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

Dwaipayan Chatterjee[1], Sabrina Schnitt[1], Paula Bigalke[1], Claudia Acquistapace[1], Susanne Crewell[1]

[1]Institute for Geophysics and Meteorology, University of Cologne, Cologne, Germany

Key Points:

- Mesoscale cloud organization can be taxonomized by a two-step deep learning approach in the feature space continuum
- Comparing seven machine-identified classes with humans' four recognized categories underlines the significance of uncertainty estimates
- New diagnostic is provided to analyze the temporal transition between regimes, as illustrated for human-labeled sugar-to-flower regimes

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

Abstract

At mesoscale, trade wind clouds organize with various spatial arrangements, shaping their effect on Earth's energy budget. Representing their fine-scale dynamics even at 1 km scale climate simulations remains challenging. However, geostationary satellites (GS) offer high-resolution cloud observation for gaining insights into trade wind cumuli from long-term records. To capture the observed organizational variability, this work proposes an integrated framework using a continuous followed by discrete self-supervised deep learning approach, which exploits cloud optical depth from GS measurements. We aim to simplify the entire mesoscale cloud spectrum by reducing the image complexity in the feature space and meaningfully partitioning it into seven classes whose connection to environmental conditions is illustrated with reanalysis data. Our framework facilitates comparing human-labeled mesoscale classes with machine-identified ones, addressing uncertainties in both methods. It advances previous methods by exploring transitions between regimes, a challenge for physical simulations, and illustrates a case study of sugar-to-flower transitions.

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) in the tropical northern Atlantic region east of Barbados. 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 physical simulation before. The clear signatures of the transition identified by our machine learning 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, though individually small (measuring in tens of meters), cover large areas of the tropical oceans, forming distinct cloud fields that span hundreds of km. They are vital in regulating the Earth's energy balance, exerting a net cooling effect by reflecting more sunlight than retaining outgoing long-wave radiation (Bony et al., 2004). However, the representation of these clouds, even in the advanced 1km scale climate simulations, is insufficient (Schneider et al., 2019). This contributes to a significant inter-model spread in predicted cloud feedback and climate sensitivity (Bony & Dufresne, 2005; Nuijens & Siebesma, 2019). To address this challenge, Bony et al. (2017) proposed the EUREC^{4}A field campaign, organized in January-February 2020, around the Barbados region of the North Atlantic Trades (NAT) (Stevens et al., 2021). This initiative aimed to enhance our understanding of shallow cloud dynamics by leveraging a diverse set of observations and thus possibly improving their representation in numerical models.

During the preparation of the campaign Stevens et al. (2020) identified four shallow convective organization regimes (*Sugar, Gravel, Flower, Fish*) (SGFF), with frequent occurrence on meso-β (20 to 200 km) and meso-α (200 to 2,000 km) spatial scale. These regimes exhibit differences in net cloud radiative feedback (Bony et al., 2020) and are related to different environmental conditions (Schulz et al., 2021). Of specific interest are transitions between different organizations, e.g., from sugar to flower, which has been studied in Large-Eddy-Simulation (LES) to understand the governing processes and prove to be difficult (Narenpitak et al., 2021; Dauhut et al., 2023).

Yet, imposing four distinct classes on the diversity of the observed organization does not cover the intermediate cloud patterns or transient states, as highlighted by LES studies. Hence, some processes critical for climate feedback may be ignored or neglected. Furthermore, recent studies trying to quantify these labeled well-organized systems find that they occur only around 50% over NAT (Janssens et al., 2021; Schulz et al., 2021; Vial et al., 2021) and some ambiguities in agreement from the labeler's side exist (Schulz, 2022).

Denby (2020) and Janssens et al. (2021) argue for a continuum of cloud organization where Denby (2020) employs an unsupervised neural network for grouping similar cloud structures and demonstrate its effectiveness via hierarchical clustering (HC) and associated radiative properties. However, their training approach involved a possibility of false negative sampling (Huynh et al., 2022), where the negative pair's distant tile (taken from a random location on a different day) does not necessarily guarantee a dissimilarity in their cloud system's structure and distribution. Further, employing high-dimensional features in HC has performance and scalability issues (Du, 2023; Gilpin et al., 2013). Janssens et al. (2021) assumes a linear combination of traditional cloud metrics for describing the cloud systems. Utilizing these metric scores and a k-means algorithm, they attempted to partition their metric space into seven arbitrary clusters, as finding meaningful cloud regimes (CRs) seemed non-trivial.

The overarching goal of our study is to develop a simplified approach to describe cloud organization from high-resolution images. In this way, it should open up new pathways to exploit the information content of existing comprehensive satellite data records. Our first objective is to develop a streamlined representation that captures the entire cloud spectrum's organizational relationships, which we call a continuum. Second, we target the four somewhat arbitrary classes from Stevens et al. (2020) and delve deeper into finding useful CRs from an interpretable continuum. We approach our objectives by developing a two-step self-supervised deep learning approach (Section 3) applied on GOES − 16 E cloud optical depth (COD) images (Section 2). Section 4.1 delves deeper into the representations and their characteristics, highlighting the differences to Denby (2023)'s approach. Our work demonstrates that the presence of derived partitions facilitates a comparison of human labels with these partitions (Section 4.2). Finally, in Section 5, we illustrate how the partitioning of the continuum supported by environmental data allows us to monitor when a particular cloud system transitions to another.

2 Satellite dataset

We use COD retrieved from GOES-16 E Advanced Baseline Imager (Schmit et al., 2005) using the daytime cloud optical and microphysical properties algorithm (DCOMP) (Walther & Heidinger, 2012) at 2 km horizontal resolution and 10 − 15 minutes temporal resolution. Our domain in NAT (5 - 20° N and 40 − 60° W) is similar to domains used in past studies (Bony et al., 2020; Schulz et al., 2021). The regional climate defines December to May as dry and June to November as wet seasons (Stevens et al., 2016). We consider November to April 2017 - 2021 as our study period. November is added to the typical dry period because we want to see how stronger convective events influence our approach.

We chose COD because it is closely related to the cloud radiative effect and mitigates solar and surface influences. The uncertainty associated with COD retrieval remains below 10% for all ranges in water clouds (see Figure 4 in Walther and Heidinger (2012)). Note that some fine-scale cloud systems, such as sugar and gravel (meso-β scale), their individual cloud cells might not be fully resolved with the spatial resolution of this product. However, since our study focuses on the organizational aspects of shallow convection clouds (spanning hundreds of km), we expect the resolution limit to have a limited impact on our study.

Representation learning, also known as feature learning, is a specialized field within machine learning that focuses on extracting meaningful features of a given dataset. To better represent the mesoscale cloud distributions, we use six images per timestamp, including an additional fixed image over the Barbados domain (see S1). Although they might overlap in some instances, random cropping aims to get mesoscale distributions as diverse as possible without human interference. Note that the Barbados domain enables comparison with ground-based measurements in future studies. To have an adequate spatial scale of typical occurring cloud fields over NAT (as discussed in Section 1), we use 256 x 256 pixels (roughly 512 square km) as also found in Muller and Held (2012). We exclude crops affected by glint or poor retrieval quality using the respective data flags. Time stamps are limited to 9 am - 3 pm Barbados local time to avoid sun glinting. We use land class data to filter out images with convection over land, specifically over the northeast of the South American continent. Finally, to mitigate uncertainties at high COD from DCOMP retrieval, COD values above a threshold of 50, already indicating deep clouds, are clipped to 50. This results in a sample size of 51,000 satellite images.

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

3 Methods

The workflow is as follows: a) A neural network (N1) ingests satellite images to continuously sort cloud organizations based on visual similarity, yielding the feature vector 'Z' (384 dimensions) for each image. b) Z is reduced to a 2-dimensional (2D) space for visualizing a continuous arrangement of images with respect to their cloud structures (continuum). c) The optimal number 'K' of meaningful clusters or CRs is derived from the 2D representation, d) A second neural network (N2) of similar architecture as N1 but constrained by 'K' classes ingests the satellite images to finally assign each image to a discrete class.

a) N1 identifies the structural similarities in the cloud systems and maps the learned visual features into the 384-dimensional feature space Z. To learn similar embeddings of semantically similar mesoscale structure and distributions, every epoch, we opt for two random global crops with a 0.75 fraction (192 x 192 pixels) within a single parent satellite image. Each crop is processed in a separate branch (called student and teacher) by a Vision Transformer (ViT), which has a sequence of self-attention (Vaswani et al., 2023) and feed-forward layers (Bebis & Georgiopoulos, 1994). Note that both branches have the same general architecture, but the parameters (weights and biases) learned during training are slightly different. As the largely overlapping global-crop pair has very similar cloud structures, the network learns their essential features and puts the pair closer to each other in the high-dimensional feature space. A cross-entropy loss function minimizes the difference in the output of both branches and yields the 384-element feature vector Z. More details on the implementation are given in S2. This way of training, unlike Denby (2020), eliminates the need for a negative pair and avoids linearized assumptions like in Janssens et al. (2021).

b) Z includes the continuously sorted representation of cloud organization. We reduce its 384 dimensions to two dimensions using the well-established t-distributed Stochastic Neighbor Embedding or t-SNE algorithm (van der Maaten & Hinton, 2008). It preserves relative local positions by using cosine distance in affinity computation and tries to retain global structure by initializing with principal components for mapping to a two-dimensional space. This proves helpful because high-dimensional data when directly applied to cluster analysis, face challenges like the curse of dimensionality (Aggarwal et al., 2001), where increased dimensions make distances between data points less meaningful. Also, the presence of noise and outliers can distort clusters, hindering the algorithm's ability to identify distinct clusters (Steinbach et al., 2004).

c) After obtaining the continuously sorted 2D representation of cloud systems (see Fig. 1.a), we intend to find optimal boundary conditions within the sorted order to derive distinct clusters (CRs). Selecting a meaningful and interpretable number of clusters is crucial to avoid over-fitting, where excessive clusters can capture noise, and also under-fitting, where too few clusters can miss significant patterns in the data. On this 2D representation space, we apply a set of three statistical approaches, namely metric scores of distortion, silhouette (Rousseeuw, 1987), and Calinski-Harabasz (Caliński & Harabasz, 1974) to identify the number of optimal classes into which the given features could be clustered. Schubert (2023) suggests taking a collective inference from these three methods to best fit the spherical k-means clustering algorithm used during the training of N2. S3 illustrates how the three metrics point to an optimal clustering of the continuum into seven CRs. Note that the choice of seven classes is robust, as illustrated by several sensitivity tests (shown in S4), such as the dimensionality-reduction technique, size of the dataset, initial weights of the network, and different global crop sizes.

d) N2 from Chatterjee, Acquistapace, et al. (2023) is similar to N1 concerning the two branches: When the feature vectors of both branches capture similar information from the global crops, the loss decreases; conversely, it increases when they diverge. However, before the cross-entropy loss is computed in each branch, a spherical k-means clustering is applied. Herein, the feature vector from the upper branch gets assigned a class (target label) based on proximity to its nearest centroid while the lower branch feature vector tries to reduce its cosine distance with the allocated centroid (predicted proxy) to reduce the loss. In this way, the network learns to allocate both global crops to the same class. After obtaining the label of each satellite image, we transfer the assigned class to the continuum space, which proves helpful because N1 has learned the sorting arrangement of keeping similar cloud systems closer. Therefore, it helps to visualize how each cluster with distinct characteristics can form a separate local region. Additionally, the N2 feature space is i) more sparse than N1 (see S2 for explanation) and ii) arranged by closeness to the centroids, which, unlike N1, may not be ideal for representing smooth transitions of cloud systems. Note that there are further differences between N1 and N2, e.g., image augmentation, which are detailed in S2.

4 Results

4.1 Continuous and discrete representations

We now analyze the diversity of cloud systems included in the satellite data record within their continuous and discrete representations. Both are visualized in 2D continuum space using the t-SNE algorithm (Section 3). The organization state captured in the satellite images changes smoothly and different cloud organizations can be identified in different areas of the continuum (Fig. 1.a). Going anticlockwise from the top, arch-shaped cloud systems lie in the top-left, followed by flower-type distributions on the left side of the continuum. Close to the flowers in the bottom-left are the flowers spreading out into stratocumulus. Note that physically simulating the transition is challenging as modeling studies struggle to capture the stratocumulus to cumulus transition (Sarkar et al., 2020), although they lie adjacent in the continuum.

The bottom part of the feature space contains long bony skeletons, i.e., fish-type cloud systems, and the bottom-right corner shows an extended part of fish-type cloud organizations delineated by unusually large cloud-free regions. The top-right region of the continuum is a collection of deep convective cells. These primarily occur in the month of November. Arc-shaped cloud systems appear on the left and top-left of the continuum. Vogel et al. (2021) suggest that the horizontal structure of mesoscale arcs is intrinsically linked to gravel, flowers, and fish. In sequence, Figure 1a shows a continuous link in the spatial arrangement of cloud systems rather than the distinct classes. This demonstrates the good performance of our continuous approach, which is further supported by the analysis of attention maps in S5. Note that any newly taken satellite image can be placed into this continuum using the weights of N1, allowing a quick assessment of its organizational status. Also, similar trajectories of subsequent images can be tracked within the continuum space.

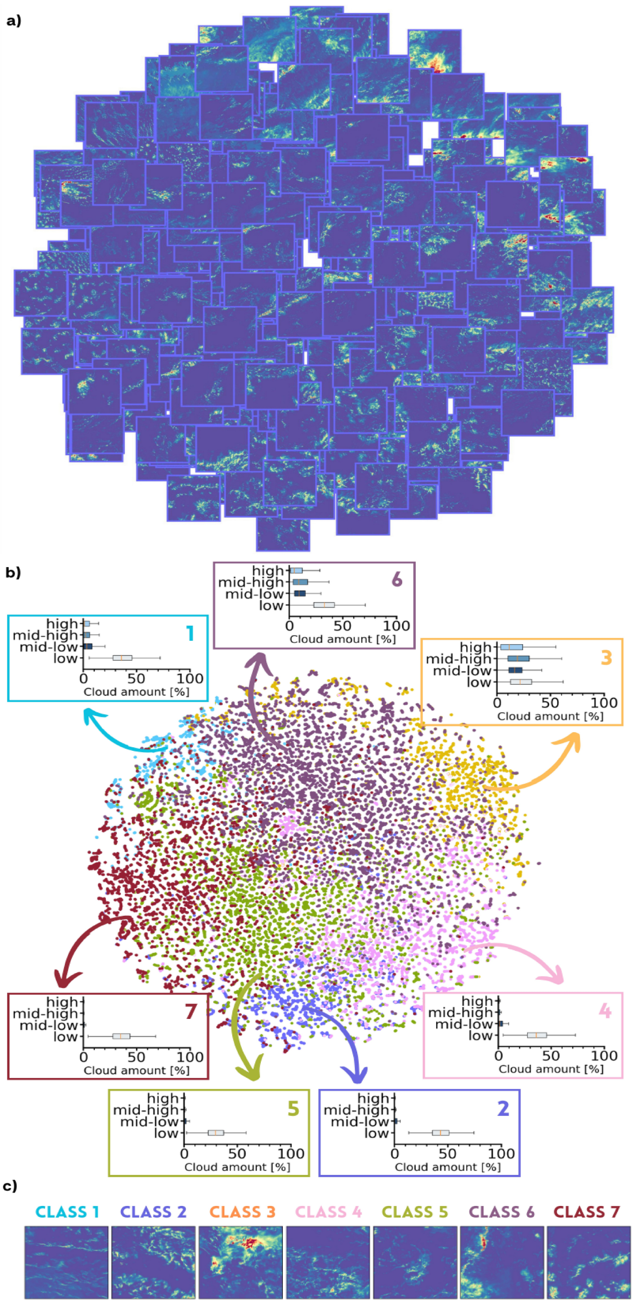
After training N2, each of the images can be attributed to one of the seven classes (refer to Section 3), revealing distinct spaces within the continuum (Fig. 1.b). To assess how well the seven classes separate, they are evaluated using cloud amounts at four different height levels from CERES data. This analysis, on the one hand, reflects how each class differs from the others, and on the other hand, it reasons for the underlying closeness of each class with neighbor classes in the continuum. The difference between the seven clusters is especially evident when looking at their centroid images (Fig. 1.c).

Deep convective class three has by far the highest cloud fraction of 76% and a third more water vapor (47.0 $\mathrm{kgm^{-2}}$) than all other classes (mean = 32.5 $\mathrm{kgm^{-2}}$). We use IWV as a fingerprint for the origin of air masses and intend to test it later to investigate the connection between CR and air mass origin. Figure 1.b already shows that class 3, which by far has the highest IWV, is also related to the deepest convection. Neighboring class six includes less frequent higher-level clouds and has a reduced CF of 59% compared to class three. All other classes are dominated by low-level clouds with lower than 50% CF. Classes one and four (neighbor to class six) still have some mid to high-level cloud amounts (below 10%). Class one can be interpreted as representing arch-shaped cloud systems, and four resembles the fish class with a more open sky (also shown by reduction in CF).

Classes two, five, and seven, being close in the continuum, have similar cloud vertical distributions and IWV ranging from 30 to 32 $\mathrm{kgm^{-2}}$; however, their organization is very different, as illustrated by the centroids (Fig. 1.c) and mean CFs (43%, 27%, and 33%, respectively). Class two primarily comprises shallow cloud cover, corresponding to cloud systems resembling fish-type formations. Class five has the lowest CF and is an intermediary class type between classes two and seven. Finally, class seven has a presence of low cloud amounts and negligible mid to higher cloud amounts, which visually resembles flower-type cloud distributions. Therefore, discretizing the continuum helps us visually find three main classes (one, two, and seven) frequently resembling features identified by humans, i.e., sugar, fish, and flower, respectively. However, it also shows the remaining diversity and their characteristics in a cohesive approach. Note that in contrast to the challenges faced by Denby (2023) or Janssens et al. (2021) in isolating meaningful clusters, our N1 + N2 framework excels in simplifying the cloud organization complexities by efficiently categorizing the continuum into seven interpretable classes.

4.2 Machine versus human labels

While we checked for visual correspondence and class-wise characteristics in Section 4.1, our framework now creates the opportunity to quantify how human labels compare to the machine's seven clusters. For this, we use the dataset from Schulz (2022), which is a 1km x 1km resolution manually labeled dataset for the NAT region and EUREC^{4}A



CF	47%	43%	76%	36%	27%	59%	33%
IWV	33.08	30.24	47.01	34.18	32.09	35.37	30.58

Figure 1. a) Visualization of four hundred randomly selected satellite images arranged in the continuum space. b) Same as a), but now, instead of an image, the discrete class determined by N2 is shown (colored). For each class, statistics on low, mid-low, mid-high, and high cloud amount (%) obtained from the CERES hourly data set are provided. c) Centroid COD images belonging to seven clusters as identified by the discrete neural network (N2). The table shows per class the average of cloud fraction (CF, %) from the GOES retrieval and integrated water vapor (IWV, kgm^{-2}) from ERA-5.

time period (47 days). Approximately 50 scientists generated the dataset by identifying mesoscale patterns (SGFF) and marking variable-sized rectangles around homogeneous organization states. Overlapping rectangles allowed a single grid point to be labeled with multiple patterns by a scientist. Individual uncertainty is expressed through each pattern's classification mask (c_m) (Schulz, 2022). For example, if a grid point is within both gravel and sugar rectangles, the c_m would be 0.5 for both and 0 for the other two patterns. Mutual agreement among scientists for each pattern at a grid point is determined by averaging c_m values, ranging from 0 to 100%.

We hypothesize patterns with higher agreement are most likely attributed to their meaningful partitions within the continuum (as discussed in Section 4.1). For each timestamp where at least one of the four patterns was identified within our domain, we select a 256 x 256-pixel satellite image centered over the area of highest human agreement. In this way, we ensure the best possible intercomparison. This leaves us with 52 samples of human-labeled satellite images (fish: 19.3%, gravel: 26.9%, flower: 28.8%, sugar: 25.0%). Note that even with the highest consensus criteria, there's still diversity in agreement. The inter-quartile agreement range is 35%, while the minimum and maximum agreements show consensus levels of 7% and 91%, respectively.

The framework classifies 40% flower-labeled cloud systems in class seven (see the hit rate for each class in Fig. 2.a) while sugar-labeled cloud systems are 31% classified in class one and 20% in class four. Gravel has a total of 44% representation in classes one and five, whereas fish annotated labels are allocated 30% in class two and 20% each in classes four and five. Further, examining example images visually (Fig. 2.a), it becomes apparent that images with lower human agreement notably diverge from the established definitions (provided in Stevens et al. (2020)) of SGFF cloud structures, in contrast to images with high human agreement.

Within the continuum (Fig. 2.b), flowers detected with high probability mostly occur in areas of class seven, which was already well reflected in the centroids. Following a similar agreement is sugar (street-type cloud systems), which can be found in areas of class one. However, 38% of sugar samples, with a low agreement, lie in classes four and five, which are extended fish and flower type classes (Section 4.1). Note that even though these samples reside in those regions of the feature space, their confidence is less than 25%. Similarly, in the gravel pattern, 21% samples belong to class six and exhibit minimal human confidence. In contrast, the rest from the gravel class are positioned between classes one and seven, suggesting that gravel cloud cell sizes fall between sugar and flower. Rightly, no human-labeled samples are found in class three, which predominantly comprise deep convective cells. Finally, the fish class exhibits relatively higher confidence in human labels, aligning well with the feature space characteristics, and lies in class two (fish) and four (extended fish-type cloud structures with large cloud-free regions). Hence, cloud systems characterized by higher agreement among human observers are situated within the designated regions, while those with lesser consensus are positioned within the ambiguous regions of the continuum.

To compensate for the limited number of human label samples, we analyze the 30 nearest satellite images to each human label as identified by N1 (Fig. 2.c). The majority of neighbors in human-identified fish-type cloud systems (more than 50%) belong to machine-identified classes two and four. The gravel regime includes members of all classes, with notable contributions from classes one, five, and seven, which exhibit cloud cell characteristics similar to gravel systems. The variability in the spread can be linked to the limited representation of gravel glass in Schulz (2022)'s dataset, as gravel occurrences were sporadic during the EUREC^{4}A campaign. Additionally, 75% of gravel labels in our sub-samples had agreement levels below 0.25. In contrast, the flower regime mainly belongs to class seven (46 %), further aligning with the high confidence of human labels. Regarding sugar-type cloud systems, 37 % of the neighbors fall into class one, while those with low human agreement are scattered across the remaining classes. Therefore, we find

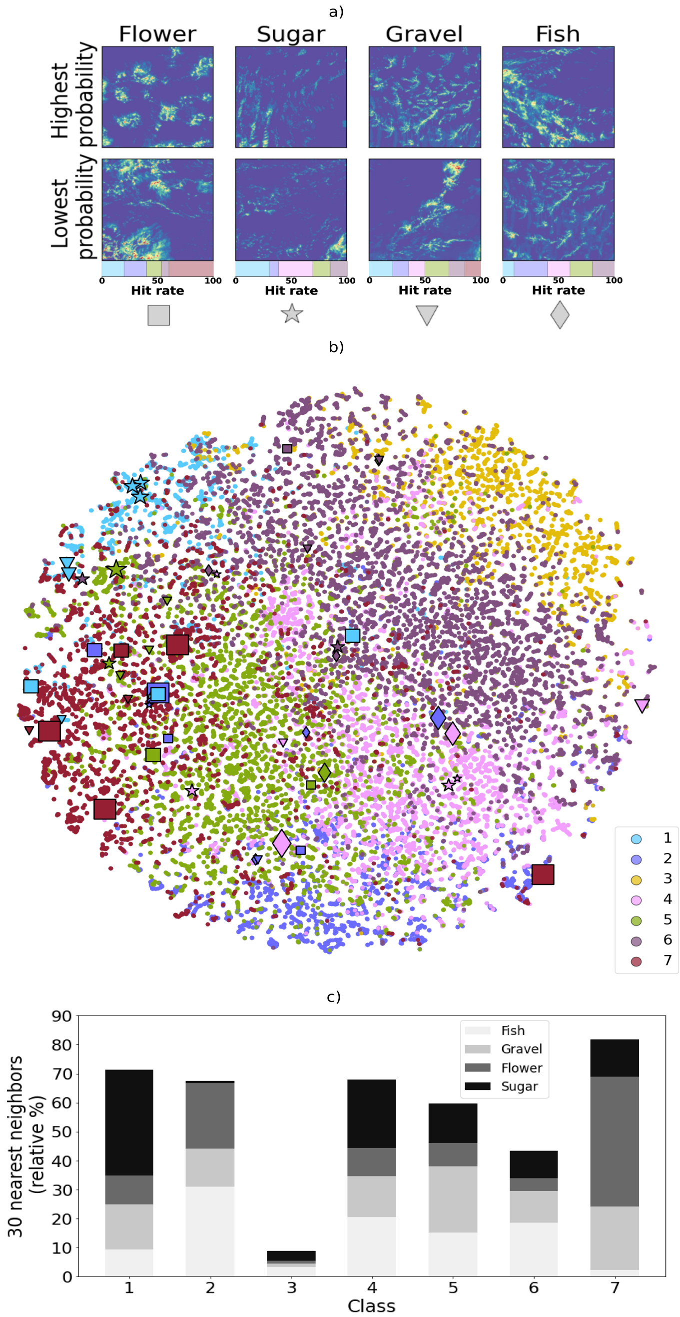


Figure 2. a) To enhance visualization and reference for human labels, each column displays 256 x 256 COD images of a specific class, with the highest and lowest human agreement shown in two rows. Below, the images in each column show the hit rate, representing the N2-predicted class for each human label. b) Continuum space colored with different classes (1-7) in the background, along with Human labels (fish, sugar, flower, gravel) in the foreground. Ascending symbol sizes with low (0-0.25), mid-low (0.25-0.50), mid-high (0.50-0.75), and high (0.75-1.00) agreement are shown. c) Relative occurrence of 30 nearest neighbors to human-labeled fish, gravel, flower, and sugar along the seven machine-labeled classes.

that machine-labeled classes of the 30 nearest neighbors encompass the human-labeled ones, especially for sugar, flower, and fish, but not so clearly for gravel.

Further, in S6, using ERA-5 large-scale environmental variables and cloud physical properties, we demonstrate that the neighbors and the human crops share a similar, homogeneous distribution of physical properties. Therefore, this analysis, for the first time, shows how to exploit the labels and physical properties of the semantically similar nearest-neighbors of any cloud system of interest further to enhance our understanding of the connection between organizations.

5 Transitions

To showcase an application that highlights the intelligible partitioning of the continuum, we explore the "sugar" to "flower" (S2F) cloud system transition on February 2, 2020. Using LES, Narenpitak et al. (2021) showed a strengthening of large-scale upward wind motion and an increase in total water path and optical depth as the transformation develops towards the flower. Here, we look at how the transition in COD is represented in the feature space. For example, where do the representations of transitions lie in the feature space? How smooth is the transition in the feature space?

Covering the temporal developments, 47 COD images were collected (after applying quality filter checks (see Section 2)), centered at 12.5° N, 50° W. They cover the time from 10:50 to 19:20 UTC, with a gap between 17:00 to 18:00 UTC likely caused by local sun glint. We ingested the available samples into the trained framework and collected their features (from N1) and machine labels (from N2).

Sugar systems comprise small and shallow clouds with a large spread of individual cloud cells in a domain, as evident in the beginning (10:50, Fig. 3.a). In contrast, flower systems appear in multiple deeper aggregates surrounded by large dry areas and are detected first in the southeast cover at 16:50 before becoming dominated at 19:20 over the full domain. In general, the transition features lie at the border of well-defined clusters one ('sugar') and cluster seven ('flower') (Fig. 3.a), and the framework is able to capture their intermediary nature as they are neither perfect sugar nor flower type. We use wind speed (vertical and horizontal) to represent changes in atmospheric dynamics and changes in cloud cover to account for the changes in mesoscale structure from the ERA-5 product. A gradual increase in vertical velocity is observed as the system transitions from S2F, and consequently, the surface wind speed gradually reduces its strength (Fig. 3.b). In addition, as expected, cloud fraction profiles show a gradual decrease as the transition progresses with time.

Sugar-type mesoscale organizations typically occur during the daytime with shallow boundary layers, while flowers occur at night with deeper boundary layers (Vial et al., 2021). We use the cosine distance between features, a unique quantifiable distance metric derived from N1, to show the gradual development of the S2F transition inside the feature space (Fig. 3.c). The transformation appears smooth initially, with relatively more significant changes occurring later (post-18:00 UTC) as the system approaches the flower state. We link the relatively high changes in cosine distance during flower stages, as opposed to initial sugar stages, to the progression of convective developments. It becomes more accelerated as the system approaches the well-defined flower state.

Therefore, the framework reveals unbiased relative changes from the point of interest (in space or time) solely based on changes captured in high-dimensional feature space. Also, unlike previous works of Denby (2020) or Janssens et al. (2021), the intelligible partitioning of the continuum allows us to see when a particular system transitions to another. S7 provides insights into the transition probability of one class transforming to another over the Barbados domain.

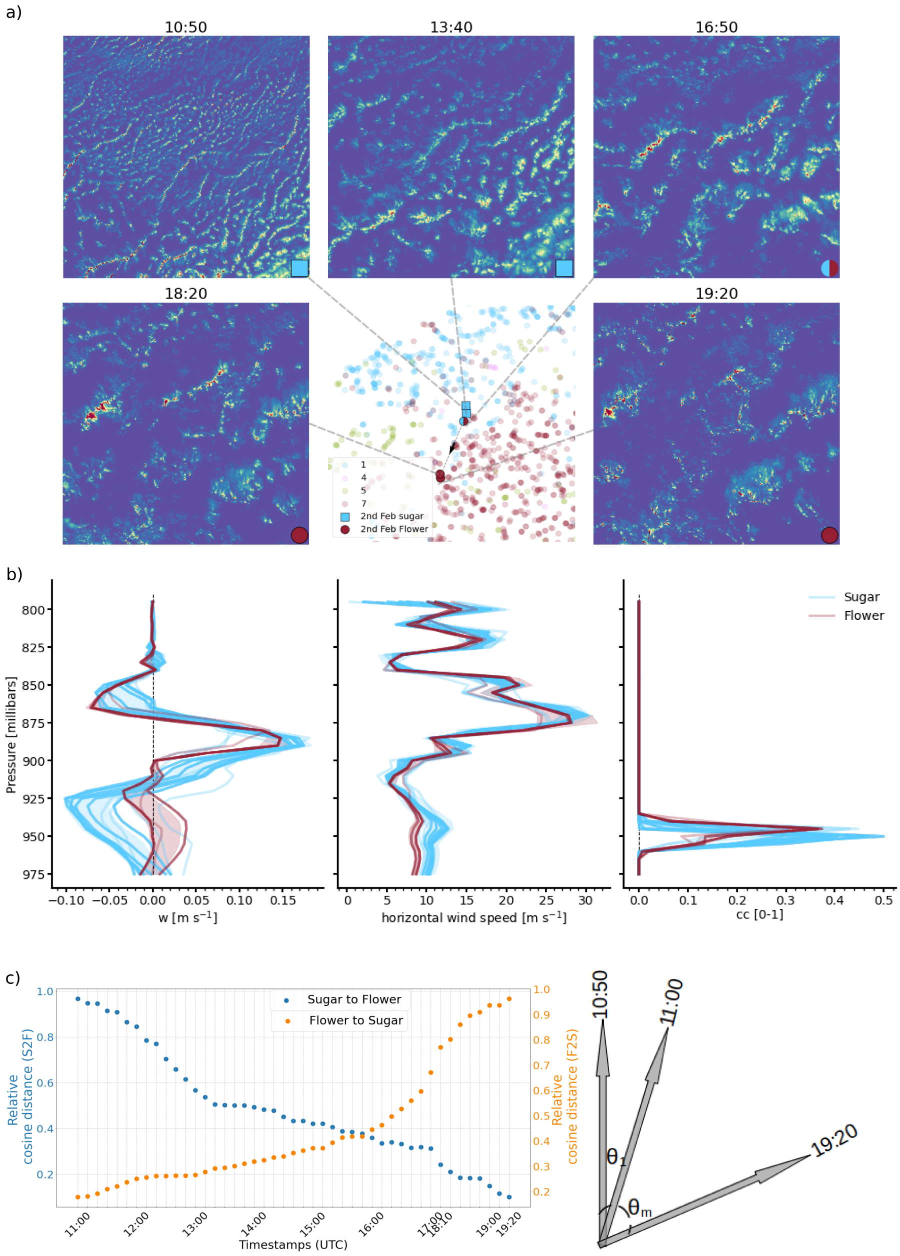


Figure 3. a) Five COD images covering the transition period between sugar and flower on the second of February 2020. Their position in the continuum is indicated in the center of the bottom row. b) Individual and standard deviation profiles of 1) vertical, 2) horizontal wind speed describing the atmospheric 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.

6 Conclusion

In this work, we develop a two-step self-supervised learning framework to study shallow convective organization properties and their transitions. By analyzing organization in a continuous approach without imposing predefined classes, we include all occurring patterns and transitional states in our analysis. Moreover, the approach shows that mesoscale cloud organizations in NAT can be partitioned into seven reasonable CRs for the time period considered. Exploiting the cloud amount at different vertical levels from CERES measurements, we show how the classes are interlinked with each other within the continuous space and thus capture the variability of tropical clouds in more detail.

We compare human-labeled cloud systems (Schulz, 2022) with the machine-identified partitions and underscore challenges in human-labeling of cloud organizations. Cloud systems with higher agreement among humans lie in the "correct" region of the feature space, while the ones with less consensus are in the "wrong" regions of the feature space. Also, the potential and interpretability of the continuum space become more evident when examining the classification and physical properties between human labels and their nearest neighbors.

Two of the seven CRs are strongly related to sugar and flower. Representing the S2F transition case study (Narenpitak et al., 2021) for February 2, 2020, in the continuum illustrates the capability to identify and represent the observed transformations smoothly in their clearly interpretable regions. We evaluate the transition's large-scale environmental parameters and observe a gradual increase in vertical wind speed and a gradual decrease in cloud amount. Finally, we showcase the use of cosine distance metric in capturing clear signatures of the S2F transition, which can help better understand cloud evolution. This is crucial for improving climate models and predicting how cloud behavior may change in a changing climate.

One of the limitations of this study is the use of only daytime cloud retrievals; hence, the organizations' nocturnal nature cannot be captured. Future studies will use infrared satellite measurements for 24-hour coverage. We aim to fine-tune our framework with ground-based observations of the EUREC^{4}A campaign and extend our analysis to a climate scale. The developed workflow could be a testing ground for investigating the newly adjusted subgrid parameterization effects in high-resolution global digital twins (Hoffmann et al., 2023) for mesoscale cloud systems or atmospheric processes at different scales.

7 Open Research

CERES, Edition-4A is available at (NASA et al., 2017), and ERA-5 reanalyses data (Hersbach et al., 2023) is available from the Copernicus climate change services. GOES-16 data has been accessed from the National Oceanic and Atmospheric Administration (NOAA), Climate Data Records (CDR) facility NOAA (2024b). Here, the COD retrieved using DCOMP algorithm (Walther & Heidinger, 2012) from GOES-16 measurements is available at NOAA (2024a). The code to produce this work and pre-trained weights of N1 and N2 can be accessed at Chatterjee, Schnitt, et al. (2023).

Acknowledgments

Dwaipayan Chatterjee's research was supported by the Federal Ministry for Environment, Nature Conservation, Nuclear Safety, and Consumer Protection. Claudia Acquistapace's (CA) research was funded by Deutsche Forschungsgemeinschaft (DFG). CA also acknowledges funding from Federal Ministry for Digital and Transport (BMDV).

References

Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space. In J. Van den Bussche & V. Vianu (Eds.), *Database theory — icdt 2001* (pp. 420–434). Berlin, Heidelberg: Springer Berlin Heidelberg.

Bebis, G., & Georgiopoulos, M. (1994). Feed-forward neural networks. *IEEE Potentials*, *13*(4), 27-31. doi: 10.1109/45.329294

Bony, S., Dufresne, J., Treut, H. L., Morcrette, J. J., & Senior, C. A. (2004). On dynamic and thermodynamic components of cloud changes. *Climate Dynamics*, *22*, 71-86. Retrieved from `https://api.semanticscholar.org/CorpusID:56077074`

Bony, S., & Dufresne, J.-L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, *32*(20). doi: 10.1029/2005GL023851

Bony, S., Schulz, H., Vial, J., & Stevens, B. (2020). Sugar, Gravel, Fish, and Flowers: Dependence of Mesoscale Patterns of Trade-Wind Clouds on Environmental Conditions. *Geophysical Research Letters*, *47*(7), e2019GL085988. doi: 10.1029/2019GL085988

Bony, S., Stevens, B., Ament, F., Bigorre, S., Chazette, P., Crewell, S., … Wirth, M. (2017, November). EUREC4A: A Field Campaign to Elucidate the Couplings Between Clouds, Convection and Circulation. *Surveys in Geophysics*, *38*(6), 1529–1568. doi: 10.1007/s10712-017-9428-0

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*(1), 1-27. Retrieved from `https://www.tandfonline.com/doi/abs/10.1080/03610927408827101` doi: 10.1080/03610927408827101

Chatterjee, D., Acquistapace, C., Deneke, H., & Crewell, S. (2023). Understanding cloud systems' structure and organization using a machine's self-learning approach. *Artificial Intelligence for the Earth Systems*, *2*(4), e220096. Retrieved from `https://journals.ametsoc.org/view/journals/aies/2/4/AIES-D-22-0096.1.xml` doi: 10.1175/AIES-D-22-0096.1

Chatterjee, D., Schnitt, S., Bigalke, P., Acquistapace, C., & Crewell, S. (2023, September). *Codes and pretrained weights for the study 'Capturing the diversity of mesoscale trade wind cumuli using complementary approaches from self- supervised deep learning'.* [Software] Zenodo. Retrieved from `https://doi.org/10.5281/zenodo.8352614` doi: 10.5281/zenodo.8352614

Dauhut, T., Couvreux, F., Bouniol, D., Beucher, F., Volkmer, L., Pörtge, V., … Wirth, M. (2023). Flower trade-wind clouds are shallow mesoscale convective systems. *Quarterly Journal of the Royal Meteorological Society*, *149*(750), 325–347. doi: 10.1002/qj.4409

Denby, L. (2020). Discovering the Importance of Mesoscale Cloud Organization Through Unsupervised Classification. *Geophysical Research Letters*, *47*(1), e2019GL085190. doi: 10.1029/2019GL085190

Denby, L. (2023). *Charting the realms of mesoscale cloud organisation using unsupervised learning.* doi: 10.48550/arXiv.2309.08567

Du, X. (2023). A robust and high-dimensional clustering algorithm based on feature weight and entropy. *Entropy*, *25*(3). Retrieved from `https://www.mdpi.com/1099-4300/25/3/510` doi: 10.3390/e25030510

Gilpin, S., Qian, B., & Davidson, I. (2013). Efficient hierarchical clustering of large high dimensional datasets. In *Proceedings of the 22nd acm international conference on information & knowledge management* (p. 1371–1380). New York, NY, USA: Association for Computing Machinery. Retrieved from `https://doi.org/10.1145/2505515.2505527` doi: 10.1145/2505515.2505527

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., … Thépaut, J.-N. (2023). *Era5 hourly data on pressure levels from 1940 to present.* [dataset] Copernicus Climate Change Service (C3S) Climate Data Store (CDS). doi: 10.24381/cds.bd0915c6

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J.-N. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi:

https://doi.org/10.1002/qj.3803

Hoffmann, J., Bauer, P., Sandu, I., Wedi, N., Geenen, T., & Thiemert, D. (2023). Destination earth – a digital twin in support of climate services. *Climate Services*, *30*, 100394. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2405880723000559` doi: https://doi.org/10.1016/j.cliser.2023.100394

Huynh, T., Kornblith, S., Walter, M. R., Maire, M., & Khademi, M. (2022). *Boosting contrastive self-supervised learning with false negative cancellation.*

Janssens, M., Vilà-Guerau de Arellano, J., Scheffer, M., Antonissen, C., Siebesma, A. P., & Glassmeier, F. (2021). Cloud Patterns in the Trades Have Four Interpretable Dimensions. *Geophysical Research Letters*, *48*(5), e2020GL091001. doi: 10.1029/2020GL091001

Muller, C. J., & Held, I. M. (2012, August). Detailed Investigation of the Self-Aggregation of Convection in Cloud-Resolving Simulations. *Journal of the Atmospheric Sciences*, *69*(8), 2551–2565. (Publisher: American Meteorological Society Section: Journal of the Atmospheric Sciences) doi: 10.1175/JAS-D-11-0257.1

Narenpitak, P., Kazil, J., Yamaguchi, T., Quinn, P., & Feingold, G. (2021). From Sugar to Flowers: A Transition of Shallow Cumulus Organization During ATOMIC. *Journal of Advances in Modeling Earth Systems*, *13*(10), e2021MS002619. doi: 10.1029/2021MS002619

NASA, LARC, SD, & ASDC. (2017, 9 19). *Ceres and geo-enhanced toa, within-atmosphere and surface fluxes, clouds and aerosols 1-hourly terra-aqua edition4a.* NASA Langley Atmospheric Science Data Center DAAC [Dataset]. Retrieved from `https://doi.org/10.5067/TERRA+AQUA/CERES/SYN1DEG-1HOUR_L3.004A`

NOAA. (2024a). *Noaa goes-16 cloud optical depth and cloud particle size distribution (codc) [dataset].* Retrieved from `https://noaa-goes16.s3.amazonaws.com/index.html#ABI-L2-CODC/` (Accessed: 2024-05-14)

NOAA. (2024b). *Noaa oceanic climate data records [dataset].* Retrieved from `https://registry.opendata.aws/noaa-cdr-oceanic/` (Accessed: 2024-05-14)

Nuijens, L., & Siebesma, A. P. (2019, June). Boundary Layer Clouds and Convection over Subtropical Oceans in our Current and in a Warmer Climate. *Current Climate Change Reports*, *5*(2), 80–94. doi: 10.1007/s40641-019-00126-x

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65. Retrieved from `https://www.sciencedirect.com/science/article/pii/0377042787901257` doi: https://doi.org/10.1016/0377-0427(87)90125-7

Sarkar, M., Zuidema, P., Albrecht, B., Ghate, V., Jensen, J., Mohrmann, J., & Wood, R. (2020). Observations pertaining to precipitation within the northeast pacific stratocumulus-to-cumulus transition. *Monthly Weather Review*, *148*(3), 1251 - 1273. doi: 10.1175/MWR-D-19-0235.1

Schmit, T. J., Gunshor, M. M., Menzel, W. P., Gurka, J. J., Li, J., & Bachmeier, A. S. (2005). Introducing the next-generation advanced baseline imager on goes-r. *Bulletin of the American Meteorological Society*, *86*(8), 1079 - 1096. Retrieved from `https://journals.ametsoc.org/view/journals/bams/86/8/bams-86-8-1079.xml` doi: https://doi.org/10.1175/BAMS-86-8-1079

Schneider, T., Kaul, C., & Pressel, K. (2019, 03). Possible climate transitions from breakup of stratocumulus decks under greenhouse warming. *Nature Geoscience*, *12*, 164-168. doi: 10.1038/s41561-019-0310-1

Schubert, E. (2023, jun). Stop using the elbow criterion for k-means and how to choose the number of clusters instead. *ACM SIGKDD Explorations Newsletter*, *25*(1), 36–42. doi: 10.1145/3606274.3606278

Schulz, H. (2022). C^3ontext: a common consensus on convective organization during the eurec^{4}a experiment. *Earth System Science Data*, *14*(3), 1233–1256. doi: 10.5194/essd-14-1233-2022

Schulz, H., Eastman, R., & Stevens, B. (2021). Characterization and Evolution of Organized Shallow Convection in the Downstream North Atlantic Trades. *Journal of Geophysical Research: Atmospheres*, *126*(17), e2021JD034575. doi: 10.1029/2021JD034575

Steinbach, M., Ertöz, L., & Kumar, V. (2004). The challenges of clustering high dimensional data. In L. T. Wille (Ed.), *New directions in statistical physics: Econophysics, bioinformatics, and pattern recognition* (pp. 273–309). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from `https://doi.org/10.1007/978-3-662-08968-2_16` doi: 10.1007/978-3-662-08968-2_16

Stevens, B., Bony, S., Brogniez, H., Hentgen, L., Hohenegger, C., Kiemle, C., … Zuidema, P. (2020). Sugar, gravel, fish and flowers: Mesoscale cloud patterns in the trade winds. *Quarterly Journal of the Royal Meteorological Society*, *146*(726), 141–152. doi: 10.1002/qj.3662

Stevens, B., Bony, S., Farrell, D., Ament, F., Blyth, A., Fairall, C., … Zöger, M. (2021, August). EUREC^{4}A. *Earth System Science Data*, *13*(8), 4067–4119. (Publisher: Copernicus GmbH) doi: 10.5194/essd-13-4067-2021

Stevens, B., Farrell, D., Hirsch, L., Jansen, F., Nuijens, L., Serikov, I., … Prospero, J. M. (2016). The barbados cloud observatory: Anchoring investigations of clouds and circulation on the edge of the itcz. *Bulletin of the American Meteorological Society*, *97*(5), 787 - 801. doi: https://doi.org/10.1175/BAMS-D-14-00247.1

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, *9*(86), 2579–2605. Retrieved from `http://jmlr.org/papers/v9/vandermaaten08a.html`

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2023). *Attention is all you need.* Retrieved from `https://arxiv.org/abs/1706.03762`

Vial, J., Vogel, R., & Schulz, H. (2021). On the daily cycle of mesoscale cloud organization in the winter trades. *Quarterly Journal of the Royal Meteorological Society*, *147*(738), 2850–2873. doi: 10.1002/qj.4103

Vogel, R., Konow, H., Schulz, H., & Zuidema, P. (2021, November). A climatology of trade-wind cumulus cold pools and their link to mesoscale cloud organization. *Atmospheric Chemistry and Physics*, *21*(21), 16609–16630. (Publisher: Copernicus GmbH) doi: 10.5194/acp-21-16609-2021

Walther, A., & Heidinger, A. K. (2012). Implementation of the daytime cloud optical and microphysical properties algorithm (dcomp) in patmos-x. *Journal of Applied Meteorology and Climatology*, *51*(7), 1371 - 1390. doi: 10.1175/JAMC-D-11-0108.1

Wielicki, B., Barkstrom, B., Harrison, E., Lee, R., Smith, G., & Cooper, J. (1996, 05). Clouds and the earth's radiant energy system (ceres): An earth observing system experiment. *Bulletin of the American Meteorological Society*, *77*, 853-868. doi: 10.1175/1520-0477(1996)077⟨0853:CATERE⟩2.0.CO;2

References From the Supporting Information

Aggarwal, C. C., Hinneburg, A., Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space. In J. Van den Bussche V. Vianu (Eds.), Database theory — icdt 2001 (pp. 420–434). Berlin, Heidelberg: Springer Berlin Heidelberg.

Bottou, L. (2012). Stochastic gradient descent tricks. In G. Montavon, G. B. Orr, K.R. Müller (Eds.), Neural networks: Tricks of the trade: Second edition (pp. 421–436). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-35289-8 25

Bridle, J. S. (1989). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In Nato neurocomputing. Retrieved from `https://api.semanticscholar.org/CorpusID:59636530`

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A. (2021). Emerging properties in self-supervised vision transformers.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale.

Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. Biological Cybernetics, 20 (3-4), 121–136. Retrieved 2022-08-30, from `http://link.springer.com/10.1007/BF00342633doi:10.1007/BF00342633`

Gao, J., Li, F., Wang, B., Liang, H. (2021). Unsupervised nonlinear adaptive manifold learning for global and local information. Tsinghua Science and Technology, 26 (2), 163-171. doi: 10.26599/TST.2019.9010049

Goyal, P., Duval, Q., Reizenstein, J., Leavitt, M., Xu, M., Lefaudeux, B., Misra, I. (2021). Vissl. https://github.com/facebookresearch/vissl.

He, K., Zhang, X., Ren, S., Sun, J. (2015, December). Deep Residual Learning for Image Recognition. Retrieved 2022-08-30, from `http://arxiv.org/abs/1512.03385`

Hendrycks, D., Gimpel, K. (2023). Gaussian error linear units (gelus). Retrieved from `https://arxiv.org/abs/1606.08415`

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., Shah, M. (2022, jan). Transformers in vision: A survey. ACM Computing Surveys, 54 (10s), 1–41. doi: 10.1145/3505244

Nie, Y., Zamzam, A. S., Brandt, A. (2021). Resampling and data augmentation for short-term PV output prediction based on an imbalanced sky images dataset using convolutional neural networks. Solar Energy, 224, 341-354. doi: https://doi.org/10.1016/j.solener.2021.05.095

Paletta, Q., Terrén-Serrano, G., Nie, Y., Li, B., Bieker, J., Zhang, W., . . . Feng, C. (2023). Advances in solar forecasting: Computer vision with deep learning. Advances in Applied Energy, 11, 100150. doi: https://doi.org/10.1016/j.adapen.2023.100150

Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986, October). Learning representations by back-propagating errors. Nature, 323 (6088), 533–536. Retrieved from `https://doi.org/10.1038/323533a0` doi: 10.1038/323533a0

Tenenbaum, J. B., de Silva, V., Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. Science, 290 (5500), 2319-2323. Retrieved from https://www.science.org/doi/abs/10.1126/science.290.5500.2319 doi: 10.1126/science.290.5500.2319

van der Maaten, L., Hinton, G. (2008). Visualizing data using t-sne. Journal of Machine Learning Research, 9 (86), 2579–2605.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I. (2023). Attention is all you need. Retrieved from `https://arxiv.org/abs/1706.03762`