

# Controls on stratocumulus texture diagnosed by deep learning

James A. Franke<sup>a,b,1</sup>, Takuya Kurihana<sup>c</sup>, Ian T. Foster<sup>c,d</sup>, and Elisabeth J. Moyer<sup>a</sup>

This manuscript was compiled on September 25, 2023

Marine low clouds—stratocumulus—represent the single largest source of uncertainty in the future climate response. Satellite observations show a vast array of fine-scale textures and properties not resolved by global climate models, making detailed cloud process analysis necessary but difficult. We show here that new self-supervised cloud classification techniques based on deep learning can aid in diagnosing drivers of observed cloud morphologies. We apply an autoencoder to two decades of MODIS multispectral observations without labels to generate the AI-driven Cloud Classification Atlas (AICCA), a set of 42 cloud classes that take texture into account, and compare observed classes with meteorological variables from ERA5 reanalysis and other satellite observations. We find that the strongest predictor of stratocumulus class is the temperature structure of the lower troposphere, and that dominant types fall into coherent and interpretable zones in the 2D space of inversion strength and near surface temperature. This relationship holds not only across the three major stratocumulus regions but also over time for any given location, suggesting these variables are fundamental physical drivers. The relationship cannot be reconstructed from mean cloud properties alone but requires information on texture. A substantial portion of variability in marine cloud textures remains unexplained, including spatially coherent transitions where parts of thick stratocumulus decks shift to more open configurations. We show that counter to expectations, rainfall does not play a clear role in these transitions and is not predictive of cloud classes. Self-supervised cloud classification shows promise to improve marine low cloud process understanding.

climate change | stratocumulus | deep learning | self-supervised classification

The response of low clouds is the largest uncertainty in projections of future climate under CO<sub>2</sub> forcing (1–4). A particular concern is the persistent marine stratocumulus decks that form in the subtropics off the West coasts of continents. These decks cover only 5% of the Earth's surface but are disproportionately important to its energy balance, cooling the planet by as much as 8K (5). Recent studies based on high-resolution (10 m) local large eddy simulations (LES) have suggested the decks may disappear under high CO<sub>2</sub> (6) and / or high sea surface temperature (7) conditions. Coarse-resolution (100 km) global climate models provide little guidance: their shortwave cloud feedbacks differ in sign (e.g. 4), though models that better capture historical cloud frequencies tend to have larger, positive values (8, 9). Global convection-permitting models (<5 km) can better capture marine stratocumulus, but their simulation periods are typically only months (10, 11, e.g.) and their resolutions are insufficient to fully resolve the relevant dynamics (12), so their output still does not fully reproduce stratocumulus textures (Fig. 1, which shows the ECMWF model (11); for another model see SI Fig. S1).

Because global scale simulations at 10 m resolution remain a long way off (13), satellite observations may be the most appropriate tool for understanding stratocumulus formation and stability. Space-borne instruments have by now captured several decades of high-resolution (to 30 m) multispectral imagery, which necessarily also reflect true underlying fine-scale cloud processes. These observational datasets present a different data challenge. Because the natural world does not permit large-scale experiments, process understanding must be derived from the complex details of natural experiments. The scale and variation of cloud observations then means that some form of dimension reduction must be applied for them to be usable. Cloud classification schemes have a long history, but do not capture the diversity of stratocumulus patterns. The most commonly-used scheme (ISCCP (14, 15)), simply assigns each observed pixel to one of nine classes (plus a mixed-layered type) based on its derived optical thickness and cloud top height. All marine stratocumulus falls in a single class (low, medium-thick) (14).

## Significance Statement

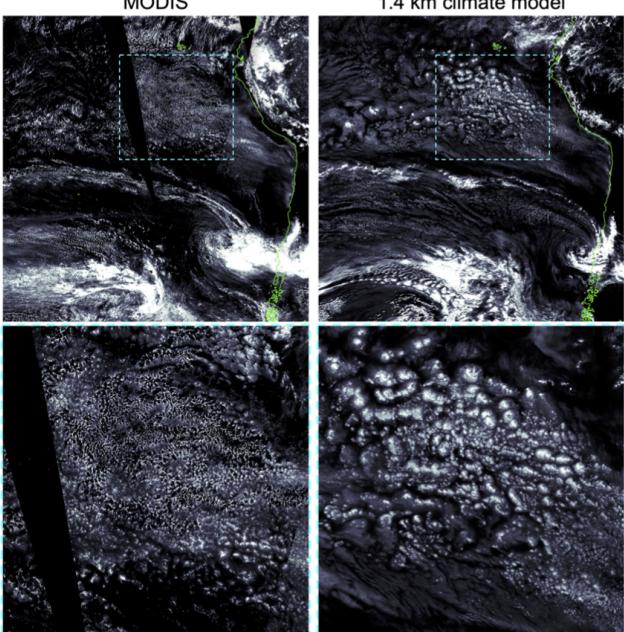
Low clouds over the ocean are critical for the Earth's climate because they cool the planet by reflecting sunlight back to space. Climate models do not simulate them well, and so cannot predict how they will change as the climate warms. Here we show that satellite images can be used to understand low cloud behavior when paired with a new automated cloud classification scheme. We find that cloud textures are broadly controlled by the temperature profile in the lower atmosphere where they form, and that, surprisingly, rainfall does not play a major role in changing those textures.

Author affiliations: <sup>a</sup>University of Chicago, Department of the Geophysical Sciences, 5734 S Ellis, Chicago, IL 60637; <sup>b</sup>Toyota Technological Institute at Chicago, 6045 S Kenwood Ave, Chicago, IL 60637; <sup>c</sup>University of Chicago, Department of Computer Science, 5730 S Ellis, Chicago, IL 60637; <sup>d</sup>Argonne National Laboratory, 9700 S Cass Ave, Lemont, IL 60439

JAF designed the study and carried out the research. TK performed the cloud classification with supervision from ITF and EJM. All authors contributed to analysis and the writing of the manuscript.

The authors declare no competing interests.

<sup>1</sup>To whom correspondence should be addressed. E-mail: jfranke@uchicago.edu



**Fig. 1.** Cloud textures off the coast of South America: liquid water path from [left] MODIS (22) observations (Nov. 1, 2018) and [right] a 1.4 km horizontal resolution simulation of the ECMWF model (11), initialized with observations on the morning of Nov. 1, 2018, so that weather patterns should match. Color intensity is identical in both images. While the high-resolution simulation captures much of the large-scale pattern (top), it produces too much self-aggregation in the main stratocumulus deck region (bottom row in).

In this work we make use of a new, deep-learning-based classification of ocean clouds, the AI-driven Cloud Classification Atlas (AICCA), in which 25 of 42 classes would clearly be considered stratocumulus in the ISCCP definition. (See Methods.) While supervised deep learning has recently been applied to cloud classification (e.g. 16–18), AICCA is instead generated with *self-supervised* classification applied 22 years of 1 km resolution cloud imagery from the Moderate Resolution Imaging Spectroradiometer (MODIS) as  $128 \times 128$  pixel patches. The resulting atlas of 200M+ classified patches reduces an 800 TB dataset to 10 GB, and the assigned classes have been shown to have consistent physical and radiative properties and to represent unique textural arrangements (19, 20). While observational studies of marine stratocumulus have largely focused on its *fractional occurrence*, and have shown that this occurrence is predictable from lower tropospheric stability (21), the AICCA classes allow asking instead what variables predict the *texture* of those clouds. Textures in turn provide insight into the processes that govern stratocumulus formation and evolution.

The evolution of marine stratocumulus textures on a timescale of days is broadly understood. Thick decks are produced in the stable subtropics by Rayleigh-Bernard convection; the decks then break up when they are advected equatorward into less stable regimes where deeper convection can occur (23, e.g.). However, this framework operating alone would produce a static spatial cloud distribution. In reality stratocumulus are highly dynamic, with complex patterns evolving on timescales of hours (24).

Research on shorter timescale stratocumulus changes has generally focused on the transition between classic closed-

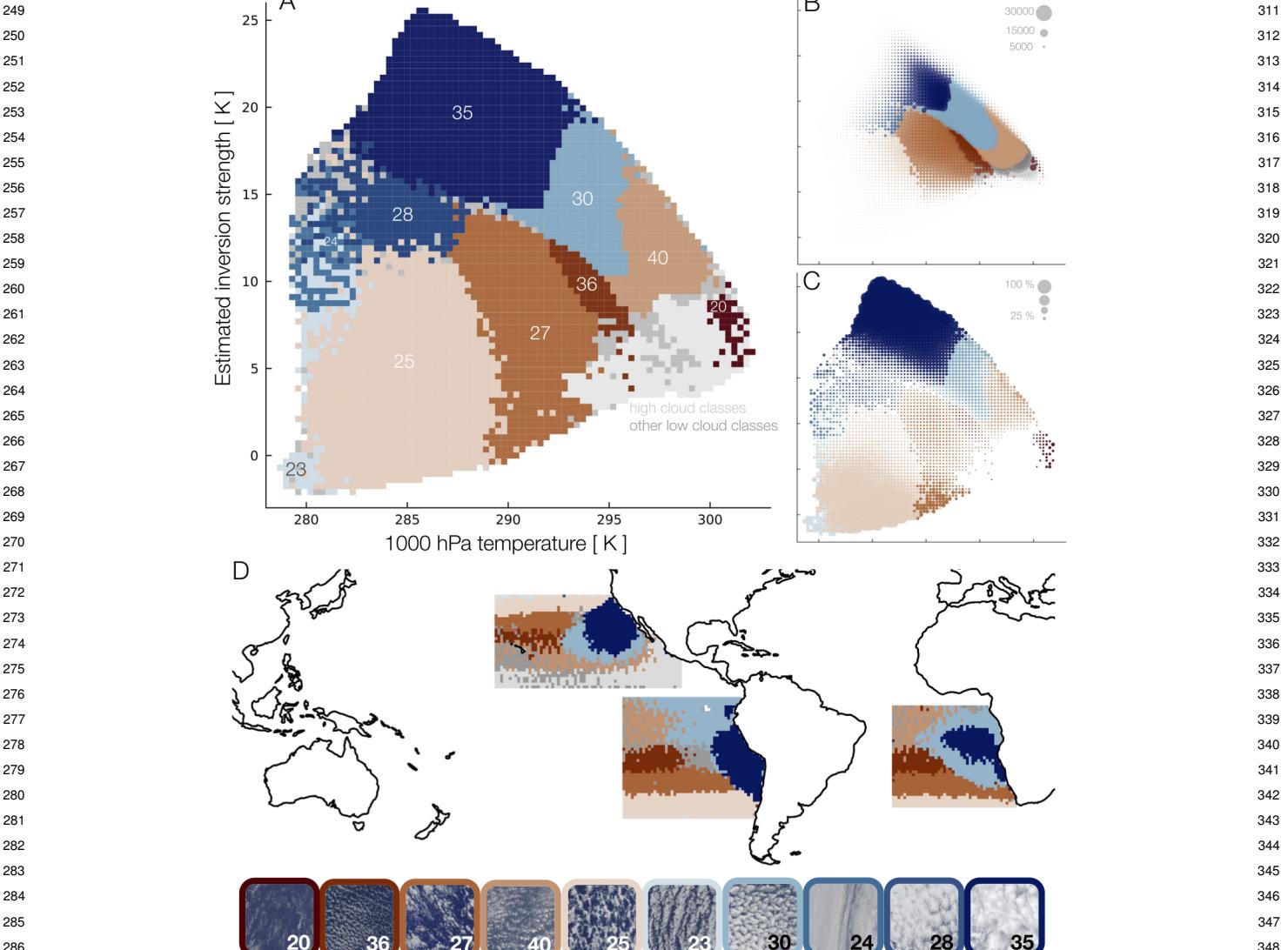
and open-cell configurations, which can occur across coherent regions even when large-scale stability remains constant (25). Many studies have suggested that rain plays a role in this transition. Observational campaigns from ships (EPIC, Eastern Pacific Investigation of Climate) and aircraft (DECS, Drizzle and Entrainment Cloud Study) found higher drizzle in parts of open-cell (“rift”) clouds (26–28), and an early LES modeling study argued that precipitation directly initiates the transition by causing downdrafts (29). Watson-Parris et al. (30) used machine learning and supervised classification to categorize 8500 occurrences of open-cell stratocumulus in otherwise unbroken decks in MODIS data and found higher rainwater path in those regions (30). On the other hand, Eastman et al. (21) cautioned that a weak statistical relationship with precipitation exists even in advective transitions where the underlying driver is clearly increasing boundary layer depth (reduced stability) (21).

We use AICCA classes here to examine the behavior of clouds in the three major subtropical stratocumulus regions (31)—the Californian, Peruvian, and Angolan, found between approximately 5 and 40 degrees North or South. We map observed cloud classes to coincident meteorological conditions and rainfall, using ERA5 reanalysis and GPM IMERG (TRMM) microwave-IR satellite precipitation measurements. The goal is to understand what factors govern stratocumulus texture, both on average and during rapid evolution.

## Results

Clouds in the subtropical stratocumulus regions are highly diverse, and the AICCA classes capture that diversity. The most dominant cloud class is the closest visually to classic closed-cell organization (#30, at 8% of cloud observations), and the next most common is a nearly uniform deck (#35, at 7%), but classes with more open textures are nearly as frequent. The 15 most common classes, all stratocumulus, collectively make up 67% of cloud observations in these regions, with the least-common of these still at 2.6%. (See Fig. 2 and SI Fig. S2 for mean class properties and SI Fig. S3 for occurrence frequencies and thumbnail images.).

To diagnose factors controlling these textures, we map the top 15 most populous classes against a battery of 68 meteorological quantities derived from reanalysis, including surface temperature, lower tropospheric stability, vertical motion, wind shear, and humidity. (See SI Table 1.) Variables are tested in pairs and scored by their ability to separate the 15 cloud classes from one another. The two 2-variable combinations with the highest predictive power across classes are inversion strength paired with either near-surface air temperature or lower tropospheric stability (See SI Table 2.). We choose the first of these pairs and show in Fig. 2a the dominant cloud classes in each part of its 2D parameter space. Domains favorable to each cloud type are both coherent in parameter space and physically interpretable. Stable conditions with high inversion strength predict more closed-cell or uniform types, as expected (blue shades in Fig. 2). Higher temperatures promote more open configurations, until some limit in inversion-strength-temperature space where stratocumulus become scarce and high clouds dominate instead. These patterns are similar in each of the three regions considered (SI Fig. S4). This separation is possibly only given



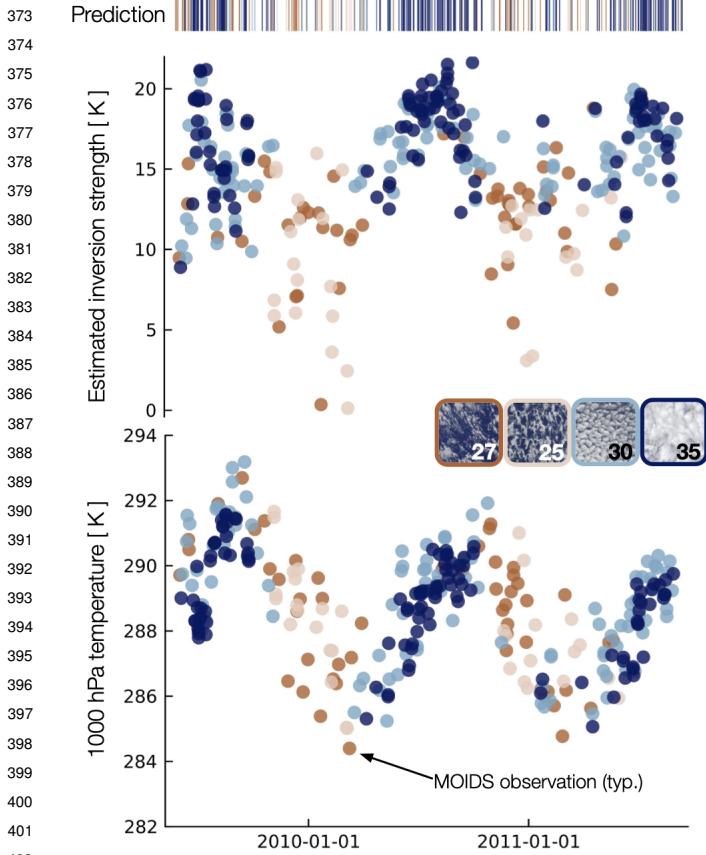
**Fig. 2.** AICCA cloud classes mapped to a meteorological domain of near-surface temperature and inversion strength, for all patches in the three main subtropical stratocumulus regions. (A) The 10 most dominant classes in meteorological space, color-coded. Thumbnail images of typical class members are shown at bottom. Nine of the 10 classes are stratocumulus; #20 is thinner and higher. All high clouds (mean cloud top pressure  $>680$  hPa) are shown in light grey. Regions of dominance are generally well-separated and coherent. (B) As in panel A but with dots scaled in size according to frequency of those meteorological conditions (for samples with clouds). Air parcels tend to move from cold and stable to warmer and less stable, and clouds textures from near-uniform to more open: #35 → #30 → #40. (C) As in panel B but with dots scaled to represent the share of all clouds by the most dominant cloud class. Very high- or low-stability conditions produce more consistent textures. (D) The three stratocumulus regions color-coded by the most dominant cloud class in each location, using the same color code. Classes are distributed in understandable geographic patterns, following large-scale environmental gradients. As expected, zones of thick stratocumulus are in the subtropics. The use of IR radiances in the classification procedure likely helps produce the strong temperature dependence in classes #25, #27, #36, and #40.

texture-based classes and cannot be reproduced using mean cloud properties alone.

Weather conditions in the stratocumulus regions do not evenly sample this entire parameter space. The predominant trajectory for air parcels moving out of subtropics is to move equatorwards, warm, and weaken in inversion strength (Fig. 2b); as this happens the stratocumulus becomes more sparse in texture and then finally breaks up. Cool surface temperatures with low inversion strength are almost never sampled, but when those conditions do occur, they are associated with a well-defined cloud class. The 10 most dominant cloud classes in parameter space are color-coded in Fig. 2. The group includes the two most frequent classes (#30

and #35), and all but one are stratocumulus and are included in the set of the 15 most common. (The sole exception is the higher and thinner class #20, whose mean characteristics put it at the ISCCP border between cumulus, stratocumulus, altocumulus, and altostratus.) In total, these 10 classes represent 42% of all cloud occurrences.

While each meteorological domain in Fig. 2a is by definition associated with a distinct cloud class, they do not explain all variance. Across all cloud occurrences, the mean “purity” across the 10 domains is 30%, i.e., 30% of cloud occurrences are in fact the dominant class associated with a given meteorological environment. However, nearly half of the “top 10” classes are located in their proper



**Fig. 3.** Timeseries of meteorological variables and observed and predicted cloud classes for a single representative location at 29N, 130W in the North Pacific basin. Top shows inversion strength and bottom near-surface temperature, with observed cloud classes shown as colors for the four main classes only (25, 27, 30, and 35). Predicted cloud classes based on EIS and T1000 are shown as colored lines in the top bar code. Cloud texture varies seasonally, broadly matches predictions, and is clearly driven by the seasonal cycle in stability. Other locations show similar features; see SI Fig. S5-8.

domain. In general, the correlation of cloud texture with environmental conditions is strongest in extreme conditions. Thick stratocumulus decks (class #35, dark blue) tend to occur in very stable conditions, and the purity of the class #35 domain is 90% at the largest inversion strengths (and 10% near the domain edge). Similarly, when low-temperature, low-stability conditions do occur, they produce class #25 with as high as 50% purity.

The dependence on meteorological conditions means stratocumulus cloud classes are geographically coherent, even though no geolocation information is provided to the autoencoder (Fig. 2d; and see SI Fig. S3 for occurrences of individual classes). The prevailing temperature and inversion strength are set by the large-scale atmospheric and ocean circulations, and in turn drive similar cloud textures over large contiguous regions. The correlation of meteorological environment to cloud texture holds not just across space but for any single location over time, suggesting it is a fundamental physical relationship (Fig. 3). As the Hadley cell shifts, the subtropics experience a seasonal cycle in inversion strength (32), and cloud textures respond.

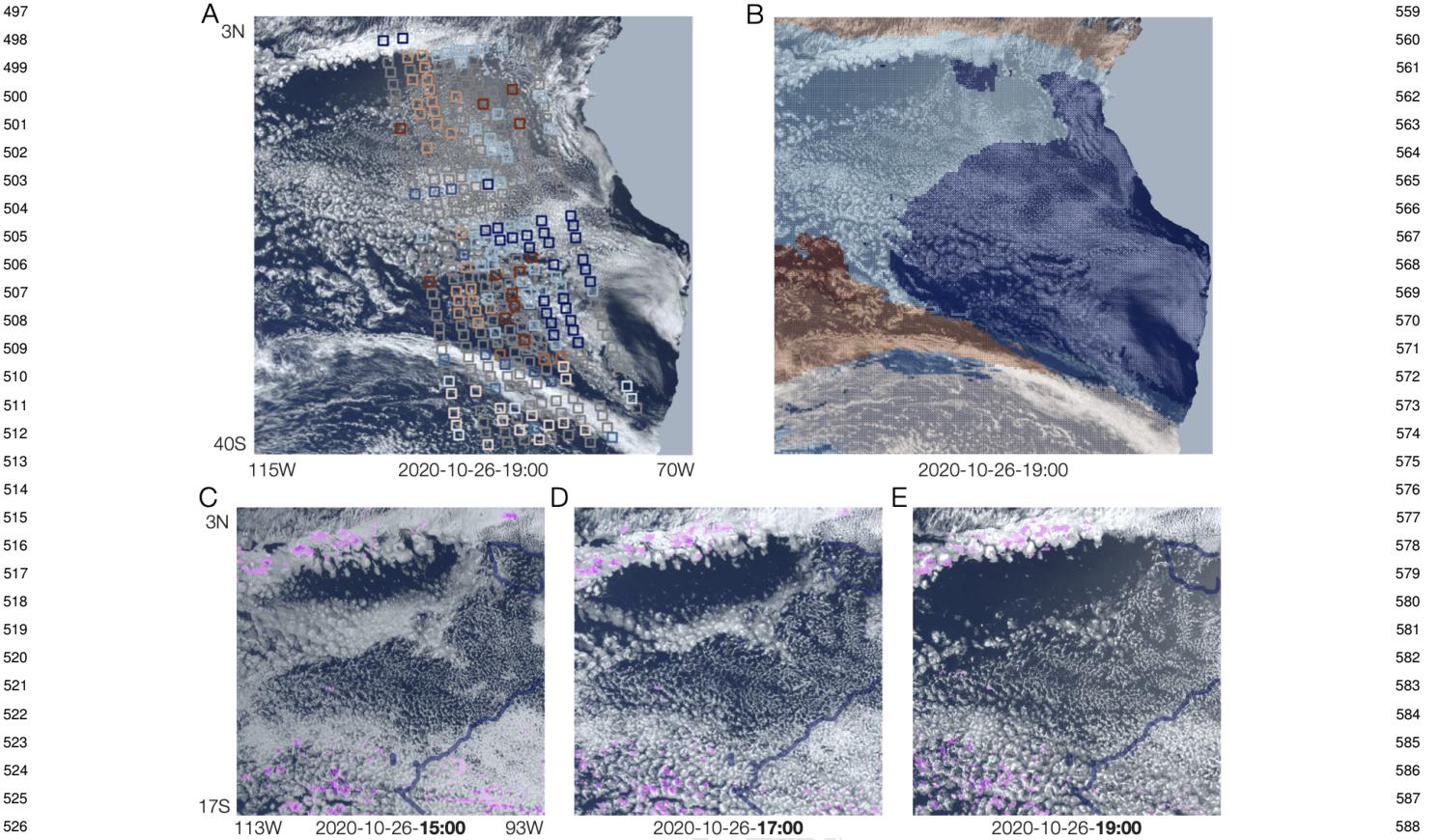
Other meteorological variables provides little additional predictive power. We repeat the prediction test used to

generate Fig. 2 but now using 3 variables, and find that no additional factor aids cloud class prediction skill by more than 2%, while many actually degrade it (SI Table S3). Relative humidity at 700 hPa provides the most benefit, and geopotential height at 700 hPa the strongest degradation. Including boundary layer depth produces little benefit since this information is already implicitly included in the inversion strength. Large-scale meteorological information therefore explains only a portion of observed variations in stratocumulus texture.

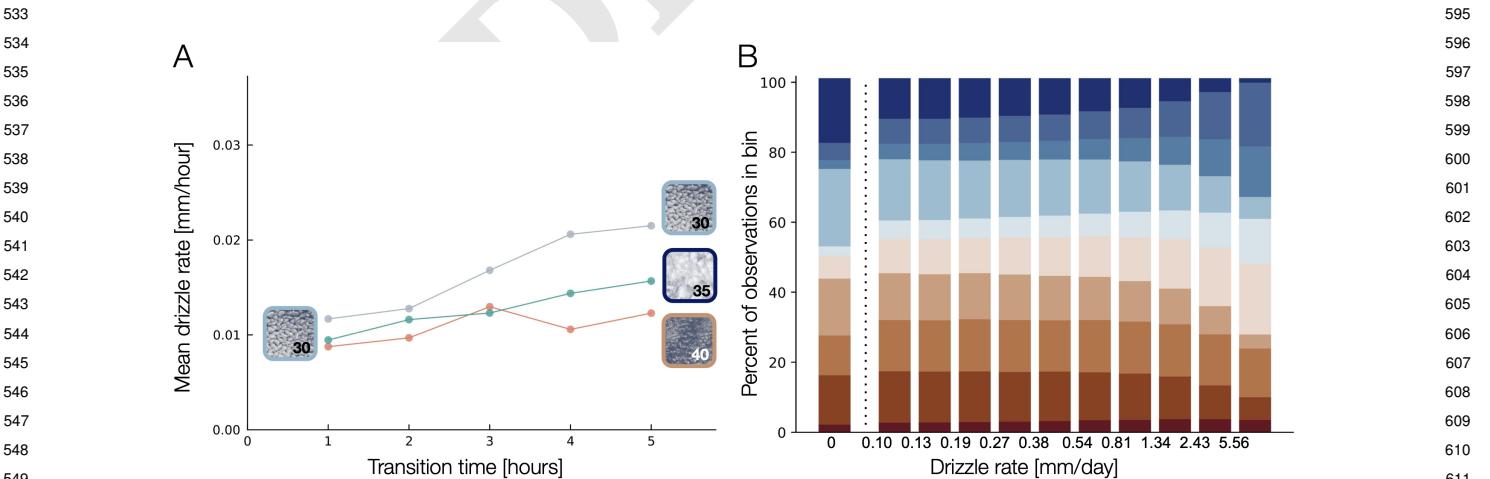
Much of the remaining, unexplained variation in stratocumulus type occurs in two broad categories: rapid transitions of cloud texture in localized contiguous regions, and a general loosening of textures over the course of each day. Fig. 4 illustrates both in GOES-16 visible images off the coast of Peru. We highlight (a) the assigned cloud classes from a MODIS overpass and (b) the predicted classes based on environmental conditions, and also show a 4-hour timeseries of one detail, an expanding “hole” in the stratocumulus deck (c–e). The meteorological cloud class predictions reproduce aspects of visual texture in this region, including large-scale patterns, but miss much fine-scale variation that is captured in the AICCA classes. The expanding “hole”, a kind of extreme endmember of a closed- to open-cell transition, occurs counter to expectations based on environmental conditions, which actually increase in inversion strength. It is also not explained by precipitation, which occurs only on its northern, upwind edge, while the “hole” expands downwind.

We then use AICCA classes to test the suggestion that precipitation drives transitions from closed- to open-cell stratocumulus (29, 33). We consider all classic closed-cell stratocumulus in the South Pacific basin (class #30, 676,000 samples), and use a Lagrangian scheme based on ERA5 horizontal winds to track air masses and identify all cases observed again within 6 hours by the TERRA satellite, which follows AQUA (3,000 samples). We bin these cases into three categories: no change (remains #30), transition to open cells (becomes any of #s 20, 25, 27, 36, or 40), or enhancement (becomes the more uniform #35). Contrary to prior suggestions, transitions to open-cell are the *least* rainy, both before and after the transition (Fig. 5a). The highest precipitation occurs when no change in texture is observed. The difference is statistically significant by a 2-sided T test.

We then check this result using a different precipitation estimate. Because the precipitation and cloud class information in Fig. 5a are derived from different satellites, on different orbits, measurements are not coincident and so hourly precipitation at each location is taken as 1/24 of the daily mean, potentially missing short-timescale changes. We therefore conduct a similar analysis with a more indirect precipitation metric derived from MODIS observations themselves: the “drizzle proxy”, based on estimated droplet concentration number and liquid water path. This proxy was used in early stratocumulus studies (e.g. 28), though more recent work has used IMERG (34). The two estimates differ by a factor of 10 (IMERG is drier) and show essentially no correlation (SI Fig. S9). Unsurprisingly, results on stratocumulus transitions using the drizzle proxy are then also different. Median estimated drizzle increases moderately for the static case (0.8 to 1.0 mm day<sup>-1</sup> for #30 to #30), increases more substantially for the thickening cases (0.6 to 1.5 mm day<sup>-1</sup>



**Fig. 4.** (A) Visible image from GOES-16 on October 26, 2020 at 19:00 UTC (~3 PM local time) with coincident MODIS-AQUA cloud classes, following the same color code used throughout, and (B) with meteorologically-predicted cloud classes. While the temperature structure predicts large-scale features of the visual texture, many features are unexplained, e.g. several large cloud-free "holes" and a 'zipper' horizontal rift feature across the middle of the domain. (C-E) Example zoom-in shows a clear sky hole expanding over the course of 4 hours from approximately 11 AM until 3 PM local time, while the local meteorologically-predicted class remains #30. (Boundary between class domains shown as blue line.) Hourly precipitation  $> 0.01 \text{ mm hour}^{-1}$  is shown in pink. Neither precipitation nor large-scale changes explain this rapid breakup of stratocumulus. Note also that cloud textures across the domain loosen over the afternoon, as is typical.



**Fig. 5.** Population level precipitation impacts. Drizzle does not aid in explaining texture. (A) Short timescale transitions (<6h) from the most populous class # are tracked in the South Pacific basin. Nearly 3,000 such transitions are observed via Lagrangian tracking between MODIS TERRA and AQUA passes. Mean precipitation rate (including zeros) is reported for three categories of transition: from #30 to an open cell class (#40, #25, #26, etc.), from #30 to thick, unbroken low cloud (#35), and from #30 to itself (null case, no transition observed). Contrary to expectation, we find no significant effect of precipitation on transitions from the most populous closed cell class. In fact, the null cases shows the strongest (albeit weak) precipitation signal. Prevailing winds advect parcels equator ward and to the west up a mean precipitation gradient which results in slight mean increase in precipitation with time. Y axis includes all observations with no precipitation as well, which dominate and thus result in very low mean drizzle rates. (B) Stacked Bar chart shows class prevalence equally divided by ~100,000 sample precipitation bins across the 10 descriptive low cloud classes. **x-axis is not uniformly spaced.** The no-precipitation bin contains two thirds of all the samples in reality, but is shown scaled for comparison. At the highest pr values, class 35 share decreases and is replaced by 30, but no overall shift to the open (red) classes is found. The example 10 low cloud classes are almost evenly distributed across the precipitation sample.

for #30 to #35), and actually decreases for the closed- to open-cell transitions (1.0 to 0.4 mm day<sup>-1</sup> for #30 to open classes). (These values are calculated only for years 2010–2013, reducing the sample size to 650 sequential observations.) While these results could imply a role for precipitation, note that they are opposite from those of prior studies, which found higher precipitation in open cells (27, 28, 30).

In general, precipitation is an unlikely candidate to drive observed stratocumulus diversity since most stratocumulus observations are dry. In the dataset over the stratocumulus region used here, clouds are near-ubiquitous but rainfall is scarce: nearly 80% of the 24 M patches are cloudy, but only 33% of those cloudy patches exhibit precipitation > 0.1 mm day<sup>-1</sup>. (This value is if anything too high; a validation exercise against gauges shows that IMERG overestimates the lightest precipitation (35).) Most transitions of stratocumulus texture therefore involve no measurable drizzle. Unsurprisingly, precipitation is not particularly predictive for cloud textures (Fig. 5b, which shows the fractional occurrences of the “top 10” stratocumulus classes in 11 precipitation bins (ranging from < 0.1 to > 5.56 mm day<sup>-1</sup>). Cloud classes are nearly evenly distributed in each bin until the highest rain rates, when some drop sharply (e.g. uniform cloud decks, class #35), but even this change does not produce an overall shift to more open textures.

## Discussion

Self-supervised cloud classification provides a new tool that can aid in understanding the processes that control stratocumulus textures and in turn their radiative impacts. This work demonstrates that AI Cloud Classification Atlas, which divides stratocumulus clouds that would traditionally be placed in a single category into 25 different classes, can be used to generate insight into factors governing their textures. These classes help identify large-scale meteorological drivers that explain at least a portion of stratocumulus diversity, since they map into coherent regions of a 2D variable space of near-surface temperature and inversion strength. These results are consistent with and expand on previous observational work focused on mean cloud fraction (21, 36–39). Furthermore, the remaining unexplained portion of textural diversity is likely driven by small-scale processes. Counter to previous work, results here suggest that precipitation does not play a strong role in driving stratocumulus transitions.

Data limitations restrict our ability to draw some conclusions about cloud processes in this work. For understanding how temperature and stability affects stratocumulus, an inherent limitation is that the clouds may feed back on local temperature structure, at minimum confusing cause and effect and at maximum also causing errors in ERA-5 reanalysis (since clouds in ERA-5 need not match those in reality). The only test for the latter effect, short of a field campaign, is to confirm that errors are not large enough to significantly affect cloud class predictions. For example, clouds of class #30 are matched to their expected domain no more frequently when ERA-5 correctly produces stratocumulus than when ERA-5 is incorrectly cloud-free (in a 1-month test case). For understanding rapid stratocumulus transitions, multiple limitations hamper understanding. No appropriate dataset exists to test the potential role of aerosols, since most techniques yield measurements only in clear sky

conditions (40). Prior work has used monthly clear-sky means or asynchronous measurements (e.g. 34). While aerosol concentrations are unlikely to change significantly on hourly timescales, the effect cannot be ruled out. Any errors in ERA-5 near-surface winds will also hamper our ability to match air parcels seen at different times. Furthermore, it is not even clear that Lagrangian analysis is the most appropriate, since some features appear stationary in space or even propagate upwind (Fig. 4c–e). Whether analysis is Lagrangian or Eulerian, the constraint of matching the sparse MODIS sampling severely reduces the sample size for analysis (by a factor of over 200 for timescales <6 hours).

While AICCA used MODIS for its global coverage and multi-decadal record, rapid transitions in subtropical stratocumulus are better studied in data from geostationary satellites, which avoid the problem of sparse overpasses. A cloud atlas built from geostationary imagery would provide fine-scale, continuous data across a single stratocumulus region, aiding in diagnosing the “holes” and rifts that develop in stratocumulus decks, and the prevalent loosening of stratocumulus textures that occurs over the course of each day, even under constant stability conditions. This diurnal evolution may be the most informative present-day analogue for proposed future CO<sub>2</sub>-induced stratocumulus breakups.

The behavior of subtropical stratocumulus remains the largest uncertainty in future climate projections. Global climate simulations suggest more favorable conditions for stratocumulus decks, at least in the near term, since inversion strength in the subtropics increases moderately even though near-surface temperatures warm (38, 41). LES simulations suggest the possibility for complete breakups, and the expected poleward expansion of the Hadley cell (42), which affects the location of stratocumulus formation, introduces new unknowns. Self-supervised cloud classification offers a new and potentially important tool for addressing these questions, by facilitating the use of decades and Petabytes of multi-spectral cloud imagery. In general, as Earth system data volumes continue to expand, data reduction techniques like self-supervised classification will become crucial aids for understanding processes in our changing climate.

## Materials and Methods

**Data Availability.** All data used in this analysis is publicly available from the following sources. Self-supervised cloud classes and MODIS cloud properties are available at [globus-labs](#). ERA5 data is available at [Copernicus](#). GPM IMERG data is available from [NASA](#). MODIS cloud properties are available from [NASA](#) and droplet concentrations are available from [CEDA](#). Codebase for analysis is available on [Github](#). 1.4 km climate model data shown for background only and is available upon request from the corresponding author.

**Self-Supervised cloud classification.** The AI-Based Cloud Classification Atlas (AICCA) employs a rotationally-invariant convolutional autoencoder to classify clouds in 2000–2022 MODIS satellite images over the global oceans from both TERRA and AQUA (19) MODIS instruments. The autoencoder is trained without explicit supervision 1 million sample 128×128 pixels patches and consisting of six near-IR and IR radiance bands from the level 1b product. Cloud patterns can occur in different orientations. Thus, we develop a rotation-invariant (RI) autoencoder that generates identical latent representations for cloud objects that differ only in their horizontal orientation. Hierarchical agglomerative clustering is then applied to the latent vectors from the autoencoder (See

745 SI Fig. S10 and (19, 20) for details) generating an information-  
746 maximizing 42 cloud classes. Inference is applied to the remaining  
747 ~200M observations with the trained model. MODIS pixels are  
748 ~1 km resolution and so the resultant classification is at ~1 degree  
749 latitude and longitude. Cloud classes are separable, stable, and  
750 physically reasonable (20, 43).

751 **Meteorological mapping.** In this analysis, we focus on the three  
752 main subtropical stratocumulus regions (31). We test 15 most  
753 populous clouds classes in these subtropical regions which are, in  
754 order: [#30, #35, #40, #36, #26, #27, #39, #32, #37, #19,  
755 #29, #25, #33, #41]. The class number is assigned based on  
756 cloud top pressure: smaller numbers represent high-altitude clouds  
757 and higher numbers denote low-altitude clouds (see SI Fig. S3).

758 Observed AICCA classes are then mapped to daily mean  
759 reanalysis meteorology variables from the ERA5 reanalysis (44).  
760 Variables are taken from the hourly product at pressure levels  
761 and aggregated to daily means. Variables include temperature,  
762 relative humidity, divergence, geopotential, potential vorticity, u,  
763 v, w components of wind at different pressure levels between 950  
764 and 700 hpa as well as surface quantities including: sea surface  
765 temperature, mean sea level pressure, significant wave height,  
766 some engineered features including: lower tropospheric stability,  
767 estimated inversion strength (EIS), and some calculated wind  
768 shears (see SI Table 1). AICCA observation patch latitude and  
769 longitude are rounded to the nearest 0.25 degree and matched with  
770 the ERA-5 value (provided at 0.25 degree spatial resolution) on  
771 the same day as the observation. Estimated inversion strength is  
772 calculated based on the method of (36). We also include daily and  
773 hourly precipitation from the GPM IMERG composite satellite  
774 Late-run product (45). Precipitation values are aggregated to  
775 0.25 degrees (from native 0.1 degrees) to match ERA5 resolution.  
776 Finally, a MODIS drizzle proxy is calculated from the method in  
777 (28) using droplet concentrations from (46) and liquid water path  
778 data from (22).

779 Meteorological drivers of low cloud texture are quantitatively  
780 tested in the following fashion. We train a binary logistic regression  
781 model with L2 penalty and limited-memory BFGS solver (in  
782 scikit-learn (47)) for pairs of 5000 randomly-selected samples from  
783 each class in the top 15 classes. Every combination of classes is

784 tested across all 68 meteorological variable combinations (over 200k  
785 combinations in total). Pairs of meteorological driver are scored  
786 based on their mean prediction skill across all pairs of cloud classes.  
787 The process is repeated for combinations of three meteorological  
788 variables. Clear sky observations are ignored in this step.

789 **Lagrangian tracking.** MODIS class observations are tracked in space  
790 in a Lagrangian framework. Air parcels (and the clouds they  
791 contain) are assumed to move with the winds at 925 hPa, following  
792 (21). We calculate forward trajectories by advecting each MODIS  
793 observation by the rate of the mean hourly wind speed in that  
794 location for one hour (from ERA5). The process is repeated  
795 until the new location and timestamp line up in the same gridcell  
796 (rounded to half degree for join, otherwise trajectory continues) and  
797 timestamp (rounded to hour) with another MODIS observation  
798 or until 24 hours have passed, at that point the trajectory is  
799 abandoned. Some transitions are observed at every timestep  
800 between 1 and 24 hours depending on location of the starting patch  
801 and the pass times of TERRA and AQUA. Total precipitation  
802 from GPM IMERG is tracked along the trajectory at each hour.  
803 Drizzle proxy values can only be recorded where MODIS samples  
804 exist at the start or end of the trajectory.

805 **ACKNOWLEDGMENTS.** JAF was supported in part by NSF  
806 NRT program (grant DGE-1735359) and by an appointment to the  
807 Intelligence Community Postdoctoral Research Fellowship Program  
808 at Toyota Technological Institute at Chicago administered by  
809 Oak Ridge Institute for Science and Education (ORISE) through  
810 an inter agency agreement between the U.S. Department of  
811 Energy and the Office of the Director of National Intelligence  
812 (ODNI). ITF was supported in part by the U.S. Department of  
813 Energy under Contract DE-AC02-06CH11357. The authors thank  
814 Valentine Anantharaj for support with 1.4 km model data and  
815 Chris Bretherton for helpful comments.

- 816 1. GL Stephens, Cloud feedbacks in the climate system: A critical review. *J. Clim.* **18**,  
817 237–273 (2005).
- 818 2. M Zhao, et al., Uncertainty in model climate sensitivity traced to representations of cumulus  
819 precipitation microphysics. *J. Clim.* **29**, 543–560 (2016).
- 820 3. SC Sherwood, et al., An assessment of Earth’s climate sensitivity using multiple lines of  
821 evidence. *Rev. Geophys.* **58** (2020).
- 822 4. MD Zelinka, et al., Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res.  
823 Lett.* **47** (2020).
- 824 5. T Schneider, CM Kaul, KG Pressel, Possible climate transitions from breakup of  
825 stratocumulus decks under greenhouse warming. *Nat. Geosci.* **12**, 163–167 (2019).
- 826 6. T Schneider, CM Kaul, KG Pressel, Solar geoengineering may not prevent strong warming  
827 from direct effects of CO<sub>2</sub> on stratocumulus cloud cover. *Proc. Natl. Acad. Sci.* **117**,  
828 30179–30185 (2020).
- 829 7. G Bellon, O Geoffroy, Stratocumulus radiative effect, multiple equilibria of the well-mixed  
830 boundary layer and transition to shallow convection. *Q. J. Royal Meteorol. Soc.* **142**,  
831 1685–1696 (2016).
- 832 8. P Kuma, FAM Bender, A Schuddeboom, AJ McDonald, Ø Seland, Machine learning of  
833 cloud types shows higher climate sensitivity is associated with lower cloud biases.  
*Atmospheric Chem. Phys.* **23**, 523–549 (2023).
- 834 9. L Bock, A Lauer, Cloud properties and their projected changes in CMIP models with  
835 low/medium/high climate sensitivity (2023) <https://doi.org/10.5194/egusphere-2023-1086>.
- 836 10. B Stevens, et al., DYAMOND: the DYnamics of the Atmospheric general circulation Modeled  
837 On Non-hydrostatic Domains. *Prog. Earth Planet. Sci.* **6**, 61 (2019).
- 838 11. NP Wedi, et al., A baseline for global weather and climate simulations at 1 km resolution. *J.  
839 Adv. Model. Earth Syst.* **12** (2020).
- 840 12. T Schneider, et al., Harnessing AI and computing to advance climate modelling and  
841 prediction. *Nat. Clim. Chang.* **13**, 887–889 (2023).
- 842 13. T Schneider, et al., Climate goals and computing the future of clouds. *Nat. Clim. Chang.* **7**,  
843 3–5 (2017).
- 844 14. WB Rossow, RA Schiffer, Advances in understanding clouds from ISCCP. *Bull. Am.  
845 Meteorol. Soc.* **80**, 2261–2287 (1999).
- 846 15. AH Young, KR Knapp, A Inamdar, W Hankins, WB Rossow, The International Satellite  
847 Cloud Climatology Project H-Series climate data record product (2018)  
<https://www.ncdc.noaa.gov/products/international-satellite-cloud-climatology>.
- 848 16. J Zhang, P Liu, F Zhang, Q Song, CloudNet: Ground-Based Cloud Classification With Deep  
849 Convolutional Neural Network. *Geophys. Res. Lett.* **45**, 8665–8672 (2018).

869	33. I Koren, G Feingold, Aerosol–cloud–precipitation system as a predator-prey problem. <i>Proc. Natl. Acad. Sci.</i> <b>108</b> , 12227–12232 (2011).	931
870	34. MW Christensen, WK Jones, P Stier, Aerosols enhance cloud lifetime and brightness along the stratus-to-cumulus transition. <i>Proc. Natl. Acad. Sci.</i> <b>117</b> , 17591–17598 (2020).	932
871	35. J Tan, WA Petersen, A Tokay, A Novel Approach to Identify Sources of Errors in IMERG for GPM Ground Validation. <i>J. Hydrometeorol.</i> <b>17</b> , 2477–2491 (2016).	933
872	36. R Wood, CS Bretherton, On the Relationship between Stratiform Low Cloud Cover and Lower-Tropospheric Stability. <i>J. Clim.</i> <b>19</b> , 6425–6432 (2006).	934
873	37. TA Myers, JR Norris, On the Relationships between Subtropical Clouds and Meteorology in Observations and CMIP3 and CMIP5 Models*. <i>J. Clim.</i> <b>28</b> , 2945–2967 (2015).	935
874	38. DT McCoy, R Eastman, DL Hartmann, R Wood, The Change in Low Cloud Cover in a Warmed Climate Inferred from AIRS, MODIS, and ERA-Interim. <i>J. Clim.</i> <b>30</b> , 3609–3620 (2017).	936
875	39. H Andersen, J Cermak, L Zipfel, TA Myers, Attribution of Observed Recent Decrease in Low Clouds Over the Northeastern Pacific to Cloud-Controlling Factors. <i>Geophys. Res. Lett.</i> <b>49</b> (2022).	937
876	40. RC Levy, et al., The Collection 6 MODIS aerosol products over land and ocean. <i>Atmospheric Meas. Tech.</i> <b>6</b> , 2989–3034 (2013).	938
877	41. X Qu, A Hall, SA Klein, PM Caldwell, The strength of the tropical inversion and its response to climate change in 18 CMIP5 models. <i>Clim. Dyn.</i> <b>45</b> , 375–396 (2015).	939
878	42. WK Lau, KM Kim, Robust hadley circulation changes and increasing global dryness due to co <sub>2</sub> warming from cmip5 model projections. <i>Proc. Natl. Acad. Sci.</i> <b>112</b> , 3630–3635 (2015).	940
879	43. T Kurihana, J Franke, I Foster, Z Wang, E Moyer, Insight into cloud processes from unsupervised classification with a rotationally invariant autoencoder. <i>arXiv preprint arXiv:2211.00860</i> (2022).	941
880	44. Hersbach, et al., The ERA5 global reanalysis. <i>Q. J. Royal Meteorol. Soc.</i> <b>146</b> , 1999–2049 (2020).	942
881	45. G Huffman, E Stocker, D Bolvin, E Nelkin, J Tan, GPP IMERG late precipitation L3 1 day 0.1 degree x 0.1 degree V06 (2019) Accessed: August 1, 2023.	943
882	46. E Gryspenert, et al., Cloud droplet number concentration, calculated from the MODIS (Moderate resolution imaging spectroradiometer) cloud optical properties retrieval and gridded using different sampling strategies (2022) <a href="https://dx.doi.org/10.5285/864a46cc65054008857ee5bb772a2a2b">https://dx.doi.org/10.5285/864a46cc65054008857ee5bb772a2a2b</a> .	944
883	47. F Pedregosa, et al., Scikit-learn: Machine learning in Python. <i>J. Mach. Learn. Res.</i> <b>12</b> , 2825–2830 (2011).	945
884		946
885		947
886		948
887		949
888		950
889		951
890		952
891		953
892		954
893		955
894		956
895		957
896		958
897		959
898		960
899		961
900		962
901		963
902		964
903		965
904		966
905		967
906		968
907		969
908		970
909		971
910		972
911		973
912		974
913		975
914		976
915		977
916		978
917		979
918		980
919		981
920		982
921		983
922		984
923		985
924		986
925		987
926		988
927		989
928		990
929		991
930		992