

1 **Capturing the diversity of mesoscale trade wind cumuli**
2 **using complementary approaches from self-supervised**
3 **deep learning**

4 **Dwaipayan Chatterjee¹, Sabrina Schnitt¹, Paula Bigalke¹, Claudia**
5 **Acquistapace¹, Susanne Crewell¹**

6 ¹Institute for Geophysics and Meteorology, University of Cologne, Cologne, Germany

7 **Key Points:**

- 8 • Mesoscale cloud organization can be taxonomized by a two-step deep learning ap-
9 proach in the feature space continuum
10 • Comparison with human-annotated labels reveals the need to include uncertainty
11 estimates of the human-derived classification
12 • Our approach can describe the temporal transition from human-labeled sugar to
13 flower regimes

Corresponding author: Dwaipayan Chatterjee, dchatter@uni-koeln.de

Abstract

At the mesoscale, trade wind clouds organize with a wide variety of spatial arrangements, which influences their effect on Earth's energy budget. Past studies used high-resolution satellite measurements and clustering/labeling techniques to classify trade wind clouds into distinct classes. However, these methods only capture a part of the observed organization variability. This work proposes an integrated framework using a continuous followed by discrete self-supervised deep learning approach based on cloud optical depth from geostationary satellite measurements. The neural network learns the semantics of cloud system structure and distribution, verified through visualizations of different layers. Our analysis compares classes defined by human labels with machine-identified classes, aiming to address the uncertainties and limitations of both approaches. Additionally, we illustrate a case study of sugar-to-flower transitions, a novel aspect not covered by existing methods.

Plain Language Summary

Clouds are a fundamental player affecting our planet's energy balance, making their accurate representation crucial in climate models. One open question is how they organize on a scale of a few 100 km (mesoscale) within the trade wind region. Satellite observations can help to categorize these clouds, but previous methods had limitations in capturing the full range of cloud arrangements and transitions between different cloud forms. We have introduced a novel approach that utilizes machine learning and geostationary satellite data to address this issue. Our machine learning model autonomously learns to recognize various cloud patterns and distributions. We conducted a comparative analysis between the categories generated by the machine and those identified by human experts to understand the strengths and weaknesses of both methods. Additionally, we explore a case study where clouds undergo a transformation, changing from a structure resembling sugar to one resembling flowers. This particular transformation was found difficult to capture with numerical simulation before. Our approach successfully captures the transition in the machine-learned feature space. Overall, the new approach can help to better understand cloud evolution, which is crucial for improving climate models and predicting how cloud behavior may change in a changing climate.

1 Introduction

Shallow convective clouds are small in individual extent but cover large areas of the tropical oceans, appearing as distinct cloud fields. Due to their radiative and precipitating properties, their representation in climate models is crucial for understanding the current large inter-model spread in predicted cloud feedback and climate sensitivity (Bony & Dufresne, 2005; Nuijens & Siebesma, 2019; Vogel et al., 2022). The EUREC⁴A field campaign (Bony et al., 2017; Stevens et al., 2021), which took place in the North Atlantic Trade (NAT) region around Barbados, aimed at investigating the interplay between clouds, convection, and circulation by deploying a large variety of observations between January and February 2020.

While shallow convection was long seen to produce randomly scattered individual clouds, further understanding has been gathered on the importance of cloud field organization for precipitation (Rauber et al., 2007; Radtke et al., 2022) through cold pool formation and rain evaporation (Seifert et al., 2015; Vogel et al., 2021). Among others, open research questions concern a detailed quantification of the role played by shallow mesoscale cloud organizations in controlling cloud amounts and their radiative response in the trades (Bony et al., 2015; Tomassini et al., 2015; Vogel et al., 2020; Vial et al., 2017).

Introducing four shallow convective organizations (*Sugar*, *Gravel*, *Flower*, *Fish*), with common occurrences on meso-beta (20 to 200 km) and meso-alpha (200 to 2,000

63 km) scales, Stevens et al. (2020) rely on human-labeled visible satellite images in the NAT
64 region. Sugar clouds consist of small, scattered clouds with a limited vertical extent, while
65 gravel clouds exhibit organized lines or arcs resembling cell-like patterns. Fish clouds dis-
66 play a network resembling fishbones with distinct cloud-free spaces, and flowers repre-
67 sent larger, stratiform cloud structures forming dispersed closed cells. These patterns
68 vary in net cloud radiative feedback (Bony et al., 2020); and, when classified by a deep
69 neural network trained on human-labeled scenes (Rasp et al., 2020), display fundamen-
70 tal differences in cloud fraction and environmental conditions (Schulz et al., 2021). The
71 four patterns exhibit a daily cycle (Vial et al., 2021) and transitions, e.g., from sugar to
72 flower, have been studied in Large-Eddy-Simulation (LES) to identify the governing pro-
73 cesses (Narenpitak et al., 2021, 2023; Dauhut et al., 2023).

74 Yet, imposing four distinct classes on the diversity of the observed organization does
75 not cover the intermediate cloud patterns or transient states, as highlighted by the LES
76 studies. Hence, some dynamic processes important for climate feedback may be ignored
77 or neglected. Also, to our knowledge, there is no description of one of such transitions
78 among different cloud regimes purely using observations. Furthermore, most of the re-
79 cent studies trying to quantify the labeled well-organized systems find that these four
80 cloud systems occur only around 50% over NAT (Janssens et al., 2021; Schulz et al., 2021;
81 Vial et al., 2021) and have some ambiguities in agreement from the labelers' side (Schulz,
82 2022). Therefore, to handle such complexity, our first objective is to develop simplified,
83 streamlined representations to effectively understand and capture the entire cloud spec-
84 trum's organizational relationships.

85 There are several different possibilities for ordering the variability of mesoscale cloud
86 systems, such as Janssens et al. (2021) who introduced a set of selected metric spaces
87 for arranging the cloud systems using object-based, scale-based, and retrieved physical-
88 based statistical properties. Utilizing the metric scores and a k-means algorithm, they
89 observe that human-defined classes have better separation starting at seven clusters. Denby
90 (2020) demonstrates that unsupervised neural network models, which involved some hu-
91 man decisions in the learning stage, can be used to distinguish mainly ten different types
92 of cloud organization and their associated radiative properties. In this work, we do not
93 aim to favor any of the presented metrics but rather search for new information purely
94 based on their organizational aspects, minimizing human intervention. Therefore, we aim
95 to identify optimal distinct classes of cloud organizations representing the full spectrum
96 and further compare them with human-identified labels.

97 Based on GOES-16 E cloud optical depth (COD) images (Sec. 2), Section 3 pro-
98 poses a two-step self-supervised deep learning approach to study shallow convection in
99 a continuous feature space, characterizing the entire diversity of occurring organizations.
100 Further, an optimized discretization of the continuous space is developed to derive a fi-
101 nite set of classes representative of the continuous spectrum. The representations and
102 their characteristics are investigated in Section 4.1, and we conduct a proof-of-concept
103 study in Section 4.2 to explore the extent of agreement between human-annotated cloud
104 organizations and machine-identified classes. Additionally, we investigate in Section 5
105 how this approach can be used as a tool to study transitions between different organ-
106 ization patterns.

107 2 Satellite dataset

108 We use COD retrieved from GOES-16 E Advanced Baseline Imager (Schmit et al.,
109 2005) by the daytime cloud optical and microphysical properties algorithm (DCOMP)
110 (Walther & Heidinger, 2012) at 2 km horizontal resolution and 10 – 15 minutes tempo-
111 ral resolution. Our domain in NAT (5 – 20° N and 40 – 60° W) is similar to domains used
112 in past studies (Bony et al., 2020; Schulz et al., 2021). The regional climate defines De-
113 cember to May as dry and June to November as wet seasons (Stevens et al., 2016). While

114 most of the studies focus on dry season shallow convections only, we include some con-
 115 tributions from the wet season by selecting the time period from November to April 2017
 116 – 2021. The purpose of choosing convective occurrences from the wet season is to see how
 117 they influence our approach.

118 COD represents the radiative properties of the cloud in the visible range, and its
 119 retrieval from DCOMP tackles the aleatoric uncertainties from the atmosphere and sur-
 120 face robustly. For example, the uncertainty associated with COD retrieval remains be-
 121 low 10% for all ranges in water clouds (see Figure 4 in Walther and Heidinger (2012)).
 122 Therefore, we exploit the COD parameter to characterize the cloud system spatio-temporal
 123 variability. Note that some fine-scale cloud systems, such as sugar and gravel, also con-
 124 tributing to the variability of mesoscale beta clouds in regional climate systems, may not
 125 be fully resolved with the spatial resolution of this product.

126 Representation learning, also known as feature learning, is a specialized field within
 127 machine learning that focuses on extracting meaningful features of a given dataset. To
 128 better represent the mesoscale cloud distributions, we use six images per timestamp, in-
 129 cluding an additional fixed image over the Barbados domain (see S1). Note that the Bar-
 130 bados domain enables comparison with ground-based measurements in future studies.
 131 To have an adequate spatial scale of typical occurring cloud fields over NAT (as discussed
 132 in Section 1), we use 256 x 256 pixels (roughly 512 square km) as also found in Muller
 133 and Held (2012). We exclude crops affected by glint or poor retrieval quality using the
 134 respective data flags. Time stamps are limited to 9 am - 3 pm Barbados local time to
 135 avoid sun glinting. We utilize land class information to avoid land convection and ver-
 136 ify whether 0.98th fraction of random crops belong to the ocean, accepting satellite crops
 137 with islands over NAT and excluding those over the northeast South American conti-
 138 nent. Finally, to mitigate uncertainties at high COD from DCOMP retrieval, COD val-
 139 ues above a threshold of 50, already indicating deep clouds, are clipped to 50. This re-
 140 sults in a sample size of 51,000 satellite images.

141 For further analysis, we make use of hourly ERA-5 (Hersbach et al., 2020) large-
 142 scale environmental parameters (horizontal and vertical wind speed, relative humidity)
 143 and cloud fraction at a spatial resolution of 0.25°. Hourly cloud amount for four verti-
 144 cal ranges (surface-700 hPa, 700 hPa-500 hPa, 500 hPa-300 hPa, 300 hPa-tropopause)
 145 is used from the Clouds and Earth’s Radiant Energy System fourth edition (CERES, Edi-
 146 tion - 4A) (Wielicki et al., 1996), characterized by a spatial resolution of 1°.

147 **3 Methods**

148 First, we develop a neural network (N1) that learns to sort the cloud organizations
 149 based on the similarity of their visual features, which we call a continuous approach in
 150 this work. The purpose is to let the network identify the structural similarities in the
 151 cloud systems and map the learned visual features in the 384-dimensional feature space.
 152 We use the software package DINO from Facebook Artificial Intelligence Research (FAIR)
 153 (Caron et al., 2021) based on PyTorch (Paszke et al., 2019) and the open-source VISSL
 154 computer vision library (Goyal et al., 2021) to adapt the network to our requirements.
 155 As a backbone neural architecture to process images, we use Vision Transformer (ViT),
 156 which has a sequence of self-attention (Vaswani et al., 2023) and feed-forward layers (Bebis
 157 & Georgiopoulos, 1994) paralleled with skip connections. This setup helps to identify
 158 long-range spatial dependencies by learning relevant information in the image (Khan et
 159 al., 2022). To focus on the structural similarities of the cloud system, every epoch, we
 160 opt for two random global crops with a 0.75 fraction (192 x 192 pixels) of the parent satel-
 161 lite image. As the largely overlapping global-crop pair has very similar cloud structures,
 162 the network learns their essential features and puts the pair closer to each other in the
 163 high-dimensional feature space. More details are given in S2.

164 After obtaining the continuously sorted representation of cloud systems (see Fig.
 165 1.a), we intend to find optimal boundary conditions within the sorted order and, based
 166 on it, train a second neural network (N2) to discretize it. As a first step, we reduce the
 167 384-dimension features of the satellite images obtained from N1 to two dimensions us-
 168 ing the well-established t-distributed Stochastic Neighbor Embedding (tSNE) algorithm
 169 (van der Maaten & Hinton, 2008). tSNE tries to preserve the relative local position be-
 170 tween features and the overall global structure of the feature distributions while map-
 171 ping on a reduced two-dimensional space. On this 2-dimensional representation space,
 172 we apply a set of three statistical approaches, namely metric scores of distortion, silhou-
 173 ette (Rousseeuw, 1987), and Calinski-Harabasz (Caliński & Harabasz, 1974) to identify
 174 the possible number of optimal classes into which the given features could be clustered.
 175 Schubert (2023) suggests taking a collective inference from these three methods to best
 176 fit the spherical k-means clustering algorithm used during the training of N2. Supple-
 177 ment 3 illustrates how the three metrics point to an optimal clustering into seven classes.

178 N2 from Chatterjee et al. (2023) learns to put each satellite image in one of the seven
 179 classes as it progressively improves the feature space’s clustering, minimizing the cross
 180 entropy between two global random crops (192 x 192) from the parent satellite image.
 181 Here, the main difference from N1 is that additional augmented image versions (random
 182 flipping and noise addition by random Gaussian blur) of global random crops (see Fig.
 183 S2.2.b) are included. Augmentations try to provide auxiliary support to the network’s
 184 generalizability and better capture the differences in diversity of the shallow cloud sys-
 185 tems (Nie et al., 2021; Paletta et al., 2023). After obtaining the label of each satellite
 186 image, we transfer the assigned class to the continuous representation space, which proves
 187 helpful because N1 has learned the sorting arrangement of keeping similar cloud systems
 188 closer. Therefore, it helps to visualize how each cluster with distinct characteristics can
 189 form a separate local region. The N2 feature space is i) more sparse than N1 (see S2 for
 190 explanation) and ii) arranged by closeness to the centroids, which, unlike N1, may not
 191 be ideal for representing smooth transitions of cloud systems.

192 4 Results

193 4.1 Continuous and discrete representations

194 To investigate how the satellite images arrange themselves in the feature space of
 195 N1, we first study the high-dimensional feature space and assess the arrangement of di-
 196 verse cloud systems inside it. We reduce the feature dimensions to a 2D space to visu-
 197 alize the continuum using the tSNE algorithm (described in Section 3). Different cloud
 198 organizations can be identified in different areas of the 2D space (Fig. 1.a). Going an-
 199 ticlockwise from the top, arch-shaped cloud systems lie in the top-left, followed by flower-
 200 type distributions on the left side of the 2D feature space. Close to the flowers in the bottom-
 201 left are the flowers spreading out into stratocumulus. Note that while modeling stud-
 202 ies suffer from capturing the transition of stratocumulus to cumulus (Sarkar et al., 2020),
 203 these cloud regimes are adjacent to one another in the 2D representation.

204 The bottom part of the feature space contains long bony skeletons, i.e., fish-type
 205 cloud systems, and the bottom-right corner shows an extended part of fish-type cloud
 206 organizations delineated by unusually large cloud-free regions. The top-right region of
 207 the 2D space is a collection of deep convective cells. These primarily occur in the month
 208 of November. Arc-shaped cloud systems appear on the left and top-left of the 2-D fea-
 209 ture space. Vogel et al. (2021) suggest that the horizontal structure of mesoscale arcs
 210 is intrinsically linked to gravel, flowers, and fish. In sequence, Figure 1a shows a contin-
 211 uous link in the spatial arrangement of cloud systems rather than the distinct classes.
 212 Additionally, in S4, we investigate how N1 is sensitive to different visual features of cloud
 213 organizations and find that the network pays attention to specific patterns in cloud or-

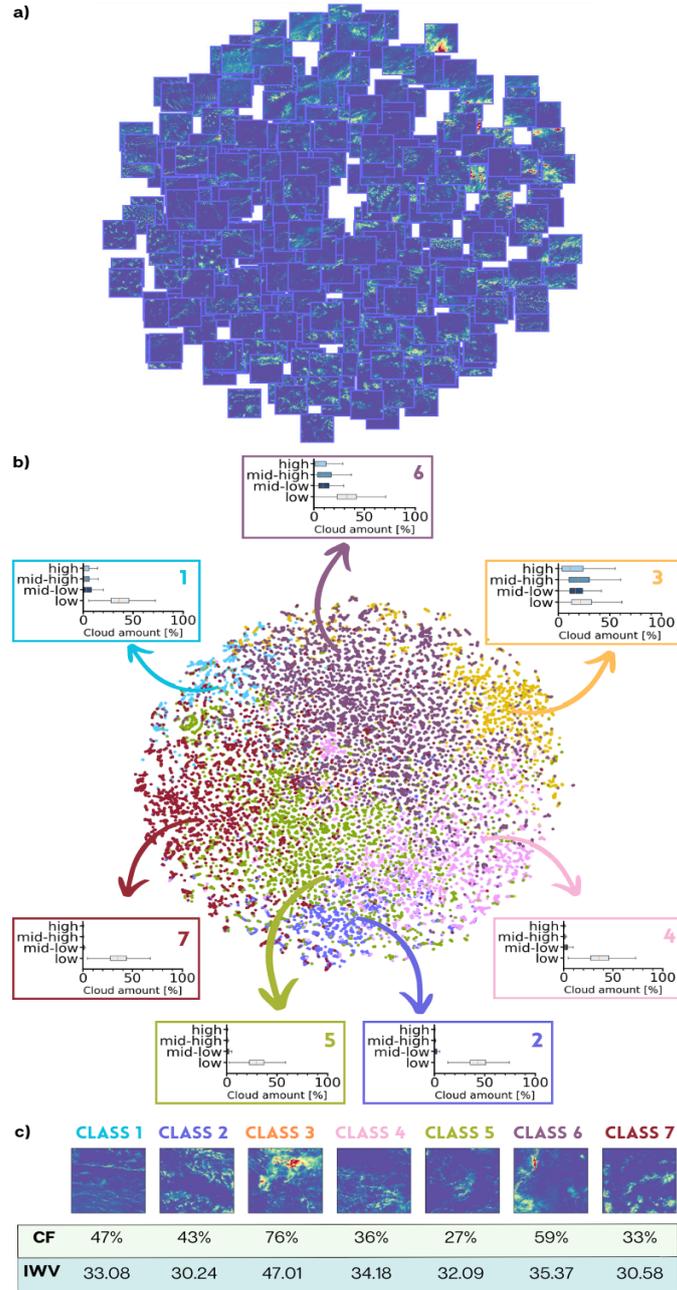


Figure 1. a) Visualization of four hundred randomly selected 256×256 satellite images arranged in the dimensionally reduced 2D continuous feature space where the closeness of one satellite image to another is learned by N1. b) Optimized classification learned by N2 provides labels overlaid on the continuous feature space to show the clustering performance. Each class shows low, mid-low, mid-high, and high cloud amounts (%) obtained from the CERES hourly data set. c) Centroid COD images belonging to seven clusters as identified by the discrete neural network (N2). The table shows per class mean of cloud fraction (CF, %) from GOES retrieval and integrated water vapor (IWV, kgm^{-2}) from ERA-5.

214 ganization, such as deep convective semantics, adjacent thin convection around deep con-
 215 vection, and clear sky features.

216 Using N2, each of the images can be attributed to one of the seven classes (refer
 217 to Section 3), revealing distinct spaces within the 2D continuous representation space
 218 (Fig. 1.b). To help investigate how well the seven classes separate, they are evaluated
 219 using cloud amount at four different height levels from CERES data. This analysis, on
 220 the one hand, reflects how each class differs from the others, and on the other hand, it
 221 reasons for the underlying closeness of each class with neighbor classes in the feature space.
 222 The difference between the seven clusters is especially evident when looking at their cen-
 223 troid images (Fig. 1.c).

224 Deep convective class three has by far the highest cloud fraction of 76% and a third
 225 more water vapor amount (47.0 kgm^{-2}) than all other classes (mean = 32.5 kgm^{-2}). Neigh-
 226 boring class six (in feature space) includes less frequent higher-level clouds and has a re-
 227 duced CF of 59% compared to class three. All other classes are dominated by low-level
 228 clouds with lower than 50% CF. Classes one and four (neighbor to class six) still have
 229 some mid to high-level cloud amount (below 10%). Class one can be interpreted as rep-
 230 resenting arch-shaped cloud systems, and four resembles the fish class with a more open
 231 sky (also shown by reduction in CF). Classes two, five, and seven, being close in the 2D
 232 feature space, have similar cloud vertical distributions and IWV ranging from 30 to 32
 233 kgm^{-2} ; however, their organization is very different, as illustrated by the centroids (Fig.
 234 1.c) and mean CFs (43%, 27%, and 33%, respectively). Class two primarily comprises
 235 shallow cloud cover, corresponding to cloud systems resembling fish-type formations. Class
 236 five has the lowest cloud fraction and is an intermediary class type between classes two
 237 and seven. Finally, class seven has a presence of low cloud amounts and negligible mid
 238 to higher cloud amounts, which visually resembles flower-type cloud distributions. There-
 239 fore, discretizing the continuous feature space helps us visually find three main classes
 240 (one, two, and seven) frequently resembling features identified by humans, i.e., sugar,
 241 fish, and flower, respectively. However, it also shows the remaining diversity and their
 242 characteristics in a cohesive approach.

243 4.2 Machine versus human labels

244 While we checked for visual correspondence and class-wise characteristics in Sec-
 245 tion 4.1, we now aim to quantify how human labels compare to the machine’s seven clus-
 246 ters. We use the seven previously identified cluster boundaries and cloud system posi-
 247 tions in the continuous feature space (N1 + N2 together defined as “framework” from
 248 now on) and the dataset by Schulz (2022), providing human labels with an agreement
 249 score ranging between 0 and 100%.

250 For each timestamp where at least one of the four patterns was identified within
 251 our domain, we select a 256 x 256-pixel satellite image centered over the area of high-
 252 est human agreement. In this way, we ensure the best possible intercomparison. Apply-
 253 ing the pre-processing (as in Section 2) leaves us with 52 samples of human-labeled satel-
 254 lite images (fish: 19.3%, gravel: 26.9%, flower: 28.8%, sugar: 25.0%). Note that the best
 255 and worst cloud organization agreements with this procedure are 91% and 7%, respec-
 256 tively. Finally, we get the feature vectors of the images corresponding to the human sam-
 257 ples from N1 and the machine-identified labels from N2.

258 The framework classifies 40% flower-labeled cloud systems in class seven (see the
 259 hit rate along each class in Fig. 2.a) while sugar-labeled cloud systems are 31% classi-
 260 fied in class one and 20% in class four. For class four, note sugar’s low agreement in Fig.
 261 2.b. Gravel has a total of 44% representation in classes one and five, whereas fish an-
 262 notated labels are allocated 30% in class two and 20% each in classes four and five. Fur-
 263 ther, looking at example images visually (Fig. 2.a), in contrast to images with high hu-
 264 man agreement, it is evident that those with lower agreement significantly deviate from

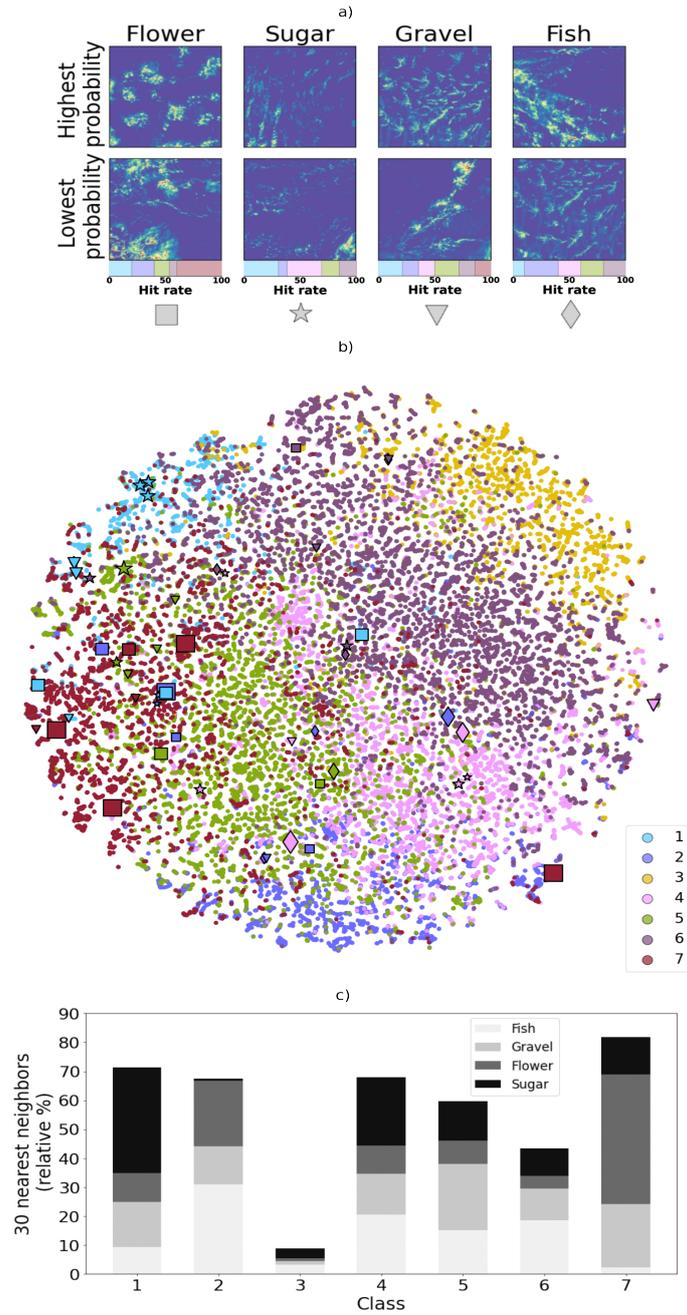


Figure 2. a) For better visualization and reference purposes of human labels, each column shows 256 x 256 COD images belonging to a certain class marked with the highest and lowest human agreement displayed along the two rows. Below, the images along each column show the proportional machine-predicted class for human labels. b) Continuous feature space colored with different classes (1-7) in the background, along with Human labels (fish, sugar, flower, gravel) in the foreground. The level of human agreement on the identified patterns is indicated by symbol size. c) Relative occurrence of 30 nearest neighbors to human-labeled fish, gravel, flower, and sugar along the seven machine-labeled classes.

265 the recognized definitions (as given in Stevens et al. (2020)) of sugar, gravel, flower, and
 266 fish cloud structures.

267 Within the 2D feature space (Fig. 2.b), flowers detected with high probability mostly
 268 occur in areas of class seven, which was already well reflected in the centroids. Follow-
 269 ing a similar agreement is sugar (street-type cloud systems), which can be found in ar-
 270 eas of class one. However, 38% of sugar samples, with a low agreement, lies in classes
 271 four and five, which are extended fish and flower type classes (Section 4.1). Thus, even
 272 though these samples reside in those regions of the feature space, their confidence is less
 273 than 25%. Rightly, no human-labeled samples are found in class three, which predom-
 274 inantly comprise deep convective cells. For the gravel pattern, 21% samples belong to
 275 class six (Fig. 2.b)) and exhibit minimal human confidence; in contrast, the rest from
 276 the gravel class are positioned between classes one and seven, suggesting that gravel cloud
 277 cell sizes fall between sugar and flower. Finally, the fish class exhibits relatively higher
 278 confidence in human labels, aligning well with the feature space characteristics, and lies
 279 in class two (fish) and four (extended fish).

280 To compensate for the limited number of human label samples, we analyze the 30
 281 nearest satellite images to each human label as identified by N1 (Fig. 2.c). This anal-
 282 ysis aims to show the generalization capacity of our approach and further enhance our
 283 understanding of the connection between organizations. The majority of neighbors in
 284 human-identified fish-type cloud systems (more than 50%) belong to machine-identified
 285 classes two and four, representing fish and extended fish-type cloud structures with large
 286 cloud-free regions. The gravel regime includes members of all classes, with notable con-
 287 tributions from classes one, five, and seven, which exhibit cloud cell characteristics sim-
 288 ilar to gravel systems. One of the reasons for the wider spread of neighbors might be due
 289 to the lower human agreement of the images labeled as gravel (75% of gravel-labeled sam-
 290 ples had agreement less than 0.25). In contrast, the flower regime mainly belongs to class
 291 seven (46 %), further aligning with the high confidence of human labels. Regarding sugar-
 292 type cloud systems, 37 % of the neighbors fall into class one, while those with low hu-
 293 man agreement are scattered across the remaining classes. Therefore, we find that machine-
 294 labeled classes encompass the human-labeled ones, especially for sugar, flower, and fish,
 295 but not so clearly for gravel.

296 Comparing human labels with their nearest neighbors shows that the framework
 297 provides more objective freedom and improves our confidence about the feature vectors
 298 allocated to images corresponding to human samples. It also shows the uncertainty as-
 299 sociated with less agreed-upon human labels. Further, in S5, using ERA-5 large-scale
 300 environmental variables and cloud physical properties, we demonstrate that both the neigh-
 301 bors and the human crops share a similar, homogeneous distribution of physical prop-
 302 erties.

303 5 Transitions

304 To showcase an application that highlights the strengths and weaknesses of the pre-
 305 sented framework, we explore the "sugar" to "flower" (S2F) cloud system transition on
 306 February 2, 2020. Using LES, Narenpitak et al. (2021) showed a strengthening of large-
 307 scale upward wind motion and an increase in total water path and optical depth as the
 308 transformation develops towards the flower. Here, we look at how the transition in COD
 309 is represented in the feature space. For example, where do the representations of tran-
 310 sitions lie in the feature space? How smooth is the transition in the feature space?

311 Covering the spatio-temporal developments, 47 COD images were collected (after
 312 applying quality filter checks (see Section 2)), centered at 12.5° N, 50° W. They cover
 313 the time from 10:50 to 19:20 UTC, with a gap between 17:00 to 18:00 UTC likely caused
 314 by local sun glint. We ingest the available samples into the trained framework, collect

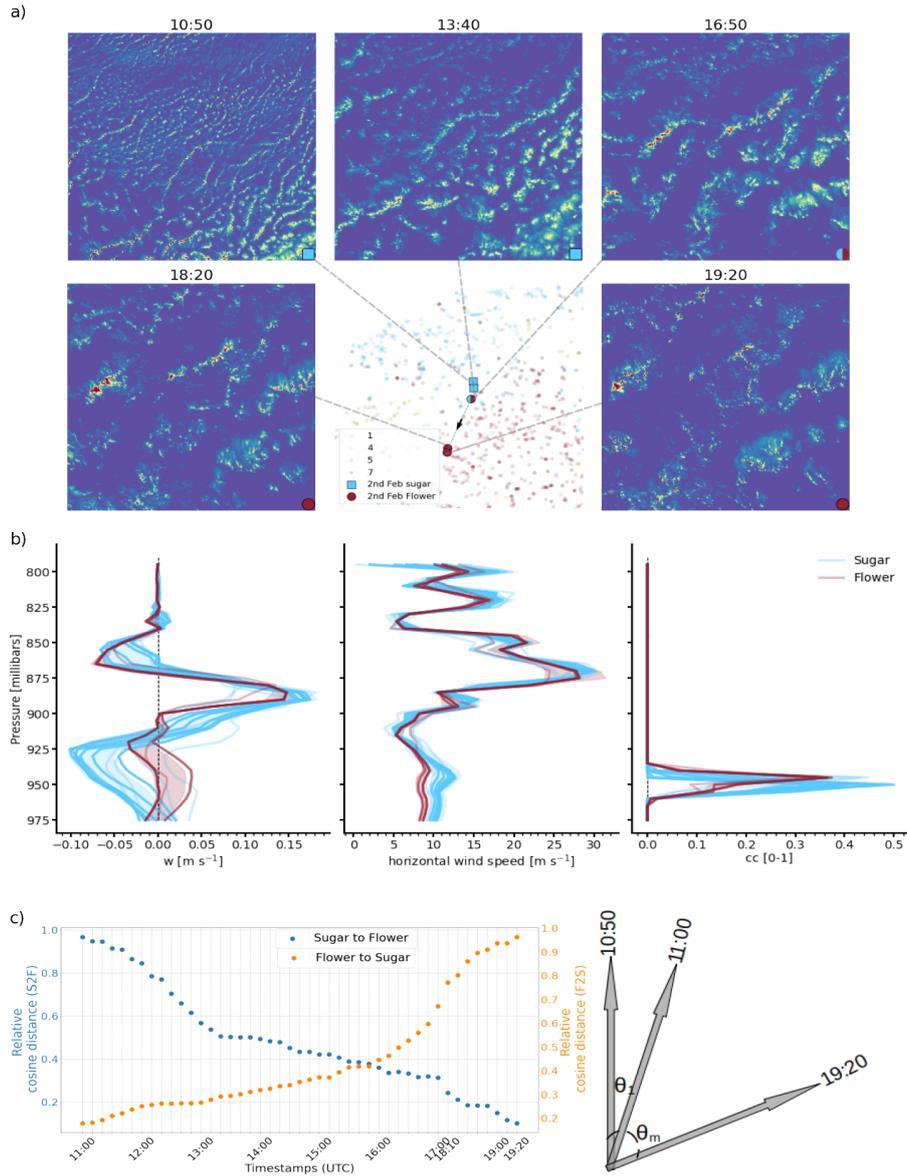


Figure 3. a) Five COD images covering the transition period between sugar and flower on the second of February 2020. Their position in the 2D feature space is indicated in the center of the bottom row. b) Individual and standard deviation profiles of 1) vertical, 2) horizontal wind speed describing the atmospheric dynamics, and 3) cloud cover showing changes in mesoscale structure of the transition samples. c) Illustration of temporal transition development inside the feature space: cosine distance of the first daytime image feature obtained at 10:50 UTC compared with the cloud system evolution features for the rest of the day (blue). The last obtained image at 19:20 UTC towards the first image (orange) and θ_m represents the increasing cosine distance.

315 their features (from N1) and machine labels (from N2), and further dimensionally reduce
 316 the features for 2D visualization.

317 Sugar systems comprise small and shallow clouds with a large spread of individ-
 318 ual cloud cells in a domain, as evident in the beginning (10:50, Fig. 3.a). In contrast,
 319 flower systems appear in multiple deeper aggregates surrounded by large dry areas and
 320 are detected first in the southeast cover at 16:50 before becoming dominated at 19:20
 321 over the full domain. In general, the transition features lie at the border of well-defined
 322 clusters one (‘sugar’) and cluster seven (‘flower’) (Fig. 3.a), and the framework is able
 323 to capture their intermediary nature as they are neither perfect sugar nor flower type.
 324 We use wind speed (vertical and horizontal) to represent changes in atmospheric dynam-
 325 ics and changes in cloud cover to account for the changes in mesoscale structure from
 326 the ERA-5 product. A gradual increase in vertical velocity is observed as the system tran-
 327 sitions from sugar to flowers, and consequently, the surface wind speed gradually reduces
 328 its strength (Fig. 3.b). In addition, as expected, cloud fraction profiles show a gradual
 329 decrease as the transition progresses with time.

330 Sugar-type mesoscale organizations typically occur during the daytime with shal-
 331 low boundary layers, while flowers occur at night with deeper boundary layers (Vial et
 332 al., 2021). We use cosine distance between the features to show the gradual development
 333 of the S2F transition inside the feature space (Fig. 3.c), which quantifies the variation
 334 in visual features of the 47 COD images. The transformation appears smooth initially,
 335 with relatively more significant changes occurring later (post-18:00 UTC) as the system
 336 approaches the flower state. We associate the relatively high changes in cosine distance
 337 compared to initial sugar stages to convective developments which are faster once the
 338 system starts approaching a well-defined flower state. This example illustrates that the
 339 framework can capture the intrinsic characteristics of S2F transitions and can be fur-
 340 ther exploited as a tool to study cloud system transformations and associated processes
 341 with large satellite datasets.

342 6 Conclusion

343 In this work, we develop and make use of a two-step self-supervised learning ap-
 344 proach to study shallow convective organization properties and their transitions. By an-
 345 alyzing organization in a continuous approach without imposing predefined classes, we
 346 include all occurring patterns and transitional states in our analysis. Moreover, the ap-
 347 proach shows that mesoscale cloud organizations in NAT can be classified into seven op-
 348 timal classes for the time period considered. Exploiting the cloud amount at different
 349 vertical levels from CERES measurements, we show how the classes are interlinked with
 350 each other within the continuous space, and thus, the feature space captures the vari-
 351 ability of tropical clouds in more detail.

352 We compare human-labeled cloud systems (Schulz, 2022) to machine-identified clus-
 353 ter regions. Cloud systems with higher agreement among humans lie in the “correct” re-
 354 gion of the feature space, while the ones with less consensus are in the “wrong” regions
 355 of the feature space. Two of the seven optimal classes are strongly related to flower and
 356 sugar, respectively. Representing the sugar-to-flower transition case study (Narenpitak
 357 et al., 2021) for February 2, 2020, in the feature space illustrates the capability to iden-
 358 tify and represent the observed transformations smoothly in their clearly interpretable
 359 regions. We evaluate the transition’s large-scale environmental parameters and observe
 360 a gradual increase in vertical wind speed and a gradual decrease in cloud amount. Fi-
 361 nally, we demonstrate the framework’s capability to capture the underlying mesoscale
 362 visual transformations, such as the transition approaching mature flower convective stages
 363 through quick changes in consecutive cosine distances.

364 One of the limitations of this study is that we use only the daytime cloud retrievals,
 365 and hence, the nocturnal nature of the organizations cannot be captured. Future stud-
 366 ies will use infrared satellite measurements for 24-hour coverage. We aim to fine-tune
 367 our framework with the ground-based observations of the EUREC⁴A campaign and fur-

368 ther extend our analysis to a climate scale. Currently, Destination Earth (Hoffmann et
 369 al., 2023) focuses on simulating high-resolution global digital twins at a 1 km grid scale.
 370 The developed workflow could be a testing ground for investigating the newly adjusted
 371 subgrid parameterization effects on mesoscale cloud systems or atmospheric processes
 372 at different scales.

373 7 Open Research

374 CERES, Edition-4A, DOI:10.5067/TERRA+AQUA/CERES/SYN1DEG-1HOUR_L3.004A)
 375 is made available by the NASA CERES group. ERA-5 reanalyses were downloaded from
 376 the Copernicus climate change services DOI:10.24381/cds.143582cf.

377 The code to produce this work and pre-trained weights of N1 and N2 can be ac-
 378 cessed at <https://doi.org/10.5281/zenodo.8352614>

379 Acknowledgments

380 Dwaipayan Chatterjee’s research was supported by the Federal Ministry for En-
 381 vironment, Nature Conservation, Nuclear Safety, and Consumer Protection. Claudia Ac-
 382 quistapace’s (CA) research was funded by Deutsche Forschungsgemeinschaft (DFG). CA
 383 also acknowledges funding from Federal Ministry for Digital and Transport (BMDV).
 384 The authors would like to thank Andi Walther for providing the GOES-16 E satellite
 385 data. This work was performed as part of the Helmholtz School for Data Science in Life,
 386 Earth, and Energy (HDS-LEE).

387 References

- 388 Bebis, G., & Georgiopoulos, M. (1994). Feed-forward neural networks. *IEEE Potentials*,
 389 *13*(4), 27-31. doi: 10.1109/45.329294
- 390 Bony, S., & Dufresne, J.-L. (2005). Marine boundary layer clouds at the heart of
 391 tropical cloud feedback uncertainties in climate models. *Geophysical Research*
 392 *Letters*, *32*(20). doi: 10.1029/2005GL023851
- 393 Bony, S., Schulz, H., Vial, J., & Stevens, B. (2020). Sugar, Gravel, Fish, and
 394 Flowers: Dependence of Mesoscale Patterns of Trade-Wind Clouds on Environ-
 395 mental Conditions. *Geophysical Research Letters*, *47*(7), e2019GL085988. doi:
 396 10.1029/2019GL085988
- 397 Bony, S., Stevens, B., Ament, F., Bigorre, S., Chazette, P., Crewell, S., ... Wirth,
 398 M. (2017, November). EUREC4A: A Field Campaign to Elucidate the Cou-
 399 plings Between Clouds, Convection and Circulation. *Surveys in Geophysics*,
 400 *38*(6), 1529–1568. doi: 10.1007/s10712-017-9428-0
- 401 Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R.,
 402 ... Webb, M. J. (2015). Clouds, circulation and climate sensitivity. *Nature*
 403 *Geoscience*, *8*, 261–268. doi: 10.1038/ngeo2398
- 404 Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analy-
 405 sis. *Communications in Statistics*, *3*(1), 1-27. Retrieved from [https://](https://www.tandfonline.com/doi/abs/10.1080/03610927408827101)
 406 www.tandfonline.com/doi/abs/10.1080/03610927408827101 doi:
 407 10.1080/03610927408827101
- 408 Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin,
 409 A. (2021). *Emerging properties in self-supervised vision transformers*.
- 410 Chatterjee, D., Acquistapace, C., Deneke, H., & Crewell, S. (2023). Understanding
 411 cloud systems structure and organization using a machine’s self-learning ap-
 412 proach. *Artificial Intelligence for the Earth Systems*. Retrieved from [https://](https://journals.ametsoc.org/view/journals/aies/aop/AIES-D-22-0096.1/AIES-D-22-0096.1.xml)
 413 [journals.ametsoc.org/view/journals/aies/aop/AIES-D-22-0096.1/](https://journals.ametsoc.org/view/journals/aies/aop/AIES-D-22-0096.1/AIES-D-22-0096.1.xml)
 414 [AIES-D-22-0096.1.xml](https://journals.ametsoc.org/view/journals/aies/aop/AIES-D-22-0096.1/AIES-D-22-0096.1.xml) doi: <https://doi.org/10.1175/AIES-D-22-0096.1>
- 415 Dauhut, T., Couvreur, F., Bouniol, D., Beucher, F., Volkmer, L., Pörtge, V., ...

- 416 Wirth, M. (2023). Flower trade-wind clouds are shallow mesoscale convective
417 systems. *Quarterly Journal of the Royal Meteorological Society*, *149*(750),
418 325–347. doi: 10.1002/qj.4409
- 419 Denby, L. (2020). Discovering the Importance of Mesoscale Cloud Organization
420 Through Unsupervised Classification. *Geophysical Research Letters*, *47*(1),
421 e2019GL085190. doi: 10.1029/2019GL085190
- 422 Goyal, P., Duval, Q., Reizenstein, J., Leavitt, M., Xu, M., Lefaudeux, B., ... Misra,
423 I. (2021). *Vissl*. <https://github.com/facebookresearch/vissl>.
- 424 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater,
425 J., ... Thépaut, J.-N. (2020). The era5 global reanalysis. *Quarterly*
426 *Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. doi:
427 <https://doi.org/10.1002/qj.3803>
- 428 Hoffmann, J., Bauer, P., Sandu, I., Wedi, N., Geenen, T., & Thiemert, D. (2023).
429 Destination earth – a digital twin in support of climate services. *Climate*
430 *Services*, *30*, 100394. Retrieved from [https://www.sciencedirect.com/
431 science/article/pii/S2405880723000559](https://www.sciencedirect.com/science/article/pii/S2405880723000559) doi: [https://doi.org/10.1016/
432 j.cliser.2023.100394](https://doi.org/10.1016/j.cliser.2023.100394)
- 433 Janssens, M., Vilà-Guerau de Arellano, J., Scheffer, M., Antonissen, C., Siebesma,
434 A. P., & Glassmeier, F. (2021). Cloud Patterns in the Trades Have Four In-
435 terpretable Dimensions. *Geophysical Research Letters*, *48*(5), e2020GL091001.
436 doi: 10.1029/2020GL091001
- 437 Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022,
438 jan). Transformers in vision: A survey. *ACM Computing Surveys*, *54*(10s),
439 1–41. doi: 10.1145/3505244
- 440 Muller, C. J., & Held, I. M. (2012, August). Detailed Investigation of the Self-
441 Aggregation of Convection in Cloud-Resolving Simulations. *Journal of the At-
442 mospheric Sciences*, *69*(8), 2551–2565. (Publisher: American Meteorological
443 Society Section: Journal of the Atmospheric Sciences) doi: 10.1175/JAS-D-11
444 -0257.1
- 445 Narenpitak, P., Kazil, J., Yamaguchi, T., Quinn, P., & Feingold, G. (2021). From
446 Sugar to Flowers: A Transition of Shallow Cumulus Organization Dur-
447 ing ATOMIC. *Journal of Advances in Modeling Earth Systems*, *13*(10),
448 e2021MS002619. doi: 10.1029/2021MS002619
- 449 Narenpitak, P., Kazil, J., Yamaguchi, T., Quinn, P. K., & Feingold, G. (2023). The
450 Sugar-To-Flower Shallow Cumulus Transition Under the Influences of Diel
451 Cycle and Free-Tropospheric Mineral Dust. *Journal of Advances in Modeling*
452 *Earth Systems*, *15*(1), e2022MS003228. doi: 10.1029/2022MS003228
- 453 Nie, Y., Zamzam, A. S., & Brandt, A. (2021). Resampling and data augmen-
454 tation for short-term pv output prediction based on an imbalanced sky images
455 dataset using convolutional neural networks. *Solar Energy*, *224*, 341–354.
456 Retrieved from [https://www.sciencedirect.com/science/article/pii/
457 S0038092X21004795](https://www.sciencedirect.com/science/article/pii/S0038092X21004795) doi: <https://doi.org/10.1016/j.solener.2021.05.095>
- 458 Nuijens, L., & Siebesma, A. P. (2019, June). Boundary Layer Clouds and Convec-
459 tion over Subtropical Oceans in our Current and in a Warmer Climate. *Cur-
460 rent Climate Change Reports*, *5*(2), 80–94. doi: 10.1007/s40641-019-00126-x
- 461 Paletta, Q., Terrén-Serrano, G., Nie, Y., Li, B., Bieker, J., Zhang, W., ... Feng,
462 C. (2023). Advances in solar forecasting: Computer vision with deep learn-
463 ing. *Advances in Applied Energy*, *11*, 100150. Retrieved from [https://
464 www.sciencedirect.com/science/article/pii/S266679242300029X](https://www.sciencedirect.com/science/article/pii/S266679242300029X) doi:
465 <https://doi.org/10.1016/j.adapen.2023.100150>
- 466 Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chin-
467 tala, S. (2019). *Pytorch: An imperative style, high-performance deep learning*
468 *library*.
- 469 Radtke, J., Naumann, A. K., Hagen, M., & Ament, F. (2022). The relationship
470 between precipitation and its spatial pattern in the trades observed during

- 471 EUREC4A. *Quarterly Journal of the Royal Meteorological Society*, 148(745),
472 1913–1928. doi: 10.1002/qj.4284
- 473 Rasp, S., Schulz, H., Bony, S., & Stevens, B. (2020, November). Combining Crowd-
474 sourcing and Deep Learning to Explore the Mesoscale Organization of Shallow
475 Convection. *Bulletin of the American Meteorological Society*, 101(11), E1980–
476 E1995. (Publisher: American Meteorological Society Section: Bulletin of the
477 American Meteorological Society) doi: 10.1175/BAMS-D-19-0324.1
- 478 Rauber, R. M., Stevens, B., Ochs, H. T., Knight, C., Albrecht, B. A., Blyth, A. M.,
479 ... Zuidema, P. (2007, December). Rain in Shallow Cumulus Over the Ocean:
480 The RICO Campaign. *Bulletin of the American Meteorological Society*, 88(12),
481 1912–1928. doi: 10.1175/BAMS-88-12-1912
- 482 Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and val-
483 idation of cluster analysis. *Journal of Computational and Applied Mathemat-*
484 *ics*, 20, 53-65. Retrieved from [https://www.sciencedirect.com/science/](https://www.sciencedirect.com/science/article/pii/0377042787901257)
485 [article/pii/0377042787901257](https://www.sciencedirect.com/science/article/pii/0377042787901257) doi: [https://doi.org/10.1016/0377-0427\(87\)](https://doi.org/10.1016/0377-0427(87)90125-7)
486 90125-7
- 487 Sarkar, M., Zuidema, P., Albrecht, B., Ghate, V., Jensen, J., Mohrmann, J., &
488 Wood, R. (2020). Observations pertaining to precipitation within the north-
489 east pacific stratocumulus-to-cumulus transition. *Monthly Weather Review*,
490 148(3), 1251 - 1273. doi: 10.1175/MWR-D-19-0235.1
- 491 Schmit, T. J., Gunshor, M. M., Menzel, W. P., Gurka, J. J., Li, J., & Bachmeier,
492 A. S. (2005). Introducing the next-generation advanced baseline imager on
493 goes-r. *Bulletin of the American Meteorological Society*, 86(8), 1079 - 1096.
494 Retrieved from [https://journals.ametsoc.org/view/journals/bams/86/8/](https://journals.ametsoc.org/view/journals/bams/86/8/bams-86-8-1079.xml)
495 [bams-86-8-1079.xml](https://journals.ametsoc.org/view/journals/bams/86/8/bams-86-8-1079.xml) doi: <https://doi.org/10.1175/BAMS-86-8-1079>
- 496 Schubert, E. (2023, jun). Stop using the elbow criterion for k-means and how to
497 choose the number of clusters instead. *ACM SIGKDD Explorations Newsletter*,
498 25(1), 36–42. doi: 10.1145/3606274.3606278
- 499 Schulz, H. (2022). C³ontext: a common consensus on convective organization during
500 the eurec⁴a experiment. *Earth System Science Data*, 14(3), 1233–1256. doi: 10
501 .5194/essd-14-1233-2022
- 502 Schulz, H., Eastman, R., & Stevens, B. (2021). Characterization and Evolution
503 of Organized Shallow Convection in the Downstream North Atlantic Trades.
504 *Journal of Geophysical Research: Atmospheres*, 126(17), e2021JD034575. doi:
505 10.1029/2021JD034575
- 506 Seifert, A., Heus, T., Pincus, R., & Stevens, B. (2015). Large-eddy simulation of
507 the transient and near-equilibrium behavior of precipitating shallow convec-
508 tion. *Journal of Advances in Modeling Earth Systems*, 7(4), 1918-1937. doi:
509 <https://doi.org/10.1002/2015MS000489>
- 510 Stevens, B., Bony, S., Brogniez, H., Hentgen, L., Hohenegger, C., Kiemle, C., ...
511 Zuidema, P. (2020). Sugar, gravel, fish and flowers: Mesoscale cloud patterns
512 in the trade winds. *Quarterly Journal of the Royal Meteorological Society*,
513 146(726), 141–152. doi: 10.1002/qj.3662
- 514 Stevens, B., Bony, S., Farrell, D., Ament, F., Blyth, A., Fairall, C., ... Zöger, M.
515 (2021, August). EUREC⁴A. *Earth System Science Data*, 13(8), 4067–4119.
516 (Publisher: Copernicus GmbH) doi: 10.5194/essd-13-4067-2021
- 517 Stevens, B., Farrell, D., Hirsch, L., Jansen, F., Nuijens, L., Serikov, I., ... Prospero,
518 J. M. (2016). The barbados cloud observatory: Anchoring investigations
519 of clouds and circulation on the edge of the itcz. *Bulletin of the Ameri-*
520 *can Meteorological Society*, 97(5), 787 - 801. doi: [https://doi.org/10.1175/](https://doi.org/10.1175/BAMS-D-14-00247.1)
521 [BAMS-D-14-00247.1](https://doi.org/10.1175/BAMS-D-14-00247.1)
- 522 Tomassini, L., Voigt, A., & Stevens, B. (2015). On the connection between tropi-
523 cal circulation, convective mixing, and climate sensitivity. *Quarterly Journal of*
524 *the Royal Meteorological Society*, 141(689), 1404-1416. doi: [https://doi.org/10](https://doi.org/10.1002/qj.2450)
525 [.1002/qj.2450](https://doi.org/10.1002/qj.2450)

- 526 van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-sne. *Journal*
 527 *of Machine Learning Research*, 9(86), 2579–2605. Retrieved from [http://jmlr](http://jmlr.org/papers/v9/vandermaaten08a.html)
 528 [.org/papers/v9/vandermaaten08a.html](http://jmlr.org/papers/v9/vandermaaten08a.html)
- 529 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N.,
 530 ... Polosukhin, I. (2023). *Attention is all you need*. Retrieved from
 531 <https://arxiv.org/abs/1706.03762>
- 532 Vial, J., Bony, S., Stevens, B., & Vogel, R. (2017). Clouds, circulation and climate
 533 sensitivity. *Surveys in geophysics*, 38(6), 1331–1353. doi: 10.1007/s10712-017-
 534 -9418-2
- 535 Vial, J., Vogel, R., & Schulz, H. (2021). On the daily cycle of mesoscale cloud organ-
 536 ization in the winter trades. *Quarterly Journal of the Royal Meteorological So-*
 537 *ciety*, 147(738), 2850–2873. doi: 10.1002/qj.4103
- 538 Vogel, R., Albright, A. L., Vial, J., George, G., Stevens, B., & Bony, S. (2022, De-
 539 cember). Strong cloud–circulation coupling explains weak trade cumulus feed-
 540 back. *Nature*, 612(7941), 696–700. (Number: 7941 Publisher: Nature Publish-
 541 ing Group) doi: 10.1038/s41586-022-05364-y
- 542 Vogel, R., Konow, H., Schulz, H., & Zuidema, P. (2021, November). A climatology
 543 of trade-wind cumulus cold pools and their link to mesoscale cloud organiza-
 544 tion. *Atmospheric Chemistry and Physics*, 21(21), 16609–16630. (Publisher:
 545 Copernicus GmbH) doi: 10.5194/acp-21-16609-2021
- 546 Vogel, R., Nuijens, L., & Stevens, B. (2020). Influence of deepening and mesoscale
 547 organization of shallow convection on stratiform cloudiness in the downstream
 548 trades. *Quarterly Journal of the Royal Meteorological Society*, 146(726), 174-
 549 185. doi: <https://doi.org/10.1002/qj.3664>
- 550 Walther, A., & Heidinger, A. K. (2012). Implementation of the daytime cloud opti-
 551 cal and microphysical properties algorithm (dcomp) in patmos-x. *Journal of*
 552 *Applied Meteorology and Climatology*, 51(7), 1371 - 1390. doi: 10.1175/JAMC
 553 -D-11-0108.1
- 554 Wielicki, B., Barkstrom, B., Harrison, E., Lee, R., Smith, G., & Cooper, J. (1996,
 555 05). Clouds and the earth’s radiant energy system (ceres): An earth observing
 556 system experiment. *Bulletin of the American Meteorological Society*, 77, 853-
 557 868. doi: 10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2

558 References From the Supporting Information

- 559 Bottou, L. (2012). Stochastic gradient descent tricks. In G. Montavon, G. B. Orr,
 560 K.R. Müller (Eds.), *Neural networks: Tricks of the trade: Second edition* (pp. 421–436).
 561 Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-35289-8_25
- 562 Bridle, J. S. (1989). Probabilistic interpretation of feedforward classification net-
 563 work outputs, with relationships to statistical pattern recognition. In *Nato neurocom-*
 564 *puting*. Retrieved from <https://api.semanticscholar.org/CorpusID:59636530>
- 565 Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin,
 566 A. (2021). Emerging properties in self-supervised vision transformers.
- 567 Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner,
 568 T., Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recog-
 569 nition at scale.
- 570 Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network.
 571 *Biological Cybernetics*, 20 (3-4), 121–136. Retrieved 2022-08-30, from [http://link.springer](http://link.springer.com/10.1007/BF00342633)
 572 [.com/10.1007/BF00342633](http://link.springer.com/10.1007/BF00342633)doi:10.1007/BF00342633
- 573 Goyal, P., Duval, Q., Reizenstein, J., Leavitt, M., Xu, M., Lefaudeux, B., Misra,
 574 I. (2021). Vissl. <https://github.com/facebookresearch/vissl>

- 575 He, K., Zhang, X., Ren, S., Sun, J. (2015, December). Deep Residual Learning for
576 Image Recognition. Retrieved 2022-08-30, from <http://arxiv.org/abs/1512.03385> (arXiv:1512.03385
577 [cs])
- 578 Hendrycks, D., Gimpel, K. (2023). Gaussian error linear units (gelus). Retrieved
579 from <https://arxiv.org/abs/1606.08415>
- 580 Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., Shah, M. (2022, jan).
581 Transformers in vision: A survey. *ACM Computing Surveys*, 54 (10s), 1–41. doi: 10.1145/3505244
- 582 Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986, October). Learning rep-
583 resentations by back-propagating errors. *Nature*, 323 (6088), 533–536. Retrieved from
584 <https://doi.org/10.1038/323533a0> doi: 10.1038/323533a0
- 585 Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Fei-Fei, L. (2015,
586 December). ImageNet Large Scale Visual Recognition Challenge. *International Journal*
587 *of Computer Vision*, 115 (3), 211–252. Retrieved 2022-08-31, from [http://link.springer](http://link.springer.com/10.1007/s11263-015-0816-y)
588 [.com/10.1007/s11263-015-0816-y](http://link.springer.com/10.1007/s11263-015-0816-y) doi: 10.1007/s11263-015-0816-y
- 589 Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . .
590 . Polosukhin, I. (2023). Attention is all you need. Retrieved from [https://arxiv.org/](https://arxiv.org/abs/1706.03762)
591 [abs/1706.03762](https://arxiv.org/abs/1706.03762)