Key Points:
- A neural network can learn without supervision to discover similar cloud structures in satellite imagery
- The groups of cloud structures produced by the neural network have distinct radiative properties and thus distinct climatic impacts
- The neural network learns to represent similarity between cloud structures allowing for the study of the temporal evolution between regimes

Correspondence to:
L. Denby,
l.c.denby@leeds.ac.uk

Discovering the Importance of Mesoscale Cloud Organization Through Unsupervised Classification

L. Denby[1]

[1]School of Earth and Environment,University of Leeds, Leeds, UK

Abstract The representation of shallow trade wind convective clouds in climate models dominates the uncertainty in climate sensitivity estimates. In particular the radiative impact of cloud spatial organization is poorly understood. This work presents the first unsupervised *neural network* model which autonomously discovers cloud organization regimes in satellite images. Trained on 10,000 GOES-16 satellite images (tropical Atlantic and boreal winter) the regimes found are shown to exist in a hierarchy of organizational scales, with sub-clusters having distinct radiative properties. The model requires no time-consuming and subjective hand-labeled data based on predefined structures allowing for objective study of very large data sets. The model enables the study of environmental conditions in different organizational regimes and in transitions between regimes and objective comparisons of model behavior with observations through cloud structures emerging in both. These abilities enable the discovery of previously unknown physical relationships in cloud processes, enabling better representation of clouds in weather and climate simulations.

1. Introduction

By eye, it is clear from satellite imagery that clouds often organize into patterns (fronts, cyclones, cellular cloud decks, etc). Each of these affect the Earth's weather and climate in different ways, through their precipitation and interaction with radiation and atmospheric circulation.

Clouds of one particular form, the shallow trade wind cumuli, substantially impact the Earth's climate due to their ubiquity (Bony et al., 2004) and their overall cooling effect, stemming from the fact that they reflect more of the Sun's short-wave radiation than the long-wave radiation emanating from the Earth's surface. Differing predictions, between different climate models, for how these clouds will respond to a warming climate account for most of the variation in climate sensitivity between models (Bony & Dufresne, 2005; Medeiros et al., 2008; Vial et al., 2013; Webb et al., 2006) and highlight an urgent need to better understand how these clouds form and interact with their environment. This is further demonstrated by the World Climate Research Programme making *the interplay between clouds, atmospheric circulation, and climate sensitivity* (Bony et al. 2015) one of their Grand Science Challenges.

These shallow clouds organize into many forms of mesoscale (meso-β, 20–200 km, Orlanski, 1975) structures, and based on Large-Eddy Simulations (LES), it appears that this spatial organization may play a key role in determining the cloud fraction (Seifert and Heus, 2013; Vogel et al., 2016), which may in turn affect their radiative impact (as has been shown to be the case for deep convection Tobin et al., 2012). However, which forms of organization occur in nature and what drives each is still poorly understood, leading to a difficulty in representing these clouds in climate models. Motivated by this challenge (and the upcoming EUREC4A field-campaign aiming to study these clouds, Bony et al., 2017), recent work by (Stevens et al., 2019) manually identified and named four distinct regimes of frequently occurring patterns: Sugar, Flower, Fish, and Gravel. This work was extended in (Rasp et al., 2019) by training a neural network to recognize these four specific forms of organization using $\approx 30,000$ hand-labeled training examples (using ≈ 250 man-hours) and allowed for the production of a global spatial map of their relative occurrence.

In current work it is not the aim to identify predefined cloud structures but instead to construct a computational model (a neural network) which can without *supervision* discover which structures exist in a given image and thereby group images containing similar structures. Instead of predicting which structure is most likely present in each input image, the model maps these images into a space where similarity between

structures can be studied. This is thus a form of *unsupervised* learning, where the model is not informed what the *correct* answer is (as would be the case for *supervised* learning) but rather the similarity in answers between different inputs is used to constrain the model's behavior.

The method by which the neural network is produced will be detailed in section 2, together with some detail of how neural networks are able to achieve such a task. The ability of the network to identify groups of similar cloud structures will be investigated in section 3, first by studying the clustering that the model produces and then by showing that the identified cloud structures have different physically relevant (radiative and morphological) properties.

2. Methods

When we see patterns in satellite images and start grouping cloud structures in our mind, what we are implicitly doing is grouping the images by visible physical characteristics (e.g., cloud size, cloud -cover, and fractal dimension). This grouping by visual characteristics amounts to placing each image into a high-dimensional space with the dimensions mapped out by the characteristics we are observing.

In a similar way the aim in this work is to produce a computational model, a neural network (detailed below), which, through training, will learn to group similar image tiles (the pixel intensities forming the input, $\mathbf{x}$) into a small neighborhood in N_d-dimensional space. Mathematically this amounts to devising a function (F_{nn}, detailed below) which for an input image $\mathbf{x}_i$ produces an N_d-dimensional vector $\mathbf{y}_{N_d} = F_{nn}(\mathbf{x}_i)$.

In machine learning nomenclature this type of output y_{N_d} is called an *embedding*, a lower-dimensional vector representation of the input data which has the desired properties that a given application might require. An example is the seminal `word2vec` paper (Mikolov et al., 2013), which introduced the concept of embeddings for words; individual words are represented as points in high-dimensional space and arithmetic with these points produces predictive relationships. For example

$$\begin{aligned} F_{nn}(''santa'') - F_{nn}(''christmas'') &\approx F_{nn}(''man''), \\ F_{nn}(''london'') - F_{nn}(''england'') &\approx F_{nn}(''copenhagen'') - F_{nn}(''denmark''). \end{aligned} \tag{1}$$

Training a model to produce an embedding is a form of unsupervised learning, which means, as mentioned previously, that the model is not trained to pick among a finite set of predefined answers (in this case types of cloud structures) but is free to utilize the embedding space to achieve the training objective (here grouping images with similar cloud structures together). In the current work the dimensions in the embedding space are automatically utilized by the model to create the required separation in N_d-dimensional space in order to represent the structural differences between images as necessary.

The unsupervised nature of the training presents a key challenge when interpreting the output of the model as the components of the embeddings do not have an a priori meaning. Instead, the embeddings produced must be studied to identify what relationships the model has learnt. The extent to which the model has produced clusters in the embedding space will be studied in section 3; interpretation of the individual embedding dimensions will be left for future work.

The specific type of computational model used in this work to produce image embeddings is a deep neural network, the application of which has recently revolutionized computer vision applications (see, e.g., Krizhevsky et al. (2012)). A neural network is mathematically the sequential application of linear and non-linear transformations ($f_1, f_2, f_3, ..., f_N$) to an input vector $\mathbf{x}$, to produce an output vector

$$\mathbf{y} = F_{nn}(\mathbf{x}) = f_N(...f_3(f_2(f_1(\mathbf{x})))).$$

These transformations make up the so-called *layers* of the network, and the notion of a *deep* neural network comes from the large number of layers used. Each layer transforms the previous layer's output vector, either through multiplication by a matrix (the contents, or *weights*, of which must be *learnt*) or through scaling by a non-linear function (e.g., $max(x, 0)$). Based on the output $\mathbf{y}$ of the neural network, a loss function ($L(\mathbf{y})$, see below) is formulated, and learning is achieved by incrementally altering the matrix contents (*weights*) so as to minimize the loss function. Mathematically this is done by applying the chain-rule to expand how the loss
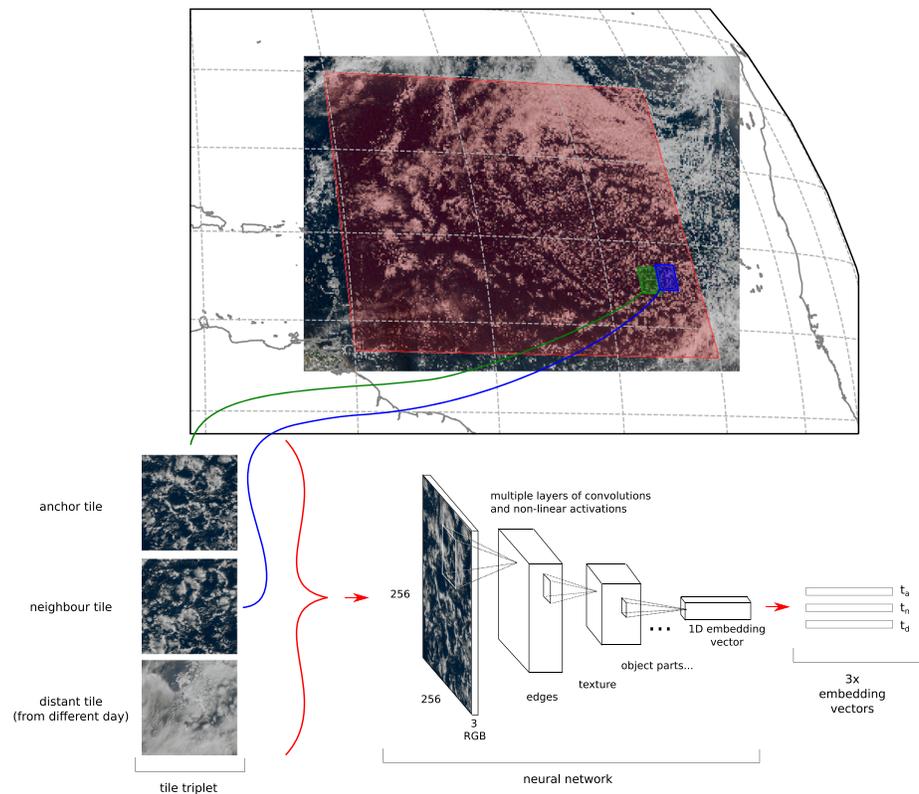
Figure 1. Overview of tile generation and neural network model training process. Top: map of the studied domain (red box) in tropical Atlantic with example *anchor* (green) and *neighbor* (blue) tiles from a single triplet highlighted (*distant* tile sampled from same domain on different day). Bottom: for each tile in the triplet the neural network produces a 1D embedding vector. The loss function ensures that *anchor* and *neighbor* tiles are close in embedding space compared to *distant* tile.

function changes given changes in the matrix weights for each training example (**x**) seen (see Goodfellow et al., 2016, for more details).

Research has shown which transformations, when applied in grouped succession, work well for particular applications. For the present work, a Convolution Neural Network (CNN) is used as these have been very successful in computer vision applications. In a CNN, successive layers build increasingly higher abstractions (edges, arcs, composite shapes, etc.) from compositing spatially localized gradient calculations over multiple layers. The particular type of CNN used in this work is a so-called residual network, ResNet-34 (implemented in the `pytorch` framework, Paszke et al., 2017).

The ability of the neural network to cluster similar images in the embedding space comes from how the neural network is trained (see figure 1 for a schematic of the training process). Instead of each training example consisting of only one input, three image tiles (a triplet) are used, and the embeddings produced from each of these three images are compared. This technique was pioneered by (Jean et al., 2018). The tiles of the triplet are sampled from satellite imagery, in such a way that two are likely to contain very similar cloud structures (the *anchor* t_a and *neighbor* t_n tiles) whereas the third (the *distant* tile, t_d) is likely to contain very different cloud structure compared to the former two. This is achieved by picking the *anchor* and *neighbor* tiles from the same image so that they are spatially close (here we use $\approx$ 50% overlap by displacing the neighbor tile's center half a tile width in a random direction from the anchor tile's center), and the *distant* tile is picked at a random location on a different day (this is in contrast to Jean et al. (2018) which used the same source image for the distant tile but sampled at a predefined distance to the anchor tile). The comparison of the three embeddings is formulated in the loss function $L(t_a, t_n, t_d)$, which encourages similar

embeddings for the *anchor* and *neighbor* but penalizes similar embedding for the *anchor* and *distant* tiles. This can mathematically be written as

$$L(t_a, t_n, t_d) = max(||F_{nn}(t_a) - F_{nn}(t_n)||_2 - ||F_{nn}(t_a) - F_{nn}(t_d)||_2, 0) + m,$$

where $||...||_2$ denotes the L2-norm (Euclidean distance) and m is the *margin* (setting the desired separation between the anchor and distant tile). We use $m = 1.0$.

The input image tiles (256×256 pixel) were sampled from RGB composites, which were generated with "satpy" (Raspaud et al., 2019) from radiance channels 1, 2, and 3 from GOES-16 observations. In this work, as the primary focus was trade wind cumulus clouds, tiles were generated from a domain spanning across the Tropical Atlantic (60° W to 30° W and 10° N to 30° N) during the boreal winter. Three months between 1 November 2018 and 31 January 2019, using observations spanning an hour over local zenith at 60° W (four sets of observations per day). From the four satellite images per day one random image per day was retained to use for studying the clustering of the model (the *study set*), the rest were used for training (*training set*). The tile size was set to $L_t = 200$ km, so that any cloud structures up to the meso-β scale could be sampled.

The embedding length was set to $N_d = 100$, as this provided the neural network with adequate space to satisfy the loss function by creating the necessary clustering in the embedding space. Below $N_d = 20$, the training speed decreased and inter-cluster variance in both cloud structures seen and metrics of their properties (section 3) increased. While fewer embedding dimensions may have been sufficient, the focus of this work was on creating a functional rather than optimal model. One-cycle learning (Smith, 2018) with adaptive learning rate was used, with max rate ($\lambda = 0.01$) identified from the largest loss rate.

3. Results

In this section, the neural network's ability to group tiles with similar cloud structures into similar regions of the embedding space will be shown. This is done first by demonstrating how the clustering can be studied and visualized, using hierarchical clustering and second by investigating the extent to which the neural network produces clusters of tiles with different radiative properties. The results shown here are based on first training the neural networking using 10,000 triplets (generated as detailed in section 2) and using the network to produce embeddings for 1,000 tiles produced from satellite imagery spanning the same period and domain but were not included in the training set.

3.1. Clustering of Cloud Structures

To assess the ability of the neural network to group images with similar cloud structures, traditional measures such as cloud cover and cloud size will be used in this section to subjectively describe and compare the clusters produced. A quantitative analysis of the clusters will be given through their radiative properties in subsection 3.2.

In figure 2 a *so-called* dendrogram (Jones et al., 2001) displays the outcome of performing hierarchical clustering on the 1,000 embeddings produced by the trained neural network model. In hierarchical clustering points are incrementally merged into progressively larger clusters, each merge combining the two clusters which minimize a given clustering metric. Here the *Ward* metric, which aims to minimize the intra-cluster variance in distance between embeddings (measured using the Euclidean distance as in the loss function used to train the model), is used. The last 12 merges are shown for brevity, with each of the resulting clusters labeled using a letter (A to L) and the number of tiles in each cluster. To demonstrate the type of cloud structure represented by each cluster, 12 random tiles from each cluster have been rendered below each of the leaf nodes of the dendrogram.

The vertical axis in the dendrogram measures the intra-cluster variance after each merge, and thus, the length of the vertical vertices connecting each merge indicates the similarity between clusters (larger intra-cluster variance meaning larger variance in distance between points in the embedding space). This similarity measure shows that the model separates some groups more strongly than others, based on common characteristics between clusters. For example, clusters A and B are much more similar to each other than to any other cluster, supported visually by the dominance of small scattered clouds as compared to the other clusters. This can be similarly observed for the smaller cellular structures in clusters G and H and to some extent the larger broken structures in clusters J, K, and L.
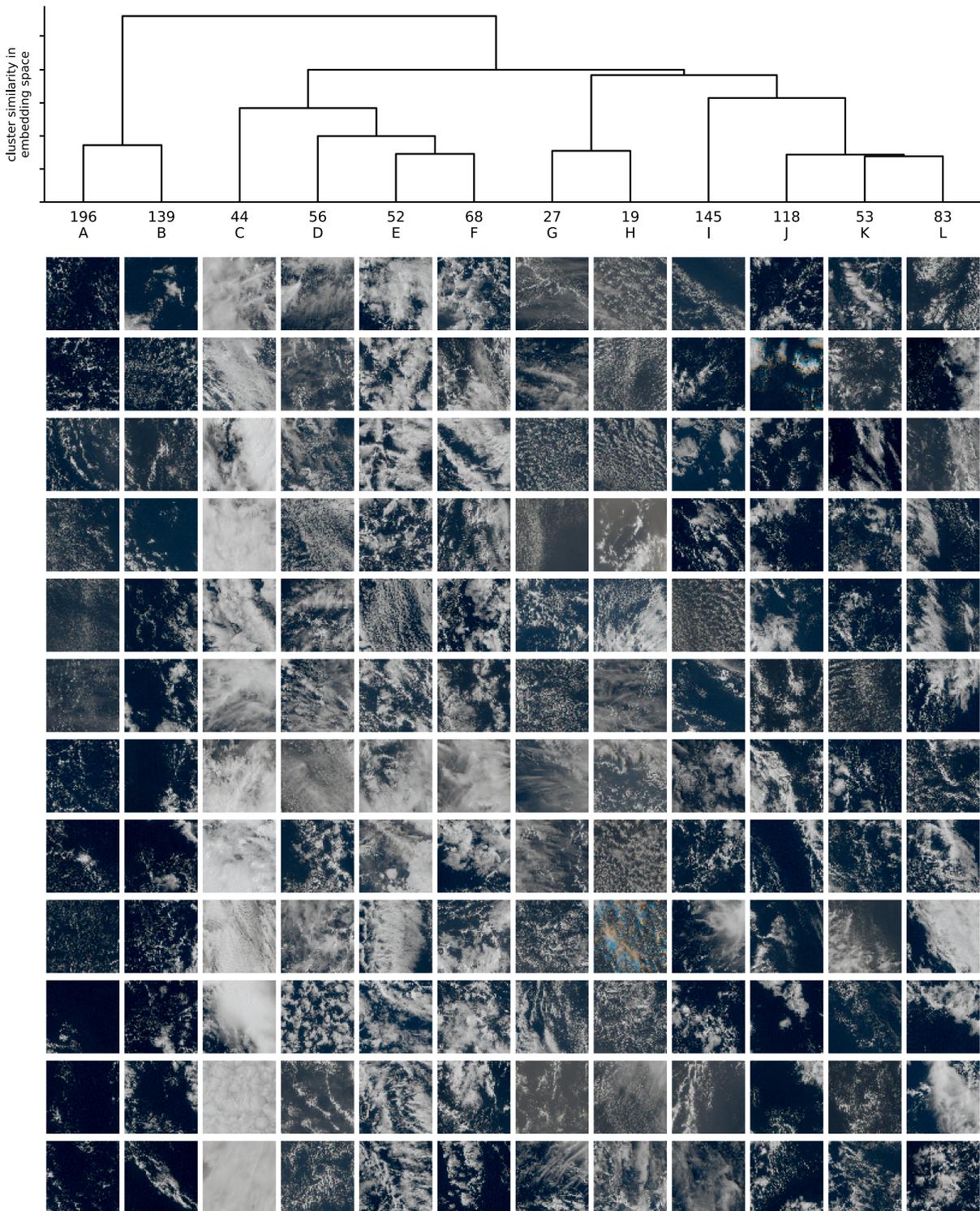
Figure 2. Demonstration of clustering for embeddings produced by the trained neural network. Top: dendrogram showing hierarchical structure of the clustering, with height representing intra-cluster variance in embedding distance in a cluster resulting from a specific merge of two child clusters. Only the clusters present before the last 12 merges are showed for brevity. Bottom: 12 random tile examples from each cluster belonging to the leaf immediately above in the dendrogram. Each leaf node cluster is annotated with the number of tiles in that cluster and a label to aid discussion. The persistence in the dendrogram indicates for example clusters A and B are much more similar to each other than to any other cluster. A number of visibly distinct structures have been identified, for example, scattered small clouds (A, B) cellular structures (C, D, G, and H in order of scale) and larger cloud cellular (D–F) and broken (I–L) cloud structures.

In general clusters A, B, G, and H contain smaller cloud structures as compared to C–F and I–L. Clusters A and B are distinguished from G and H by the latter containing a more regular cellular pattern of clouds where clouds in A and B are more scattered. Similarly D–F contain more cellular structures as compared to the broken larger clouds in I–L. Another feature that separates the different clusters is the amount of cloud cover, for example, clusters A and B have very little cloud, whereas cluster C has almost complete cloud cover. The aim here is not to quantify exactly what features the model has used to create different clusters, as these visible features are likely not independent but simply to describe the clusters and assert that they are in fact different. Although by eye the model may appear to have misclassified some images the strength of the model lies in applying it effortlessly to very large data sets and providing insight through the statistical properties of each cluster identified (as will be shown in the next section).

At first glance it appears that another feature used by the model is the presence or absence of a light-gray shading (which is likely high-level dust being swept from the coast of Africa) seen in G and H. However, the fact that the clusters G and H contain both tiles with and without shading suggests that the model is in fact able to discern the underlying structure even when it is obscured. The prevalence of a dust-like overlay in clusters G and H may instead indicate that the cloud structures in these clusters are more common in regions where dust is prevalent.

The relative number of tiles in each cluster provides an indication of the relative frequency of different cloud structures in the studied region. It suggests that scattered small clouds (clusters A and B) are much more common than full cloud cover with occasional cellular structures (cluster C). A systematic study of the climatology of different cloud structures will be carried out in future work.

The hierarchical clustering used here becomes a means to study the *continuum* of different cloud structures, which to a lesser extent is dominated by discrete forms of cloud structure but instead can be viewed as different types of structures and sub-types within these and so on. Depending on the specific cloud type of interest, a branch of the dendrogram may be isolated and studied further through further sub-clustering, adding more training data to improve the statistics where necessary.

Figure 2 demonstrates that the neural network *without supervision* discovers different types of cloud structures and groups images with similar features together in the embedding space. As there was no pre-selection to determine which tiles to use and so many tiles contain many different scales of cloud structures, it is remarkable how well the clustering works. Through experimentation with training data used, model architecture (depth), embedding length and further development of means to study the embedding space, this can be used to study cloud structures across a hierarchy of scales in any region across the globe.

3.2. Radiative Properties of Clouds

A key aspect of clouds is their impact on the Earth's radiation budget and so showing that different clusters have different cloud radiative properties, in particular their differing effects on outgoing short-wave and long-wave radiation, provides a clear measure of merit for the neural network based unsupervised classification. At the time of writing, no long-wave (LW) and short-wave (SW) products were available for GOES-16 and so the raw band-centered radiances are utilized here by choosing bands representative of SW and LW radiation. For SW channel 1 ($f_1 = 0.47$ μm) which is in the visible frequency range is used and for LW channel 9 ($f_9 = 6.95$ μm) in the long-wave infrared range is used. Although channel 1 was also used in generation of composite RGB images used for training the model, the image tiles used to study the behavior of the model here have not been seen by the model.

Figure 3 plots the per-tile mean for all 1,000 tiles in the *study* data set. Alongside this, the per-cluster mean is plotted for the tiles belonging to each cluster shown in figure 2, with the error in the mean for each cluster represented by x-y error bars. For each of these clusters, the tile nearest to the mean (distance normalized by the variance in the two channels) is added as an annotation, so that the radiative properties of the clusters may be related to the cloud structures.

Although there is a large degree of overlap in the radiative properties of the clusters identified with the unsupervised neural network their mean radiative properties (and so climatic impact) are clearly distinct. If this was not the case, the error bars for each cluster would overlap across clusters, which, although not shown in this work, was verified by randomly shuffling which tiles were assigned to each cluster.

The separation in radiative properties shows that with increasing cloud cover there is in general a decreasing amount of long-wave emissions to space and an increase of short-wave emissions. As well as this general
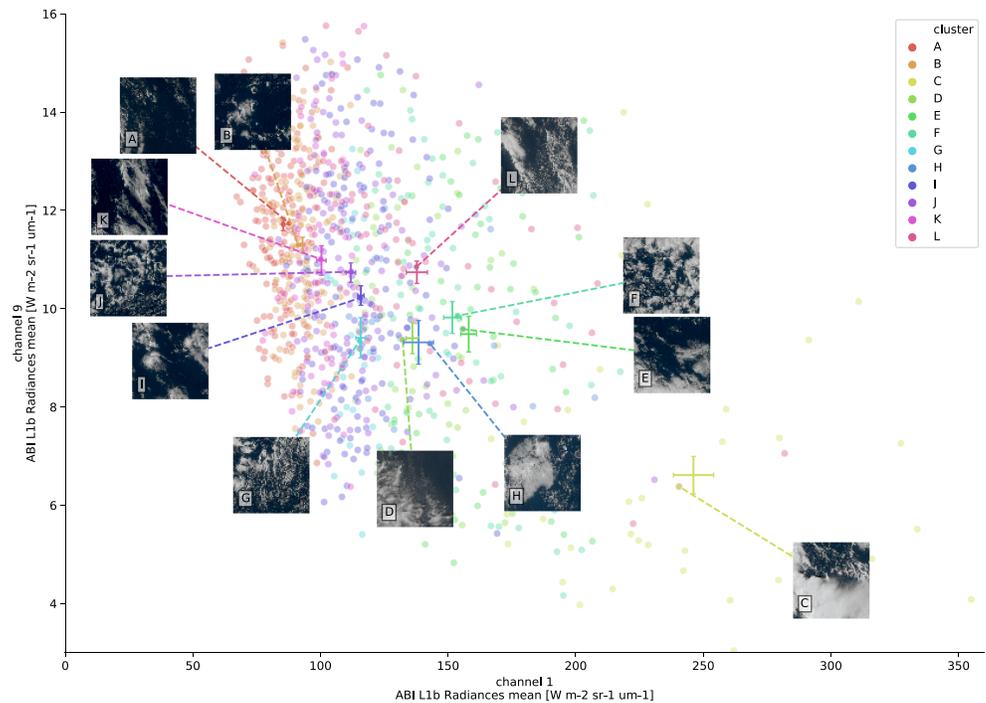
Figure 3. Per-tile mean radiance of channels 1 (in short-wave range) and 9 (in long-wave range) for all 1000 tiles in *study* set together with the per-cluster mean and error (standard error of the mean) across all tiles in each cluster produced by hierarchical clustering (figure 2) colored by cluster. Each cluster is annotated with the nearest tile. The clusters show a clear separation between different cloud structures, more cloudy tiles having less long-wave and more short-wave emission is radiated into space, and each cluster has distinct radiative properties.

trend, the cellular structure and organization of clouds also appear to be important (compare for example clusters G and H), this will be studied further and quantified in further work.

The fact that each cluster containing tiles with similar cloud structures has specific radiative properties suggests that with the neural network it is possible to automatically identify and group cloud structures by physically meaningful properties. As mentioned above the further study of each cluster can simply be achieved through providing more training and test data from the ample satellite observations available.

3.3. Spatial Structure and Organization of Clouds

To further support that the neural network has produced clustering in the embedding space representing distinct organizational states, the I_{org} spatial organization metric (Tompkins and Semie, 2017) and Minkowski-Bouligand fractal dimension D_{MB} (Falconer, 2004) were calculated for all tiles in the study set (plotted by color figure 4 with mean and standard error of the mean for each cluster). I_{org} quantifies whether the spatial distribution of clouds is regular ($I_{org} < 0.5$), random ($I_{org} = 0.5$), or organized (I_{org}), by comparing the nearest-neighbor cumulative distribution function to that of a random process. D_{MB} quantifies the spatial form of clouds by measuring the fractal dimension, expected to be between that of a line ($D_{MB} = 1$) and a plane ($D_{MB} = 2$), by counting the number of successively smaller boxes that cover the cloud mask. Both metrics require that a cloud mask is defined; here the tile RGB images were converted to gray-scale, and 50% intensity was used.

The clusters of cloud types identified by the neural network are seen to occupy different parts of the I_{org}-D_{MB} space, and the cluster means are clearly separated, affirming that the neural network has learnt distinct cloud regimes. Comparing the tiles highlighted for each cluster, as expected, more elongated and larger structures have larger fractal dimension and more regularly spaced cloud patterns have smaller I_{org}. Comparing the cluster means (e.g., clusters G, I, and J or clusters A, J, and F) it is clear that some forms of cloud structures can only be distinguished by considering both metrics together. This highlights a further limitation of applying existing metrics of organization based on inter-cloud distance (from cloud centers), to
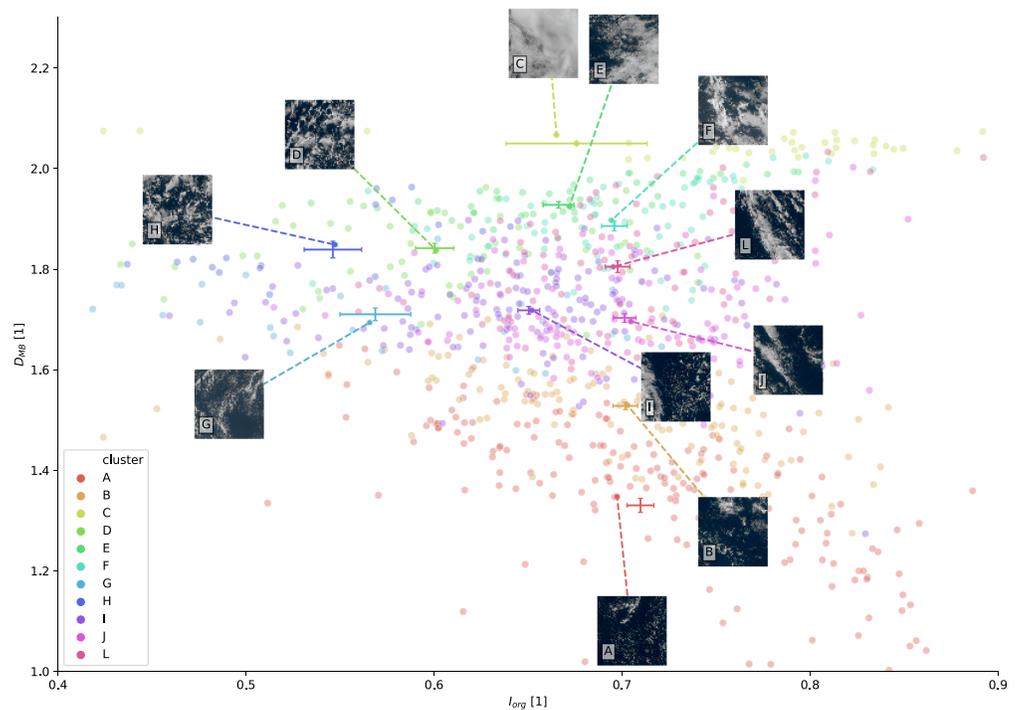
Figure 4. Cloud organization (I_{org}) and fractal dimension (D_{MB}) for all 1,000 tiles in *study* set produced by hierarchical clustering (figure 2), colored by cluster, together with per-cluster mean and standard error in the mean. Each cluster is annotated with nearest tile.

quantify the type of cloud structures present, as these do not take into account the spatial form (e.g., fractal dimension) of clouds. Only by producing more complicated metrics (with further inherent thresholds) or by applying a neural network which autonomously learns which spatial features to extract and use for clustering can full identification of distinct regimes be achieved.

4. Conclusions

This work demonstrates that a neural network can, without labeled training data, automatically discover different forms of cloud organization and through this learn to group input images containing similar cloud structures together. This was demonstrated by first training a neural network with 10,000 triplets of input image tiles sampled from 3 months of GOES-16 data and then studying the clustering of predictions produced by the network on 1,000 image tiles not seen during training.

Using hierarchical clustering, it was shown that not only does the model identify different cloud structures and group images with similar structures but also generates a representation where the similarity between different structures can be studied. Through the use of two channels of the GOES-16 radiance measurements in the long- and short-wave frequency range, as well as common metrics of cloud structure, it was shown that different cloud structures have distinct radiative and morphological properties. This indicates that the neural network discovers cloud structures which have unique physical characteristics.

To the author's knowledge, this work presents the first unsupervised approach to the identification of cloud structures from satellite imagery. The benefits of using an unsupervised model are numerous, stemming from the fact that the model (1) can be applied to any spatial data set with limited effort as no hand-labeling is required; (2) automatically discovers the types of structures present in the input; and (3) produces a representation of the similarity between these structures. These properties open up a number of new opportunities to provide further understanding.

First, the application to new spatial domains without needing hand-labeled training data and the automatic discovery of the cloud structure types present enables systematic study of very disparate cloud types (e.g., mesoscale convective systems, cold-air outbreaks, cirrus, and ship-tracks) and regimes of cloud organization

across the Earth. For example, having quantified the type of cloud structures present at a given location and time, the environmental conditions may be extracted from reanalysis data sets and analyzed to provide insight into what conditions are required for different forms of convective organization to occur.

Second, the temporal evolution of one type of cloud structure into another (e.g., closed- to open-cell strato-cumulus) can be studied by combining the model's ability to represent similarity between cloud structures together with observations from high-temporal resolution geostationary satellite images (GOES-16 $\Delta t = 10$ min). This can be done by studying the trajectory of a given spatial domain through the embedding space of the neural network (see section 2 for details), enabling quantitative analysis of the transition between different regimes.

Third, the model can be applied to simulations of weather and climate as well as to satellite observations. Through the ability to quantify scales and structures in simulation output, the model will provide quantitative objective measures of similarity across models and with observations, which will facilitate model inter-comparison (by quantifying the structures present across modeling hierarchies) and will help bridge the challenging gap in comparison of simulations with reality.

Finally, by using a model which itself identifies the distinguishing features that separate different types of clouds structures, rather than features that are obviously visible to the human eye (e.g., cloud size and fractal dimension) it is possible for the model to learn features that are as of yet not known. As an example, the model could be trained on all 13 radiance channels from GOES-16, rather than the composite RGB channels used in this work and through this discover structures invisible to the human eye. This may lead to new insight into cloud processes and better methods for satellite retrieval of cloud properties.

In addition, the technique presented in this work is applicable beyond the study of clouds and their interaction with the environment and can be applied to autonomously discover coherent structures of a given length-scale in any 2D geophysical field. This opens up the possibility of identifying structures which drive physical behavior which are currently unknown, and only appear when different fields are composited together, and eliminates the need for hand-labeling of training data to study patterns in geophysical data.

In summary this paper presents a tool with the potential to provide new insight into clouds and their interaction with the environment by enabling the systematic study of very large data sets from models and observations. The source code and training data in this work will be released to be freely used by the community.

References

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*, 1–4. https://doi.org/10.1029/2005GL023851

Bony, S., Dufresne, J. L., Le Treut, H., Morcrette, J. J., & Senior, C. (2004). On dynamic and thermodynamic components of cloud changes. *Climate Dynamics*, *22*(2-3), 71–86. https://doi.org/10.1007/s00382-003-0369-6

Bony, S., Stevens, B., Ament, F., Bigorre, S., Chazette, P., Crewell, S., et al. (2017). EUREC4A: A field campaign to elucidate the couplings between clouds, convection and circulation. *Surveys in Geophysics*, *38*(6), 1529–1568. https://doi.org/10.1007/s10712-017-9428-0

Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., et al. (2015). Clouds, circulation and climate sensitivity. *Nature Geoscience*, *8*(4), 261–268. https://doi.org/10.1038/ngeo2398

Falconer, K. (2004). *Fractal geometry: Mathematical foundations and applications*. Chichester, UK: John Wiley and Sons.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT Press.

Jean, N., Wang S., Samar A., Azzari G., Lobell D., & Ermon, S. (2018), Tile2Vec: Unsupervised representation learning for spatially distributed data.

Jones, E., Oliphant T., Peterson, P., & others (2001). SciPy: Open source scientific tools for Python. https://github.com/takluyver/scipy.org-new/blob/master/www/scipylib/citing.rstciting-scipy-tools

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, & Weinberger, K. Q. (Eds.), *Advances in neural information processing systems 25* (pp. 1097–1105). Inc: Curran Associates.

Medeiros, B., Stevens, B., Held, I. M., Zhao, M., Williamson, D. L., Olson, J. G., & Bretherton, C. S. (2008). Aquaplanets, climate sensitivity, and low clouds. *Journal of Climate*, *21*(19), 4974–4991. https://doi.org/10.1175/2008JCLI1995.1

Mikolov, T., Chen K., Corrado G., & Dean, J. (2013). Efficient estimation of word representations in vector space, pp. 1–12.

Orlanski, I. (1975). A rational subdivision of scales for atmospheric processes. *Bulletin of the American Meteorological Society*, *56*(5), 527–530.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., et al. (2017). Automatic differentiation in PyTorch, in NIPS Autodiff Workshop.

Rasp, S., Schulz H., Bony S., & Stevens B. (2019). Combining crowd-sourcing and deep learning to understand meso-scale organization of shallow convection, pp. 1–12.

Raspaud, M., Hoese, D., Lahtinen, P., Dybbroe, A., Finkensieper, S., Roberts, W., et al. (2019). pytroll/satpy: Version 0.18.1 (2019/11/07). https://doi.org/10.5281/zenodo.3531917

Seifert, A., & Heus, T. (2013). Large-eddy simulation of organized precipitating trade wind cumulus clouds. *Atmospheric Chemistry and Physics*, *13*(11), 5631–5645. https://doi.org/10.5194/acp-13-5631-2013

Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay, pp. 1–21.

Stevens, B., Bony, S., Brogniez, H., Hentgen, L., Hohenegger, C., Kiemle, C., et al. (2019). Sugar, gravel, fish, and flowers—Mesoscale cloud patterns in the tradewinds. *Quarterly Journal of the Royal Meteorological Society*.

Tobin, I., Bony, S., & Roca, R. (2012). Observational evidence for relationships between the degree of aggregation of deep convection, water vapor, surface fluxes, and radiation. *Journal of Climate*, *25*(20), 6885–6904. https://doi.org/10.1175/JCLI-D-11-00258.1

Tompkins, A. M., & Semie, A. G. (2017). Organization of tropical convection in low vertical wind shears: Role of updraft entrainment. *Journal of Advances in Modeling Earth Systems*, *9*, 1046–1068. https://doi.org/10.1002/2016MS000802

Vial, J., Dufresne, J. L., & Bony, S. (2013). On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Climate Dynamics*, *41*(11-12), 3339–3362. https://doi.org/10.1007/s00382-013-1725-9

Vogel, R., Nuijens, L., & Stevens, B. (2016). The role of precipitation and spatial organization in the response of trade-wind clouds to warming. *Journal of Advances in Modeling Earth Systems*, *8*, 843–862. https://doi.org/10.1002/2015MS000568

Webb, M. J., Senior, C. A., Sexton, D. M., Ingram, W. J., Williams, K. D., Ringer, M. A., et al. (2006). On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles. *Climate Dynamics*, *27*(1), 17–38. https://doi.org/10.1016/j.pbiomolbio.2006.02.004