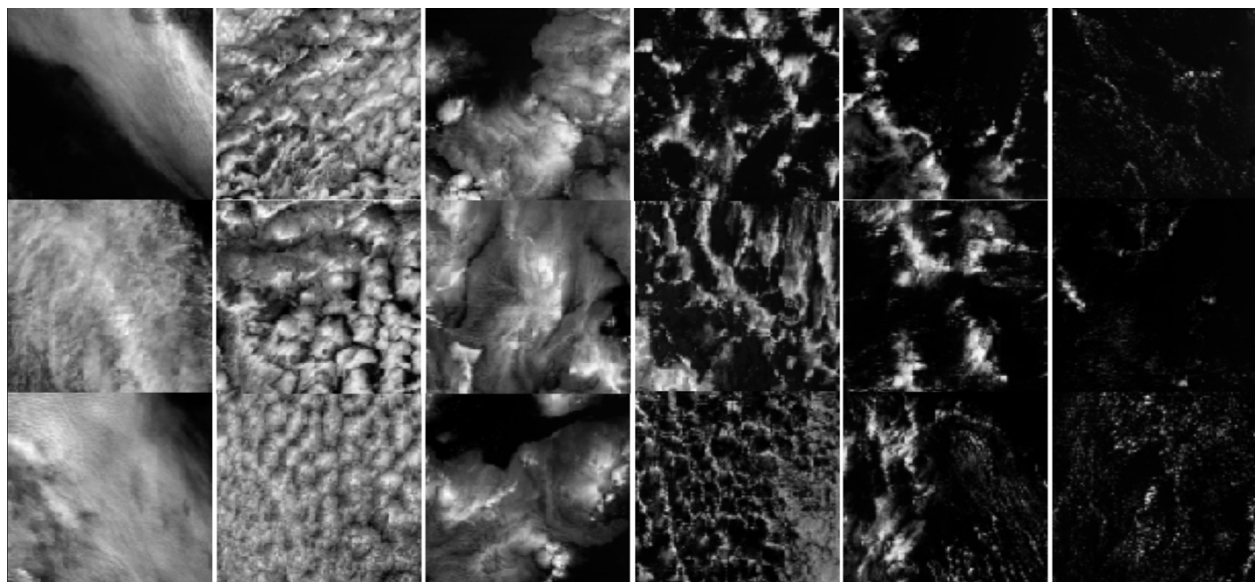
313
314 Figure 1: A full disk image of GOES-16 on Aug 6, 2018 and six scenes of MODIS images at smaller
315 scales representing different morphology types at corresponding locations in the GOES image.
316 Except scene 1, all scenes are from the same day. Scene 1 is from a different day because there
317 was no representative stratus scenes on this day in the Southeast Pacific region.

318
319



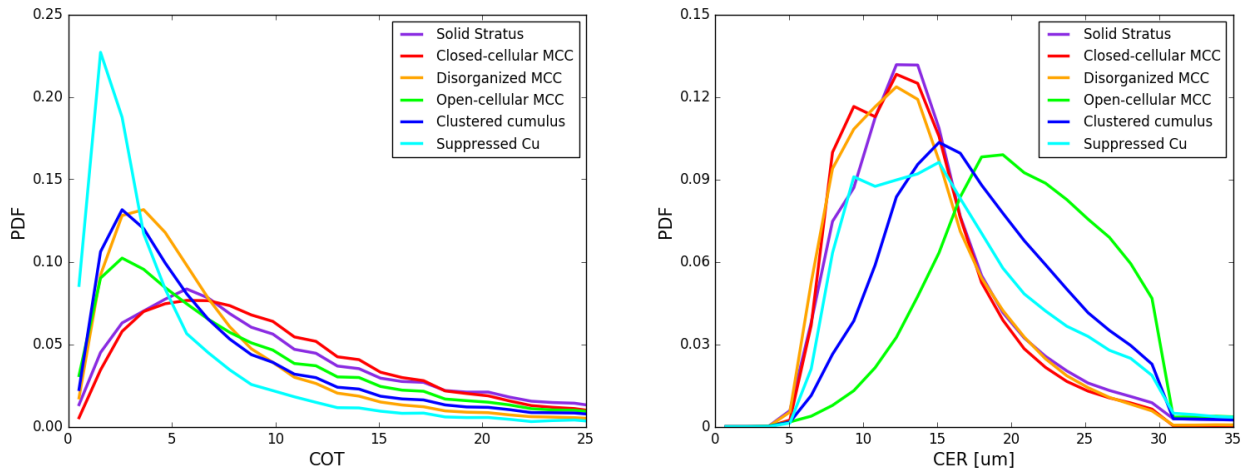
320
 321 Figure 2: the Zooniverse interface for manual labelling. The center image is made up of five
 322 panels. Panel a shows the full granule (usually 2030x1350 pixels) true color image for large
 323 context. Panel b shows a portion of the granule immediately surrounding the scene to be
 324 labelled, outlined by the red square. Panel c shows the visible scene image while panels d and e
 325 show the cloud top height and LWP fields in the scene to be labelled. The panels to the right of
 326 the center image show labelling choices. The tutorial document is available by clicking on the
 327 'FIELD GUIDE' tab on the right side. Additional options for scenes with heavily mixed types,
 328 scenes with sea ice, or scenes with other issues are found in the 'other' menu. The image is a
 329 screenshot of our Zooniverse project.

330
 331

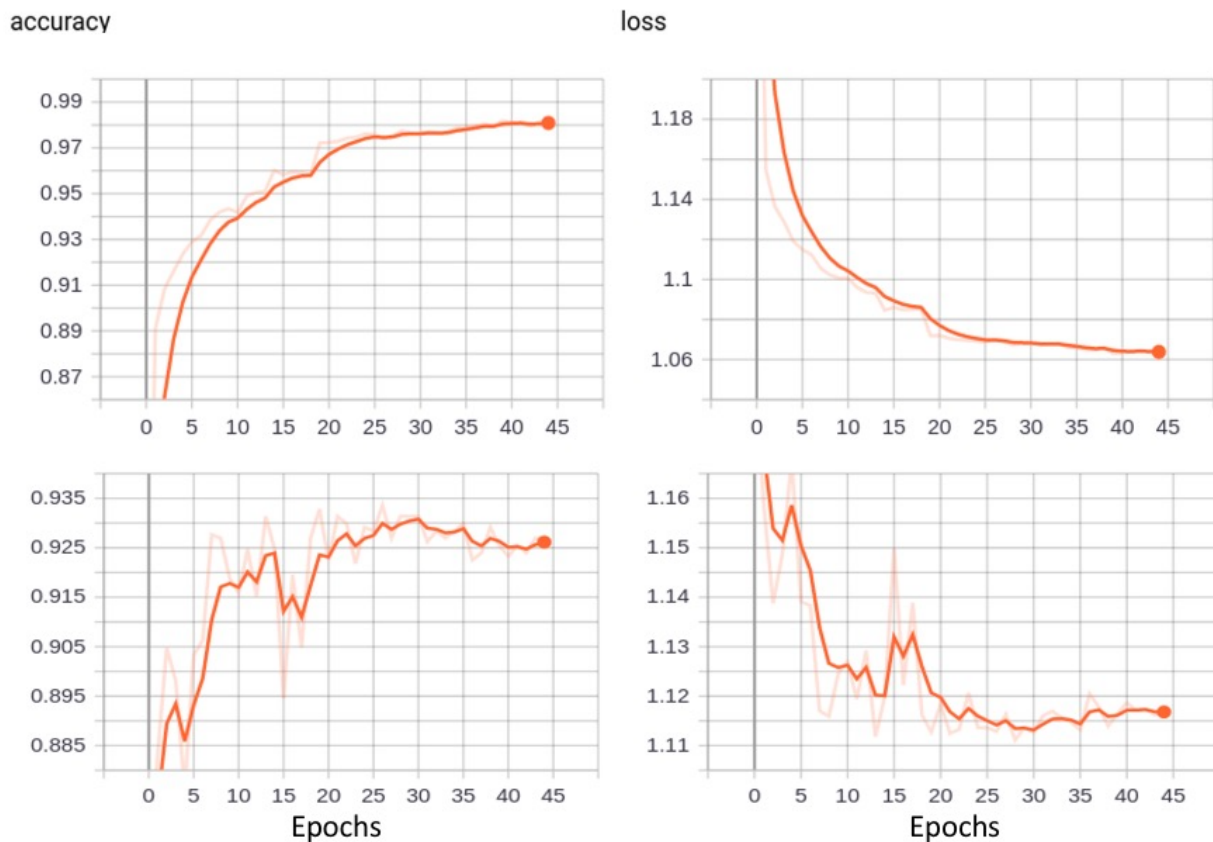


332
 333

334 Figure 3: Example scenes of MODIS single channel images for the six different types. From left
 335 to right: stratus, closed cellular, disorganized cellular, open cellular, clustered cumulus, and
 336 suppressed cumulus types. Images taken by the NASA MODIS.
 337

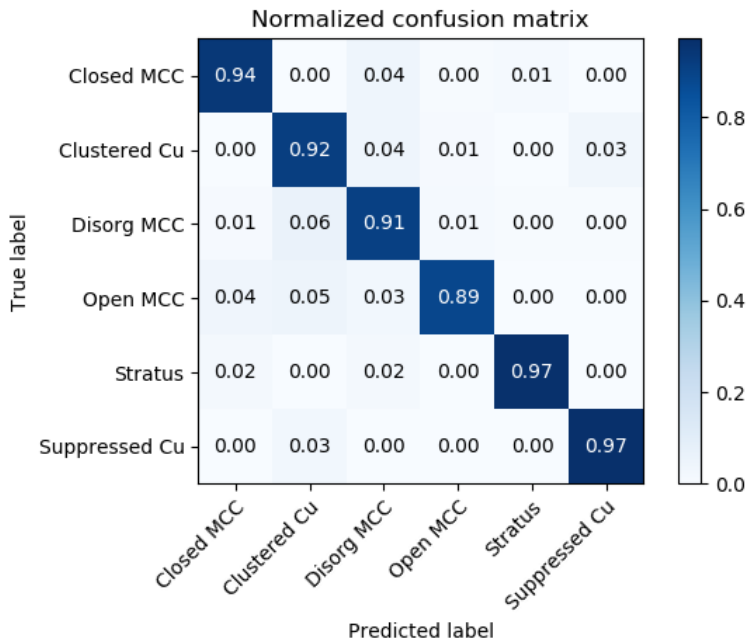


338
 339
 340 Figure 4: PDFs of cloud optical depth and cloud effective radius for six morphology types. We
 341 randomly selected 1000 samples for each type and mean distributions are shown here.
 342 Significant overlaps are observed for PDFs of both variables among different morphology types.
 343
 344

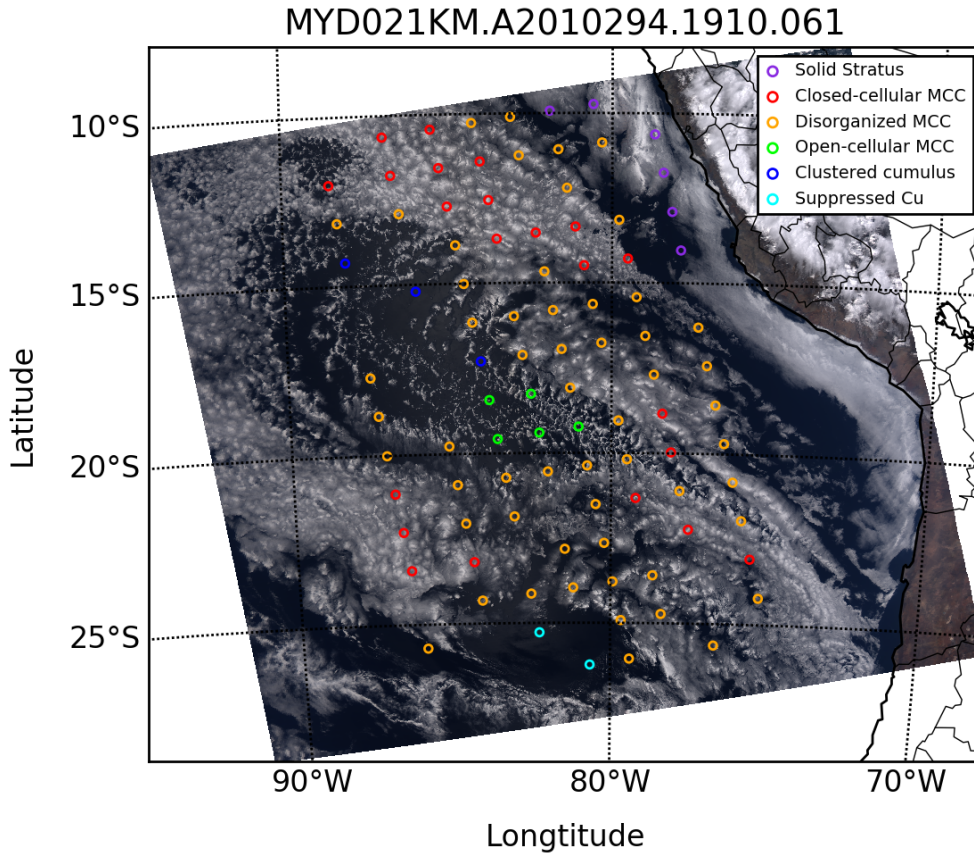


345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000

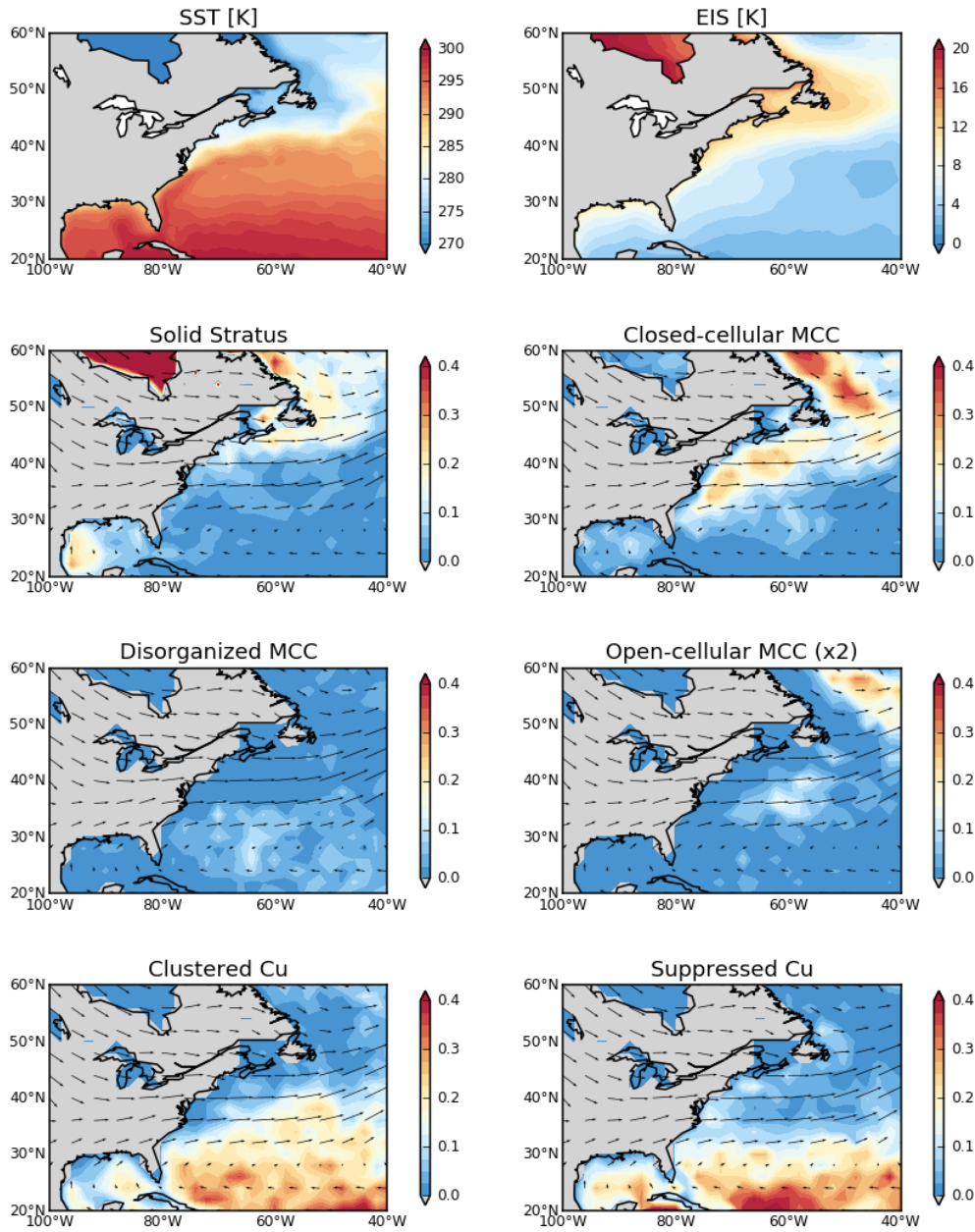
346 Figure 5: Training (upper two panels) and validation (lower ones) accuracy and loss trajectories.
 347 By around epoch 30, the validation accuracy peaks while validation loss bottoms out and the
 348 training loss and accuracy asymptotically reach their minimum and maximum, respectively,
 349 which indicates further training may be overfitting the model.
 350



351
 352 Figure 6: Confusion matrix of the model predictions on test data.

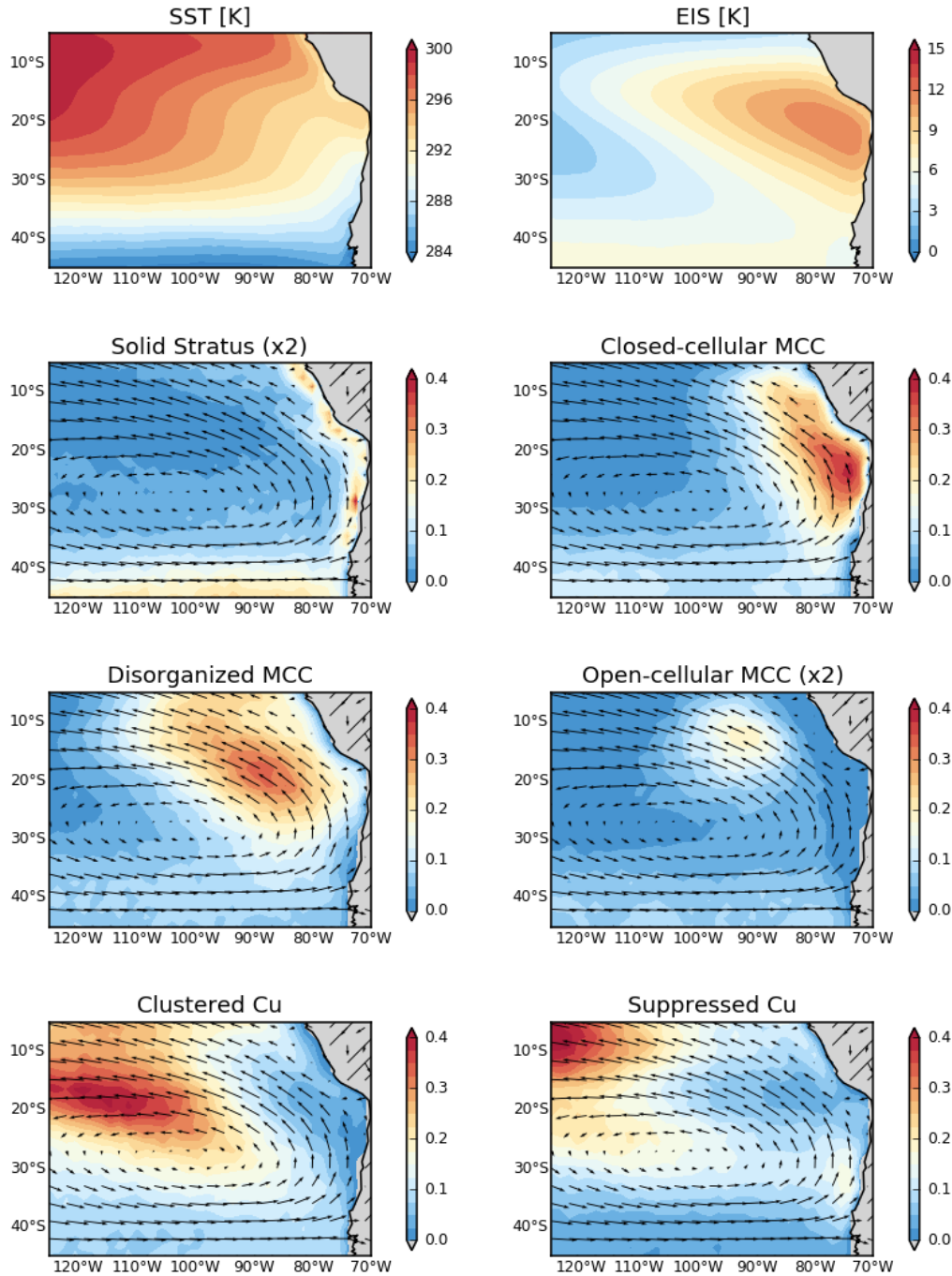


353
 354 Figure 7: An example granule illustrating the results of the classification algorithm. This is quite
 355 a complex granule with different morphology types mixed together. The left and right margins
 356 are not classified because current algorithm filters out scenes whose sensor viewing zenith
 357 angles are greater than 45 degrees. The image is taken by NASA MODIS.
 358



359
 360
 361
 362
 363
 364
 365
 366

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.



367
 368 Figure 9: Frequency distributions of various morphology types obtained from the classification
 369 algorithm in the subtropical eastern Pacific off the coast of South America for the period 2003-
 370 2018. The top two panels show the SST and EIS climatology from MERRA-2 for the same period.
 371 Note the doubling of scale on the stratus and open-cellular types.

372

373 **6. Author Contribution**

374 T. Y. implemented the method to train the network model. H. S., J. M., and T.Y. prepared the
375 training data. All co-authors contributed to compiling the training dataset. T. Y. wrote the
376 manuscript with contributions from all co-authors.
377

378 **7. Reference**

- 379 Agee, E. M., & Dowell, K. E. (1974). Observational Studies of Mesoscale Cellular Convection.
380 *Journal of Applied Meteorology*, 13(1), 46–53. [https://doi.org/10.1175/1520-](https://doi.org/10.1175/1520-0450(1974)013<0046:OSOMCC>2.0.CO;2)
381 [0450\(1974\)013<0046:OSOMCC>2.0.CO;2](https://doi.org/10.1175/1520-0450(1974)013<0046:OSOMCC>2.0.CO;2)
- 382 Atkinson, B. W., & Zhang, W. J. (1996). Mesoscale shallow convection in the atmosphere.
383 *Reviews Of Geophysics*, 34(4), 403–431. <https://doi.org/10.1029/96RG02623>
- 384 Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, & Li Fei-Fei. (2009). ImageNet: A large-scale
385 hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern*
386 *Recognition* (pp. 248–255). <https://doi.org/10.1109/CVPR.2009.5206848>
- 387 Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The
388 Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2).
389 *Journal of Climate*, 30(14), 5419–5454. <https://doi.org/10.1175/JCLI-D-16-0758.1>
- 390 LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
391 <https://doi.org/10.1038/nature14539>
- 392 McCoy, I. L., Wood, R., & Fletcher, J. K. (2017). Identifying Meteorological Controls on Open and
393 Closed Mesoscale Cellular Convection Associated with Marine Cold Air Outbreaks.
394 *Journal of Geophysical Research: Atmospheres*, 122(21), 11,678–11,702.
395 <https://doi.org/10.1002/2017JD027031>

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

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

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

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

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

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

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

417 Wood, Robert, & Hartmann, D. L. (2006). Spatial variability of liquid water path in marine low
418 cloud: The importance of mesoscale cellular convection. *Journal Of Climate*, 19(9),
419 1748–1764.

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

422 Yuan, T., & Oreopoulos, L. (2013). On the global character of overlap between low and high
423 clouds. *Geophysical Research Letters*, 40(19), 5320–5326.

424 <https://doi.org/10.1002/grl.50871>

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451