

# 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)), assigns each observed pixel to one of nine classes (plus a mixed-layered type) based on its derived optical thickness and cloud top height. In this scheme, all marine stratocumulus falls in a single class (low, medium-thick) (14).

## Significance Statement

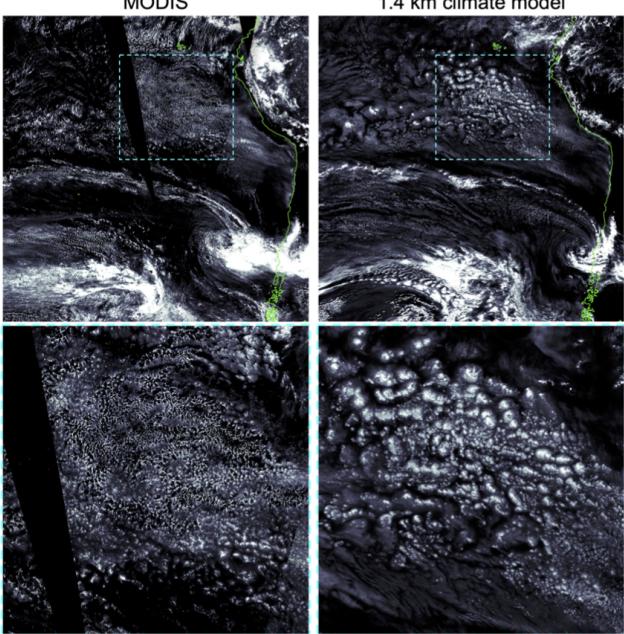
Low clouds over the ocean are critical for the Earth's climate because they cool the planet by reflecting sunlight back to space. Climate models do not simulate them well, and so cannot predict how they will change as the climate warms. Here we show that satellite images can be used to understand low 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 observations (Nov. 1, 2018) and [right] a 1.4 km horizontal resolution simulation of the ECMWF model, initialized with observations in the morning of Nov. 1, 2018 (11), 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.) AICCA is generated by self-supervised classification of 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 (16, 17). 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 (18), 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 (19, 20, 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 (20).

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

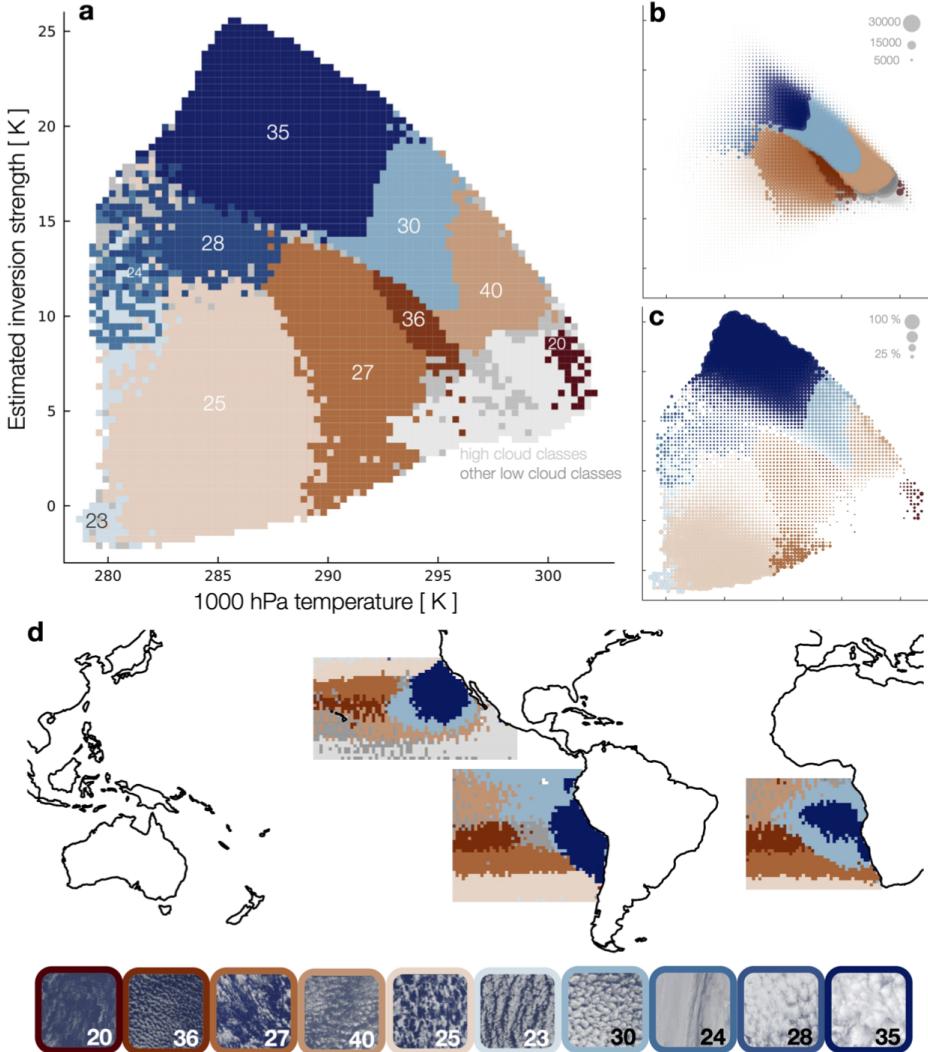
(21). Many studies have suggested that rain plays a role in this transition. Observational campaigns from ships (EPIC, Eastern Pacific Investigation of Climate) and aircraft (DECS, Drizzle and Entrainment Cloud Study) found higher drizzle in parts of open-cell (“rift”) clouds (22–24), and an early LES modeling study argued that precipitation directly initiates the transition by causing downdrafts (25). Watson-Parris et al. (26) 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 (26). On the other hand, Eastman et al. (18) 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) (18).

We use AICCA classes here to examine the behavior of clouds in the three major subtropical stratocumulus regions (27)—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 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. We choose the first of these pairs and show in Fig. 2a the 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. S5). This separation is possibly only given texture-based classes and cannot be reproduced using mean cloud properties alone.



**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.

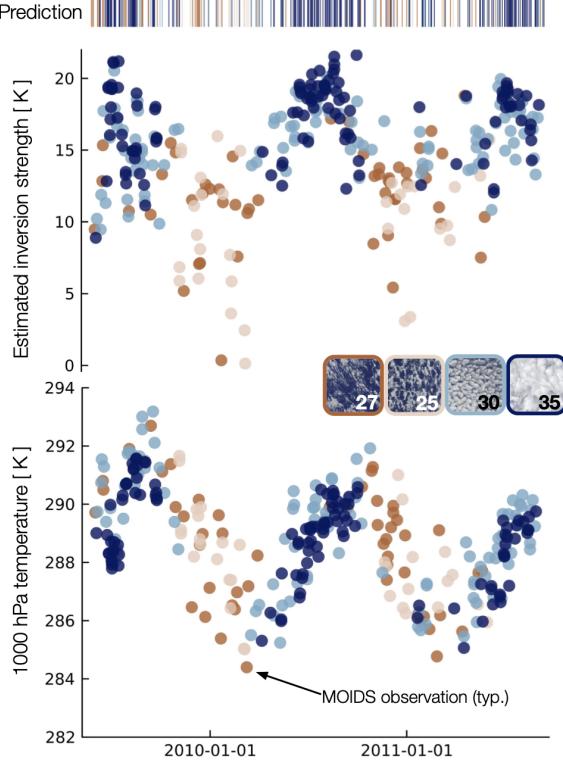
Weather conditions in the stratocumulus regions do not evenly sample this entire parameter space. The predominant trajectory for air parcels moving out of subtropics is to move equatorwards, warm, and weaken in inversion strength (Fig. 2b); as this happens the stratocumulus becomes more sparse in texture 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. 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. 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.

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). This coherence reflects the dependence on meteorological conditions discussed above. 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 (28), and cloud textures respond. Note that the seasonal cycle in cloud texture leads that in temperature, contradicting proposed suggestions that dense stratocumulus feeds back on near-surface temperature,



**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 See SI Fig. S5-9.

cooling it (). Instead, the densest cloud classes begin appearing in Spring as stability increases, while temperatures are still rising.

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 show a 4-hour timeseries of one detail, an expanding “hole” in the stratocumulus deck (c–e). The meteorological cloud class predictions reproduce some aspects of visual texture in this region, including large-

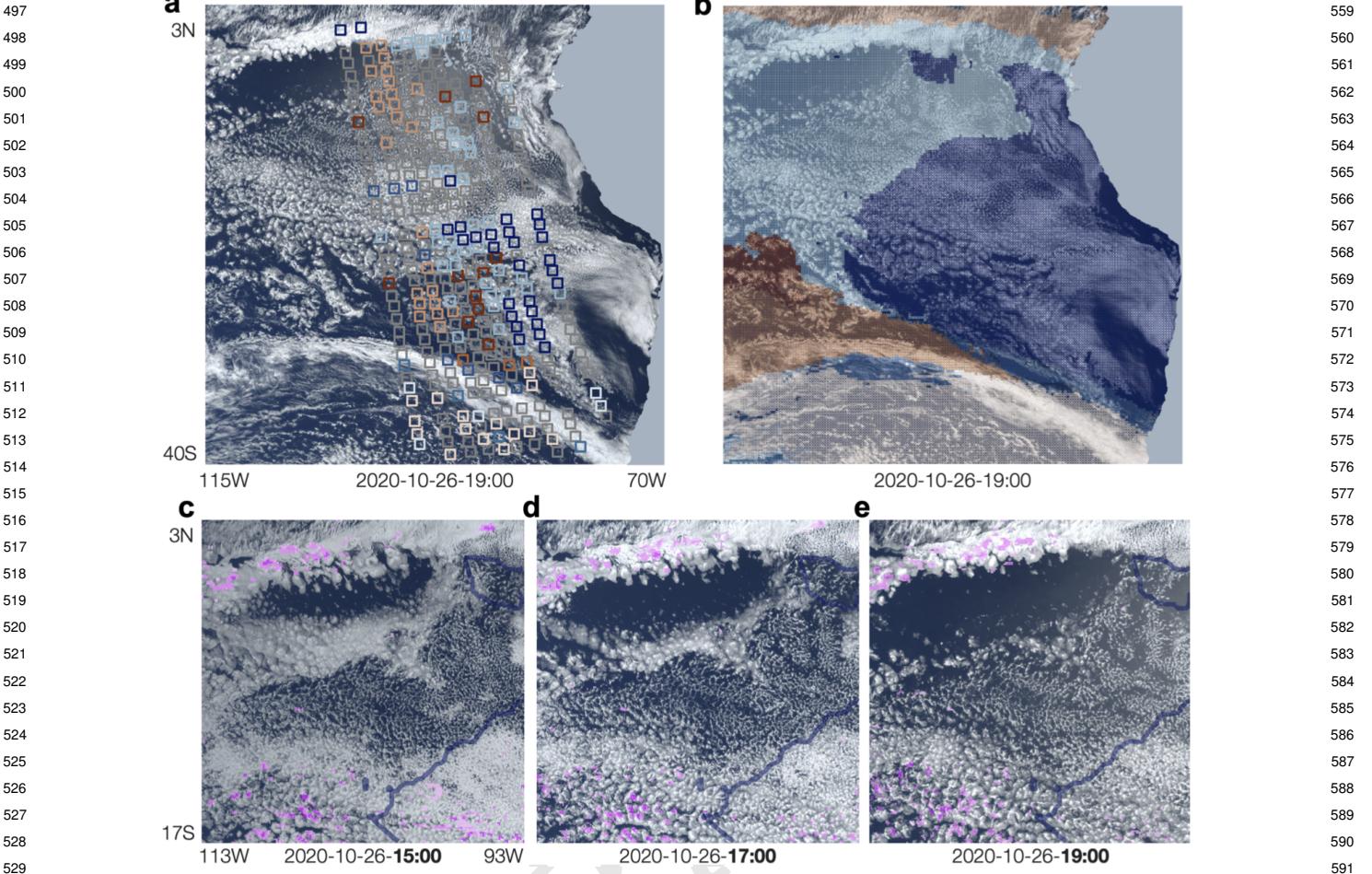
scale patterns, but miss much fine-scale variation that is captured in the AICCA classes. The expanding “hole”, a kind of extreme endmember of a closed- to open-cell transition, occurs counter to any expectations based on environmental conditions, which actually increase in inversion strength. It is also not explained by precipitation, which occurs only on its northern, upwind edge, while the “hole” expands downwind.

Precipitation has been proposed as driver for stratocumulus morphological transitions (25, 29). In our single case study (Fig. 4), precipitation does not explain the breakup from thick clouds into clear sky in otherwise stable conditions. To test the effect of precipitation on low cloud transitions more broadly, we track observed MODIS classes with a Lagrangian scheme based on ERA5 horizontal winds across the southern Pacific basin only for relatively short term transitions (less than 6 hours). We focus on transitions from the most populous class, #30, classical closed cell and bin into three separate cases: cloud degradation, enhancement, and no change. Transitions from 30 to much sparser classes (20, 25, 27, 36, 40) are less rainy than apparent cloud enhancement transitions from 30 to the unbroken class #35, though this is statistically insignificant (2-sided T test). The case with the highest rainfall is actually the no-transition case (#30 remains #30, Fig. 5, panel a). We repeat this test with the MODIS drizzle proxy, calculated from MODIS estimated droplet concentration number and liquid water path following previous work and find almost the same result. Across a 2010-2013 subset of our Lagrangian sample, drizzle increases moderately ( $0.8 \text{ mm day}^{-1}$  to  $1.0 \text{ mm day}^{-1}$  median) for the #30 to #30 case, more substantially for the 30 to 35 case ( $0.6 \text{ mm day}^{-1}$  to  $1.5 \text{ mm day}^{-1}$ ), and decreases for the #30 to open case ( $1.0 \text{ mm day}^{-1}$  to  $0.4 \text{ mm day}^{-1}$ ). We find weak evidence that, class #30 samples which end up transitioning to open classes are somewhat more rainy than those #30s which either remain unchanged or thicken into #35. However, resultant open cell classes have the least rain overall ( $0.4 \text{ mm day}^{-1}$ ), contrary to previous work.

Precipitation is not particularly predictive for snapshot cloud textures across the three sub-tropical regions in general. Including precipitation measurements in the logistic regression improves prediction only 0.01% on average for the top 15 classes and actually slightly degrades prediction for separating the top four (25, 27, 30, 35). This is perhaps unsurprising, given how little it rains in these stratocumulus regions. Two thirds of all samples experience very little precipitation at all (less than  $0.1 \text{ mm day}^{-1}$ ) so it cannot help to separate class variation in the vast majority of classes. In a population-level sample across 1 million observed classes, precipitation is associated with changes in cloud type distribution, but only to a small degree and only for the highest levels of precipitation (greater than  $5 \text{ mm day}^{-1}$ , Fig. 5b). Classes are nearly evenly-distributed across the precipitation sample space.

## Discussion

Self-supervised cloud classification can reveal processes driving cloud formation and dynamics. The diversity of low cloud morphology in the subtropics is highlighted by the approach. Our atomic cloud classification algorithm identifies over 25 cloud types in what would classically be termed stratocumulus. A large portion of this diversity in taxonomy



**Fig. 4.** **a** Visible light image from GOES-16 on October 26, 2020 at 19:00 UTC (approximately 3 PM in the afternoon local time) with coincident MODIS-AQUA pass cloud classes and **b** with meteorological predicted class plotted above in color. While the temperature structure predicts some of the visual texture on the larger-scale, many of the textures are unexplained by this framework. **[c-e]** Example zoom-in shows clear sky hole expanding over the course of 4 hours from approximately 11 AM until 3 PM local time. Temperature-structure predicted class remains #30 for this time. Satellite microwave hourly precipitation measurements shown in pink. Precipitation does not explain rapid cloud disappearance in the center of the domain, nor do changes in the ERA temperature structure. Note that cloud textures across the domain loosen over this time period as is typical.

is easily explained by the temperature structure of the lower troposphere, confirming previous observational work focused on mean cloud fraction (18, 30–32, e.g.). Contrary to previous work, we find precipitation plays little role in the texture or low clouds or their apparent transitions. However, much of the variation remains unexplained by this simple framework (see supplemental bar chart). We test the following possibilities to explain this discrepancy: additional meteorological drivers, hysteresis or lag effects, regional differences, and within cloud class variation.

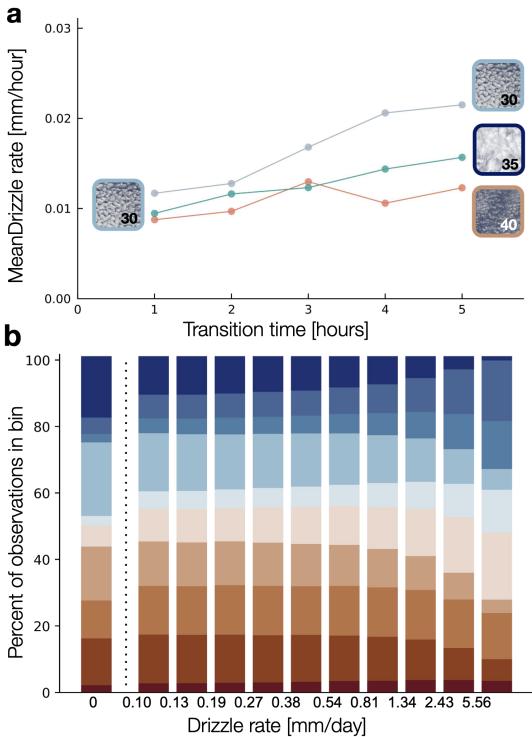
Determining the drivers of physical processes from correlations of cloud properties with environmental conditions is inherently difficult. Clouds may also feed back on their local conditions to some degree, including reducing sea surface temperatures () and increasing the stability of the lower atmosphere ( ).

However, the assumption that clouds are following their environment appears reasonable to the first order. The major regions favorable for stratocumulus persistence are set by large-scale circulation patterns, including the descending Hadley cell and wind-driven ocean upwelling (20). If clouds

were able to make favorable conditions for themselves, we would find low marine clouds persisting in other locations as well. Nevertheless, we can only report correlations between the meteorology and cloud type. with the MODIS satellite sampling rate (which provides at most two observations in one location in a day in this part of the world).

Additional meteorological variables – we test up to 68 but can only improve prediction about 1–2% by adding additional information. Likewise, we find no regional differences in prediction skill or errors or in the power of additional meteorological variables. See SI Fig. S4–5. Hysteresis and lag effects from being advocated into a less stable or warmer domain also seem unlikely. Qualitatively visual analysis of geostationary images shows clouds seem to change locally on timescales much shorter than advection to other regions. This makes hysteresis seem unlikely. MODIS temporal sampling limits our ability to study this effect. This should be quantified in future work with higher satellite temporal sampling.

Cloud texture variation within a class can be substantial (the classification is not perfect). However, this does not



**Fig. 5.** Population level precipitation impacts. Drizzle does not aid in class prediction. **a** Short timescale transitions (<6h) from the most populous class # are tracked in the south pacific basin. 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.

track with the unexplained class variation in meteorological space. Those classes which are further from the cluster center in latent space (from the autoencoder bottleneck) are not necessarily more likely to be located outside their region of temperature space, with a few potential exceptions. See SI Fig. 10.

Aerosols are clearly critical to cloud processes. However, we simply do not have sufficient large scale aerosol observational data to include in our framework. Daily aerosol data observational data from satellites is only possible in cloud-free sky (33)—which clearly biases our sample in an irreconcilable way. Consequently, most satellite-based cloud research uses data products at the monthly timescale (34, 35). Using monthly data in our analysis, we can only say that classes 30 and 35 increase in share with increasing AOD, but cannot say anything about the short term dynamics evident in our sample.

Remaining unexplained variance is likely due to small scale processes, not reliably represented in ERA5 (and aerosols). While we cannot quantity test for these small scale processes with our framework, the different textures of cloud classes imply differentiating small scale processes. Colder surface, unstable stratocumulus (e.g., classes 25, 27) show more distinctive punched-out patterns. Whereas warm weather stable broken-up stratocumulus (36, 40) have a more homogeneous, softer texture which might imply a more homogeneous small scale calculation pattern rather than stochastic convection/precipitation.

Precipitation is in general low in these regions. GPM IMERG microwave satellites may struggle to measure the low levels of drizzle found in these regions. GPM IMERG does report daily precipitation values below 0.1 mm/day, which is well below the drizzle-proxy levels reported in previous open cell stratocumulus studies (24, 26). In the LES modeling studies (25), simulated precipitation in open cells is well over. A rigorous evaluation of GPM IMERG over the subtropical ocean is not available at this time. Rain gauge comparisons in other parts of the world have found an *overestimation* of the low end of the precipitation distribution (36). Previous work has employed a MODIS drizzle proxy, which is a deterministic function of liquid water path and droplet concentration (24) from MODIS observations (droplet concentration is approximated from difference in MODIS brightness temperature). We compare the drizzle proxy to the GPM IMERG measurements directly and find the drizzle proxy shows considerable rain where GPM IMERG does not (strong low end bias). Others have employed CALIPSO satellite measurements of rain water path, which would greatly reduce our MODIS based classification sample size and may show the same missing drizzle as GPM IMERG (37). Nevertheless, findings based on CALIPSO (26) report open cells are at least twice as rainy as closed cells.

Several limitations with the Lagrangian tracking scheme are also identified. MODIS sampling in this region, coupled with daytime-only observations, mean that sometimes 24 hours have passed until there is a return observation. In this time, many unseen transitions may have taken place in between the two observations. Geostationary satellite images with higher temporal sampling images should improve short term transitions analysis. A second issue is the idea that clouds always move with the prevailing winds. We do observe some apparent transitions ‘up wind’ (see Fig. 4, bottom row). Future work could employ geostationary satellite retrievals to overcome some of the sampling limitations of MODIS.

Expected climate change will modify the temperature structure of the lower atmosphere in the subtropics including increasing sea surface temperature and poleward expansion of the Hadley cell (38). EIS is projected to increase moderately (39) under expected climate change but will trade off to some degree with warmer SSTs. Observational research carefully coupled with model circulation changes should be able to inform on future cloud response to warming where model simulated clouds themselves fall short.

## Materials and Methods

**Data Availability.** All data used in this analysis is publicly available from the following sources. ERA5 data is available at [Copernicus](#). GPM IMERG data is available from [NASA](#). MODIS cloud

745 properties are available from [NASA](#) and droplet concentrations are  
746 available from [CEDA](#). Self-supervised cloud classes and MODIS  
747 cloud properties are available at [globus-labs](#). Codebase for  
748 analysis is available on [Github](#). 1.4 km climate model data  
749 shown for background only and is available upon request from the  
750 corresponding author.

751 **Self-Supervised cloud classification.** AICCA employs a rotationally-  
752 invariant autoencoder to classify clouds in 2000-2022 MODIS satel-  
753 lite images over the global oceans from both TERRA and AQUA  
754 ([16](#)). The autoencoder is trained without explicit supervision on  
755 seven near-IR and IR bands

756 Hierarchical agglomerative clustering is applied to the latent  
757 space of the autoencoder. This process generates and information-  
758 maximizing 42 cloud classes which are separable, stable, and  
759 physically reasonable ([17](#), [40](#)). See SI Fig. 14 and ([17](#), [40](#)) for  
760 methodological details.

761 An autoencoder is trained on 1 million sample  $128 \times 128$  pixels  
762 patches with six infrared radiance bands from both the TERRA  
763 and AQUA instrument. Hierarchical agglomerative clustering  
764 is then applied to the latent vectors from the autoencoder (See  
765 supplemental Fig. S14 and ([16](#), [17](#)) for details) generating an  
766 informationally-maximizing 42 cloud classes. Inference is applied  
767 to the remaining observations with the trained model and the  
768 resultant AI Cloud Classification Atlas [AICCA] consists of  $\sim 200M$   
769 labeled cloud type observations at approximately 1 degree latitude  
770 and longitude resolution over the global oceans.

771 **Meteorological mapping.** In this analysis, we focus on the three  
772 main subtropical stratocumulus regions ([27](#)). We look at the 15  
773 most populous clouds classes in these subtropical regions which  
774 are, in order: [30, 35, 40, 36, 26, 27, 39, 32, 37, 19, 29, 25, 33, 41].  
775 The class number is assigned based on cloud top pressure: smaller  
776 numbers represent high-altitude clouds and higher numbers denote  
777 low-altitude clouds.

778 Observed AICCA classes are then mapped against daily mean  
779 reanalysis meteorology variables from the ERA5 reanalysis ([41](#)).  
780 Variables are taken from the hourly product at pressure levels  
781 and averaged to daily means. They include temperature, relative  
782 humidity, divergence, geopotential, potential vorticity, u, v, w  
783 components of wind at different pressure levels between 950 and 700  
784 hpa as well as surface quantities including: sea surface temperature,  
785 mean sea level pressure, significant wave height, some engineered  
786 features including: lower tropospheric stability, estimated inversion  
787 strength (EIS), and some calculated wind shears (see supplemental  
788 table). AICCA observation patch latitude and longitude is rounded  
789 to the nearest 0.25 degree and matched with the ERA-5 value  
790 (provided at 0.25 degree spatial resolution) on the same day.  
791 Estimated inversion strength is calculated based on the method  
792

793 of ([30](#)). We also include daily and hourly precipitation from the  
794 GPM IMERG composite satellite product ([37](#)). Precipitation values  
795 are aggregated to 0.25 degrees (from native 0.1 degrees) to match  
796 ERA5 resolution. Finally, MODIS drizzle proxy is calculated from  
797 the method in ([24](#)) using droplet concentrations from ([42](#)) and  
798 liquid water path data from ([43](#)).

799 Meteorological drivers of low cloud texture are tested in the  
800 following fashion. We train a binary logistic regression model with  
801 L2 penalty and Limited-memory BFGS solver (in [scikit-learn](#) ([44](#)))  
802 for pairs of 5000 randomly-selected samples from each class in the  
803 top 15 classes. Every combination of classes is tested across all 68  
804 meteorological variables (207298 combinations in total). Pairs of  
805 meteorological driver are scored based on their mean prediction  
806 skill across all pairs of cloud classes. The process is repeated  
807 for combinations of three meteorological variables. Cloud free  
808 conditions are ignored in this step.

809 **Lagrangian tracking.** MODIS class observations are tracked in space  
810 in a Lagrangian framework. Air parcels (and the clouds they  
811 contain) are assumed to move with the winds at 925 hPa, following  
812 ([18](#)). We calculate forward trajectories by advecting each MODIS  
813 observation by the rate of the mean hourly wind speed in that  
814 location for one hour. The process is repeated until the new  
815 location and timestamp line up in the same gridcell (rounded to  
816

817 half degree for join, otherwise trajectory continues) and timestamp  
818 (rounded to hour) with another MODIS observation or until 24  
819 hours have passed, at that point the trajectory is abandoned.  
820 Transitions are observed in 1 hour and all the way up to 24 hours  
821 depending on location and the pass times of TERRA and AQUA.  
822 Total precipitation from GPM-IMERG along the trajectory is  
823 tracked at each hour. Drizzle proxy values can only be recorded  
824 where MODIS samples exist at the start or end of the trajectory.

825 **ACKNOWLEDGMENTS.** JAF was supported in part by NSF  
826 NRT program (grant DGE-1735359) and by an appointment to the  
827 Intelligence Community Postdoctoral Research Fellowship Program  
828 at Toyota Technological Institute at Chicago administered by Oak  
829 Ridge Institute for Science and Education (ORISE) through an  
830 inter agency agreement between the U.S. Department of Energy  
831 and the Office of the Director of National Intelligence (ODNI).  
832 TK was supported in part by the NSF NRT program (grant DGE-  
833 1735359). ITF was supported in part by the U.S. Department of  
834 Energy under Contract DE-AC02-06CH11357. The authors thank  
835 Valentine Anantharaj for support with 1.4 km model data and  
836 Chris Bretherton for helpful comments.

- 837 1. GL Stephens, Cloud feedbacks in the climate system: A critical review. *J. Clim.* **18**,  
838 237–273 (2005).
- 839 2. M Zhao, et al., Uncertainty in model climate sensitivity traced to representations of cumulus  
840 precipitation microphysics. *J. Clim.* **29**, 543–560 (2016).
- 841 3. SC Sherwood, et al., An assessment of Earth’s climate sensitivity using multiple lines of  
842 evidence. *Rev. Geophys.* **58** (2020).
- 843 4. MD Zelinka, et al., Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res.*  
844 *Lett.* **47** (2020).
- 845 5. T Schneider, CM Kaul, KG Pressel, Possible climate transitions from breakup of  
846 stratocumulus decks under greenhouse warming. *Nat. Geosci.* **12**, 163–167 (2019).
- 847 6. T Schneider, CM Kaul, KG Pressel, Solar geoengineering may not prevent strong warming  
848 from direct effects of CO<sub>2</sub> on stratocumulus cloud cover. *Proc. Natl. Acad. Sci.* **117**,  
849 30179–30185 (2020).
- 850 7. G Bellon, O Geoffroy, Stratocumulus radiative effect, multiple equilibria of the well-mixed  
851 boundary layer and transition to shallow convection. *Q. J. Royal Meteorol. Soc.* **142**,  
852 1685–1696 (2016).
- 853 8. P Kuma, FAM Bender, A Schuddeboom, AJ McDonald, Ø Seland, Machine learning of  
854 cloud types shows higher climate sensitivity is associated with lower cloud biases.  
855 *Atmospheric Chem. Phys.* **23**, 523–549 (2023).
- 856 9. L Bock, A Lauer, Cloud properties and their projected changes in CMIP models with  
857 low/medium/high climate sensitivity (2023) <https://doi.org/10.5194/egusphere-2023-1086>.
- 858 10. B Stevens, et al., DYAMOND: the DYnamics of the Atmospheric general circulation Modeled  
859 On Non-hydrostatic Domains. *Prog. Earth Planet. Sci.* **6**, 61 (2019).
- 860 11. NP Wedi, et al., A baseline for global weather and climate simulations at 1 km resolution. *J.*  
861 *Adv. Model. Earth Syst.* **12** (2020).
- 862 12. T Schneider, et al., Harnessing AI and computing to advance climate modelling and  
863 prediction. *Nat. Clim. Chang.* **13**, 887–889 (2023).

- 869 24. R Wood, et al., Open cellular structure in marine stratocumulus sheets. *J. Geophys. Res.*  
870 **113**, D12207 (2008). 931
- 871 25. G Feingold, I Koren, H Wang, H Xue, WA Brewer, Precipitation-generated oscillations in  
872 open cellular cloud fields. *Nature* **466**, 849–852 (2010). 932
- 873 26. D Watson-Parris, SA Sutherland, MW Christensen, R Eastman, P Stier, A large-scale  
874 analysis of pockets of open cells and their radiative impact. *Geophys. Res. Lett.* **48** (2021). 933
- 875 27. S Klein, D Hartmann, The seasonal cycle of low stratiform clouds. *J. Clim.* **6**, 1586–1606  
876 (1993). 935
- 877 28. HS Baker, C Mbengue, T Woolings, Seasonal sensitivity of the Hadley cell and  
878 cross-hemispheric responses to diabatic heating in an idealized GCM. *Geophys. Res. Lett.*  
879 **45**, 2533–2541 (2018). 936
- 880 29. I Koren, G Feingold, Aerosol–cloud–precipitation system as a predator-prey problem. *Proc.  
881 Natl. Acad. Sci.* **108**, 12227–12232 (2011). 937
- 882 30. R Wood, CS Bretherton, On the Relationship between Stratiform Low Cloud Cover and  
883 Lower-Tropospheric Stability. *J. Clim.* **19**, 6425–6432 (2006). 938
- 884 31. TA Myers, JR Norris, On the Relationships between Subtropical Clouds and Meteorology in  
885 Observations and CMIP3 and CMIP5 Models\*. *J. Clim.* **28**, 2945–2967 (2015). 939
- 886 32. DT McCoy, R Eastman, DL Hartmann, R Wood, The Change in Low Cloud Cover in a  
887 Warmed Climate Inferred from AIRS, MODIS, and ERA-Interim. *J. Clim.* **30**, 3609–3620  
888 (2017). 940
- 889 33. RC Levy, et al., The Collection 6 MODIS aerosol products over land and ocean.  
890 *Atmospheric Meas. Tech.* **6**, 2989–3034 (2013). 941
- 891 34. DT McCoy, et al., Natural aerosols explain seasonal and spatial patterns of Southern Ocean  
892 cloud albedo. *Sci. Adv.* **1**, e1500157 (2015). 942
- 893 35. Y Chen, et al., Machine learning reveals climate forcing from aerosols is dominated by  
894 increased cloud cover. *Nat. Geosci.* **15**, 609–614 (2022). 943
- 895 36. Y Xin, et al., Evaluation of IMERG and ERA5 precipitation products over the Mongolian  
896 Plateau. *Sci. Reports* **12**, 21776 (2022). 944
- 897 37. G Huffman, E Stocker, D Bolvin, E Nelkin, J Tan, GPP IMERG late precipitation L3 1 day 0.1  
898 degree x 0.1 degree V06 (2019) Accessed: August 1, 2023. 945
- 899 38. WK Lau, KM Kim, Robust hadley circulation changes and increasing global dryness due to  
900 co2 warming from cmip5 model projections. *Proc. Natl. Acad. Sci.* **112**, 3630–3635 (2015). 946
- 901 39. X Qu, A Hall, SA Klein, PM Caldwell, The strength of the tropical inversion and its response  
902 to climate change in 18 CMIP5 models. *Clim. Dyn.* **45**, 375–396 (2015). 947
- 903 40. T Kurihara, J Franke, I Foster, Z Wang, E Moyer, Insight into cloud processes from  
904 unsupervised classification with a rotationally invariant autoencoder. *arXiv preprint  
905 arXiv:2211.00860* (2022). 948
- 906 41. Hersbach, et al., The ERA5 global reanalysis. *Q. J. Royal Meteorol. Soc.* **146**, 1999–2049  
907 (2020). 949
- 908 42. E Grisepeerdt, et al., Cloud droplet number concentration, calculated from the MODIS  
909 (Moderate resolution imaging spectroradiometer) cloud optical properties retrieval and  
910 gridded using different sampling strategies (2022)  
911 <https://dx.doi.org/10.5285/864a46cc65054008857ee5bb772a2a2b>. 950
- 912 43. S Khanal, Z Wang, JR French, Improving middle and high latitude cloud liquid water path  
913 measurements from MODIS. *Atmospheric Res.* **243**, 105033 (2020). 951
- 914 44. F Pedregosa, et al., Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**,  
915 2825–2830 (2011). 952
- 916 904 953
- 917 905 954
- 918 906 955
- 919 907 956
- 920 908 957
- 921 909 958
- 922 910 959
- 923 911 960
- 924 912 961
- 925 913 962
- 926 914 963
- 927 915 964
- 928 916 965
- 929 917 966
- 930 918 967
- 931 919 968
- 932 920 969
- 933 921 970
- 934 922 971
- 935 923 972
- 936 924 973
- 937 925 974
- 938 926 975
- 939 927 976
- 940 928 977
- 941 929 978
- 942 930 979
- 943 931 980
- 944 932 981
- 945 933 982
- 946 934 983
- 947 935 984
- 948 936 985
- 949 937 986
- 950 938 987
- 951 939 988
- 952 940 989
- 953 941 990
- 954 942 991
- 955 943 992