Controls on stratocumulus texture diagnosed by deep learning

James A. Franke[a,b,1], **Takuya Kurihana**[c], **Ian T. Foster**[c,d], and **Elisabeth J. Moyer**[a]

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_2 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_2 (6) and / or high sea surface temperature (7) conditions. Coarse-resolution (100 km) global climate models provide little guidance: their shortwave cloud feedbacks differ even 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 (Figure 1, which shows the 1.4 km ECMWF model(11); for other models see SI Figure 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 pure 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: [a]University of Chicago, Department of the Geophysical Sciences, 5734 S Ellis, Chicago, IL 60637; [b]Toyota Technological Institute at Chicago, 6045 S Kenwood Ave, Chicago, IL 60637; [c]University of Chicago, Department of Computer Science, 5730 S Ellis, Chicago, IL 60637; [d]Argonne National Laboratory, 9700 S Cass Ave, Lemont, IL 60439
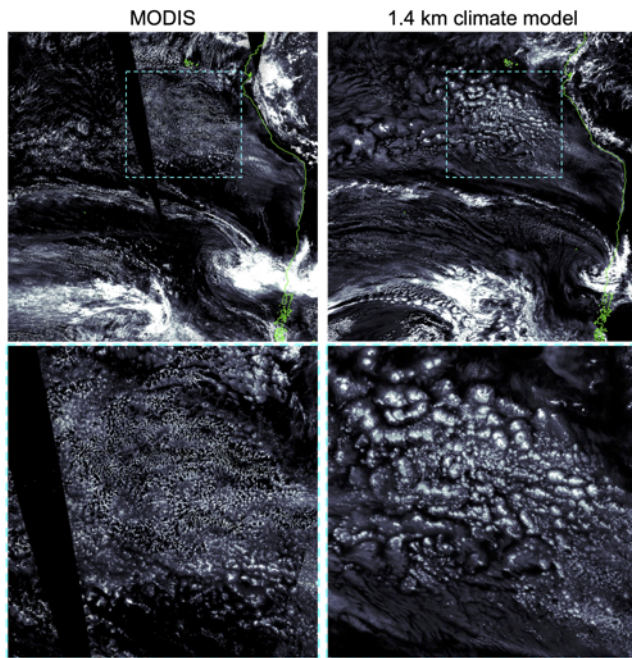
Fig. 1. Cloud textures off the coast of South America: liquid water path from **[left]** MODIS observations (Nov. 1, 2018) and **[right]** a 1.4 km horizontal resolution simulation of the ECMWF model, initialized with observations in the morning of Nov. 1, 2018 (11), so that weather patterns should match. 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 zoom 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 defintion. (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 128×128 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, 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 GMP 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 Figure 2 and SI Figure S2 for mean class properties and SI Figure 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 Figure 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 Figure 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 Figure S5). This separation is possibly only given texture-based classes and cannot be reproduced using mean cloud properties alone.
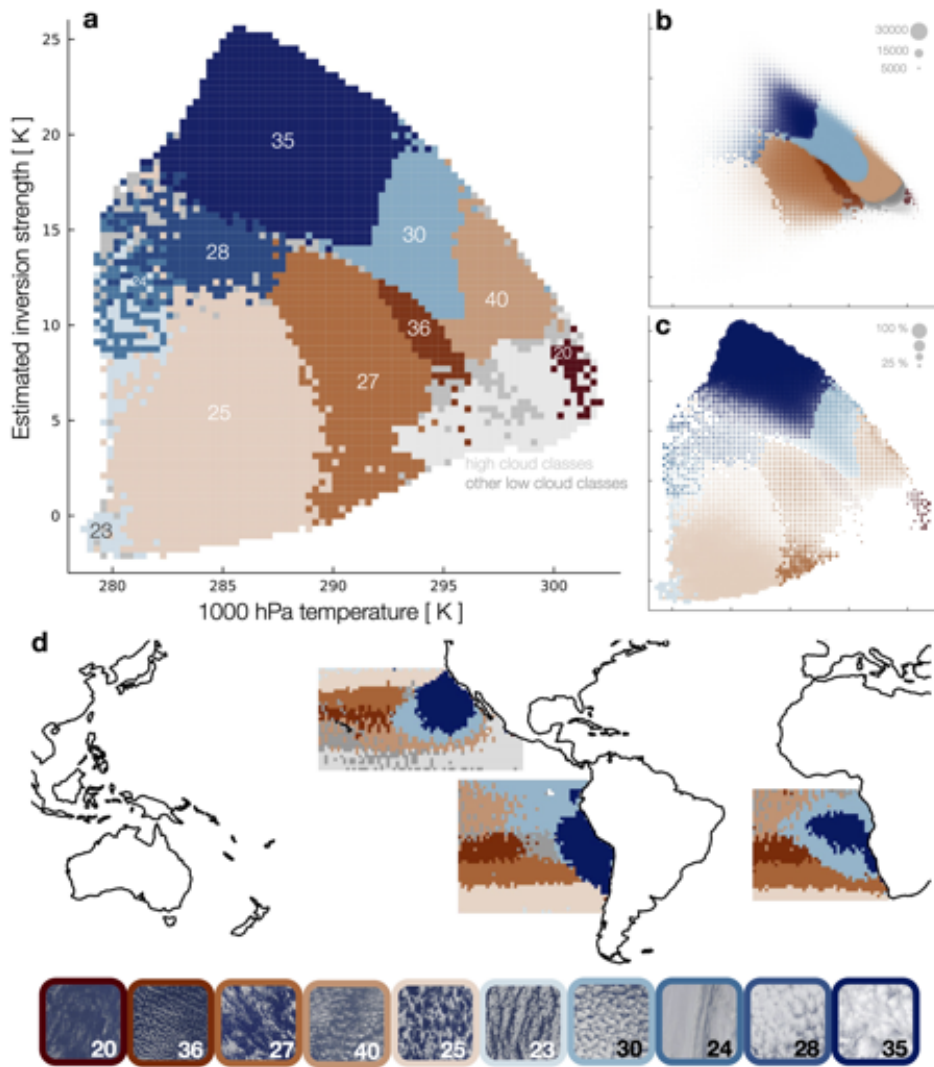
Fig. 2. AICCA cloud classes mapped to a meteorological domain of near-surface temperature and inversion strength, for all patches in the three main subtropical stratocumulus regions. **a** The 10 most dominant classes in meteorological space, color-coded. Thumbnail images of typical class members are 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 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 (Figure 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 Figure 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 Figure 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 (Figure 2d; and see SI Figure 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 (Figure 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,
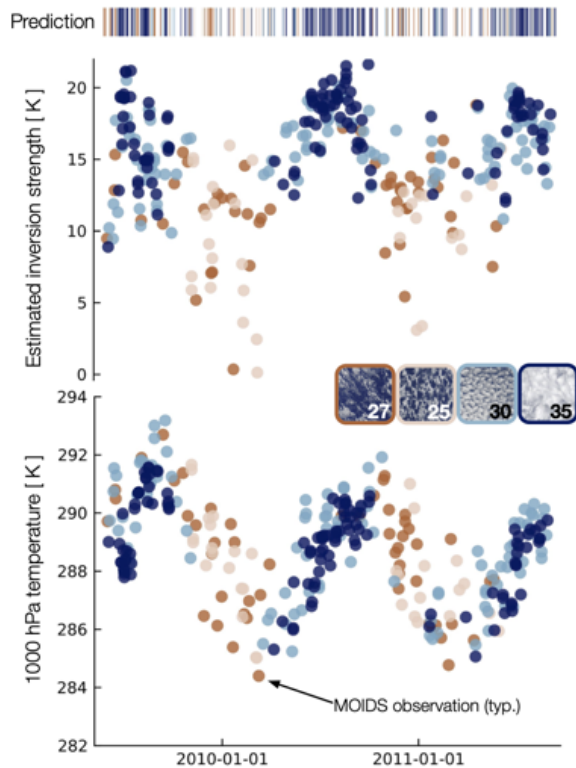
Franke *et al.*

PNAS — **September 28, 2023** — vol. XXX — no. XX — **3**

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 Figures 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 Figure 2 XX 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 XX). 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. Figure 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 the explanation for...XXX for does not explain the breakup in otherwise stable conditions. Precipitation is not particularly predictive for cloud textures across the three sub-tropical regions in general. Including precipitation measured in the logistic regression improves prediction only 0.01% on average for the top 15 classes and actually slightly degrades prediction for the top four (25,27,30,35). This is perhaps unsurprising, given how little it rains in these regions (). Two thirds of all samples experience very little precipitation (less than 0.1 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 samples, 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 mm day⁻1, Figure 5b). Classes are nearly evenly distributed across the precipitation sample space.

Precipitation may have a lag effect and play a role not in static observed classes but in the transitions between classes. To test this we track transitions between 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. Transitions from class 30 are binned into three separate cases: cloud degradation, enhancement, and no change. The case with the highest rainfall is actually the no-transition case (#30 remains #30, Figure 5, panel a). Transitions from 30 to much sparser classes (20, 25, 27, 36, 40) are actually less rainy than apparent cloud enhancement transitions from 30 to the unbroken class #35, though this is statistically insignificant (2-sided T test). The Y axis included all no precipitation observations as well, which dominate and thus result in very low mean drizzle rates.

We also test this relationship with the MODIS drizzle proxy, calculated from MODIS estimated droplet concentration number and liquid water path following previous work. We find almost the same answer. Open cell type classes are not more rainy than closed cell classes. Across a 2010-2013 subset of our Lagrangian transition sample, drizzle increases moderately (0.8 mm day⁻1 to 1.0 mm⁻1 median) for 30 to 30 cases, and more substantially for the 30 to 35 case. However, class 30 samples transition to open classes are are somewhat more rainy (median 1.0 mm day⁻1) than those 30s which either remain unchained or thicken into 35. Resultant open cell classes have the least rain (0.4 mm day⁻1), contrary to previous work. We also 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).

Discussion

Self-supervised cloud classification can reveal processes driving cloud formation and dynamics. The diversity of low
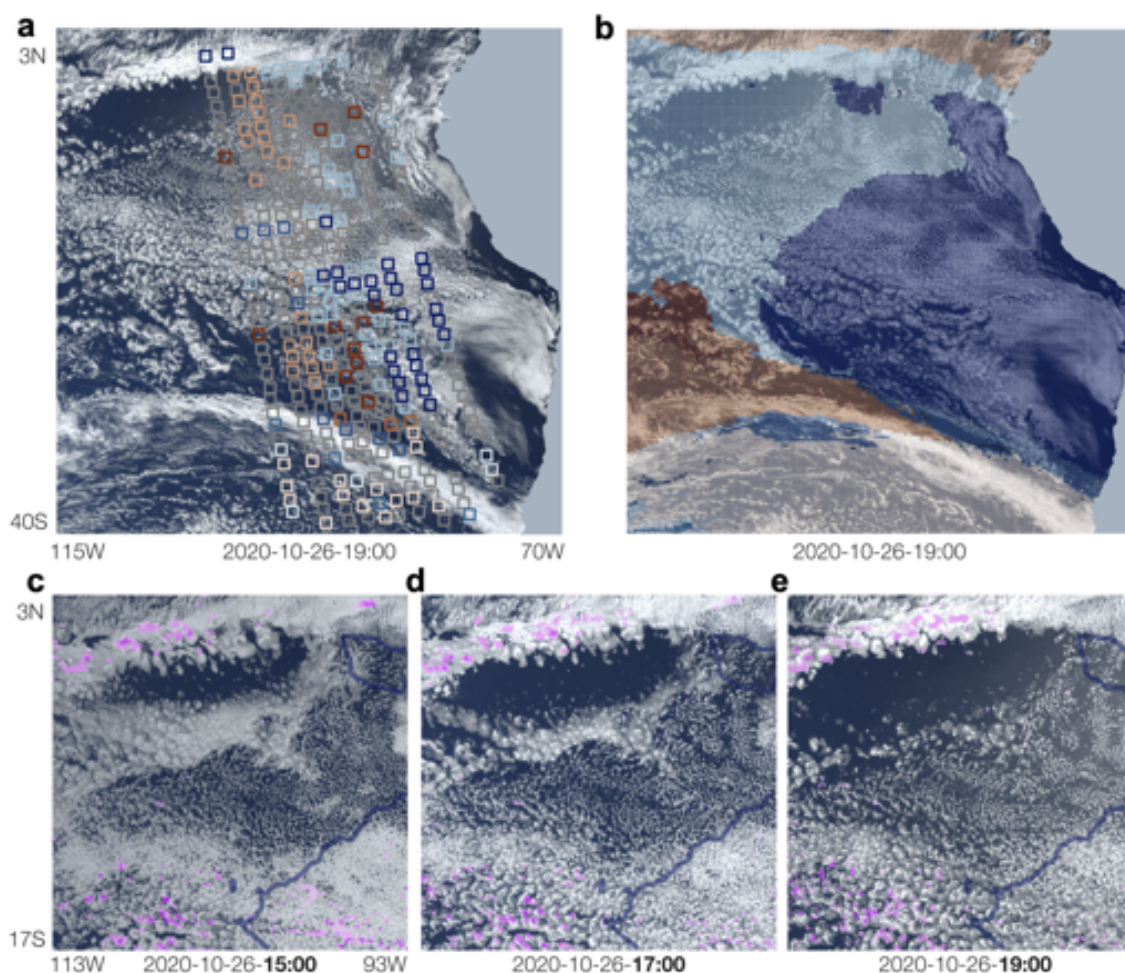
Fig. 4. [Top row] Visible light image from GOES-16 on October 26, 2020 at 19:00 UTC (approximately 3 PM in the afternoon local time) with coincident MODIS-AQUA pass cloud classes (left) and with meteorological predicted class plotted above in color (right). While the temperature structure predicts some of the visual texture on the larger-scale, many of the textures are unexplained by this framework. **[Bottom row]** Example zoom-in shows clear sky hole expanding over the course of 4 hours from approximately 11 AM local time until 3 PM. 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.

cloud morphology in the subtropics is highlighted by the approach. Our atomic cloud classification algorithm identifies over 25 cloud types in what what would classically be termed stratocumulus. A large portion of this diversity in taxonomy is easily explained by the temperature structure of the lower troposphere, confirming previous observational work focused on mean cloud fraction. 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, hysterisis 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 Figures 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
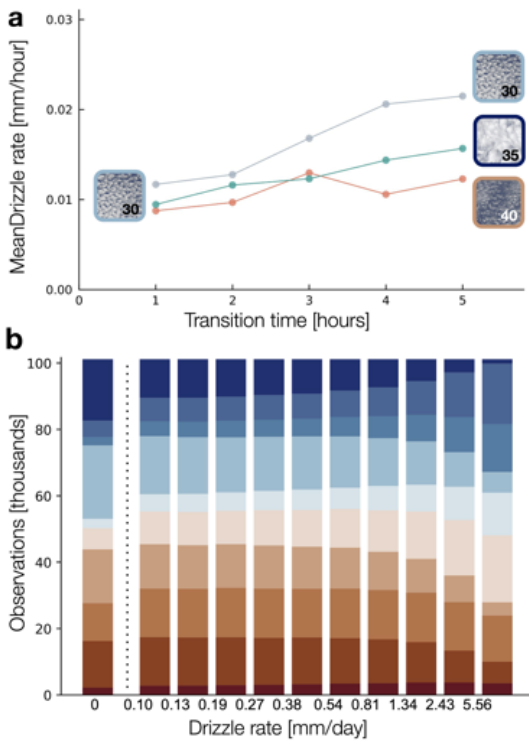
Fig. 5. Population level precipitation impacts. Drizzle does not aid in class prediction. [Top] Short timescale transitions (<6h) from the most populous class # are tracked in the south pacific basin. 11,000 such transitions are observed via a Lagrangian tracking scheme between MODIS passes. Mean precipitation 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. [bottom] 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.

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 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 necessary more likely to be located outside their region of temperature space, with a few potential exceptions. See SI Figure 10.

Aerosols are clearly critical to cloud processes. However, we simply do not have sufficient large scale aerosol observational data to include in or framework. Daily aerosol data observational data from satellites is only possible in cloud-free sky (29)—which clearly biases our sample in an irreconcilable way. Consequently, most satellite-based cloud research uses data products are at the monthly timescale(30, 31). 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 like due to small scale processes, not reliable 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 rigours 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 (32). 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). 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 (33). 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 take 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 Figure 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 (34). However, ENSO, Walker circulation changes, and many other feature of topical and subtropical climate change are less well known. 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

available from CEDA. Self-supervised cloud classes and MODIS cloud properties are available at globus-labs. 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. AICCA employs a rotationally-invariant autoencoder to classify clouds in 2000-2022 MODIS satellite images over the global oceans from both TERRA and AQUA (16). The autoencoder is trained without explicit supervision on seven near-IR and IR bands

Hierarchical agglomerative clustering is applied to the latent space of the autoencoder. This process generates and information-maximizing 42 cloud classes which are separable, stable, and physically reasonable (17, 35). See SI Figure 12 and (17, 35) for methodological details.

An autoencoder is trained on 1 million sample 128×128 pixels patches with six infrared radiance bands from both the TERRA and AQUA instrument. Hierarchical agglomerative clustering is then applied to the latent vectors from the autoencoder (See supplemental Figure 1 and (16, 17) for details) generating an informationally-maximizing 42 cloud classes. Inference is applied to the remaining observations with the trained model and the resultant AI Cloud Classification Atlas [AICCA] consists of ~200M labeled cloud type observations at approximately 1 degree latitude and longitude resolution over the global oceans.

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

Observed AICCA classes are then mapped against daily mean reanalysis meteorology variables from the ERA5 reanalysis (36). Variables are taken from the hourly product at pressure levels and averaged to daily means. They include temperature, relative humidity, divergence, geopotential, potential vorticity, u, v, w components of wind at different pressure levels between 950 and 700 hpa as well as surface quantities including: sea surface temperature, mean sea level pressure, significant wave height, some engineered features including: lower tropospheric stability, estimated inversion strength (EIS), and some calculated wind shears (see supplemental table). AICCA observation patch latitude and longitude is rounded to the nearest 0.25 degree and matched with the ERA-5 value (provided at 0.25 degree spatial resolution) on the same day. Estimated inversion strength is calculated based on (37). We also include daily and hourly precipitation from the GPM IMERG composite satellite product (33). Precipation values are aggregated to 0.25 degrees (from native 0.1 degrees) to match ERA5 resolution. Finally, MODIS drizzle proxy is calculated from the method in (24) using droplet concentrations from (38) and liquid water path data from (39).

Meteorological drivers of low cloud texture are tested in the following fashion. We train a binary logistic regression model with L2 penalty and Limited-memory BFGS solver (in *scikit-learn* (40)) for pairs of 5000 randomly-selected samples from each class in the top 15 classes. Every combination of classes is tested across all 68 meteorological variables (207298 combinations in total). Pairs of meteorological driver are scored based on their mean prediction skill across all pairs of cloud classes. The process is repeated for combinations of three meteorological variables. Cloud free conditions are ignored in this step.

Lagrangian tracking. MODIS class observations are tracked in space in a Lagrangian framework. Air parcels (and the clouds they contain) are assumed to move with the winds at 925 hPa, following (18). We calculate forward trajectories by advecting each MODIS observation by the rate of the mean hourly wind speed in that location for one hour. The process is repeated until the new location and timestamp line up in the same gridcell (rounded to half degree for join, otherwise trajectory continues) and timestamp (rounded to hour) with another MODIS observation or until 24 hours have passed, at that point the trajectory is abandoned. Transitions are observed in 1 hour and all the way up to 24 hours depending on location and the pass times of TERRA and AQUA. Total precipitation from GPM-IMERG along the trajectory is tracked at each hour. Drizzle proxy values can only be recorded where MODIS samples exist at the start or end of the trajectory.

1. GL Stephens, Cloud feedbacks in the climate system: A critical review. *J. Clim.* **18**, 237–273 (2005).
2. M Zhao, et al., Uncertainty in model climate sensitivity traced to representations of cumulus precipitation microphysics. *J. Clim.* **29**, 543–560 (2016).
3. SC Sherwood, et al., An assessment of Earth's climate sensitivity using multiple lines of evidence. *Rev. Geophys.* **58** (2020).
4. MD Zelinka, et al., Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.* **47** (2020).
5. T Schneider, CM Kaul, KG Pressel, Possible climate transitions from breakup of stratocumulus decks under greenhouse warming. *Nat. Geosci.* **12**, 163–167 (2019).
6. T Schneider, CM Kaul, KG Pressel, Solar geoengineering may not prevent strong warming from direct effects of CO2 on stratocumulus cloud cover. *Proc. Natl. Acad. Sci.* **117**, 30179–30185 (2020).
7. G Bellon, O Geoffroy, Stratocumulus radiative effect, multiple equilibria of the well-mixed boundary layer and transition to shallow convection. *Q. J. Royal Meteorol. Soc.* **142**, 1685–1696 (2016).
8. P Kuma, FAM Bender, A Schuddeboom, AJ McDonald, Ø Seland, Machine learning of cloud types shows higher climate sensitivity is associated with lower cloud biases. *Atmospheric Chem. Phys.* **23**, 523–549 (2023).
9. L Bock, A Lauer, Cloud properties and their projected changes in CMIP models with low/medium/high climate sensitivity (2023) https://doi.org/10.5194/egusphere-2023-1086.
10. B Stevens, et al., DYAMOND: the DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Prog. Earth Planet. Sci.* **6**, 61 (2019).
11. NP Wedi, et al., A baseline for global weather and climate simulations at 1 km resolution. *J. Adv. Model. Earth Syst.* **12** (2020).
12. T Schneider, et al., Harnessing AI and computing to advance climate modelling and prediction. *Nat. Clim. Chang.* **13**, 887–889 (2023).
13. T Schneider, et al., Climate goals and computing the future of clouds. *Nat. Clim. Chang.* **7**, 3–5 (2017).
14. WB Rossow, RA Schiffer, Advances in understanding clouds from ISCCP. *Bull. Am. Meteorol. Soc.* **80**, 2261–2287 (1999).
15. AH Young, KR Knapp, A Inamdar, W Hankins, WB Rossow, The International Satellite Cloud Climatology Project H-Series climate data record product (2018) https://www.ncei.noaa.gov/products/international-satellite-cloud-climatology.
16. T Kurihana, E Moyer, R Willett, D Gilton, I Foster, Data-driven cloud clustering via a rotationally invariant autoencoder. *IEEE Transactions on Geosci. Remote. Sens.* **60**, 1–25 (2021).
17. T Kurihana, EJ Moyer, IT Foster, AICCA: AI-Driven Cloud Classification Atlas. *Remote. Sens.* **14**, 5690 (2022).
18. R Eastman, R Wood, Factors controlling low-cloud evolution over the eastern subtropical oceans: A Lagrangian perspective using the A-Train satellites. *J. Atmospheric Sci.* **73**, 331–351 (2016).
19. D Randall, Conditional stability of the first kind upside-down. *Am. Meteorol. Soc.* **37**, 125–130 (1980).
20. R Wood, Stratocumulus clouds. *Mon. Weather. Rev.* **140**, 2373–2423 (2012).
21. EM Agee, TS Chen, KE Dowell, A review of mesoscale cellular convection. *Bull. Am. Meteorol. Soc.* **54**, 1004–1012 (1973).
22. B Stevens, et al., Pockets of open cells and drizzle in marine stratocumulus. *Bull. Am. Meteorol. Soc.* **86**, 51–58 (2005).
23. TM Sharon, et al., Aerosol and cloud microphysical characteristics of rifts and gradients in maritime stratocumulus clouds. *J. Atmospheric Sci.* **63**, 983–997 (2006).
24. R Wood, et al., Open cellular structure in marine stratocumulus sheets. *J. Geophys. Res.* **113**, D12207 (2008).

Franke *et al.*

PNAS — **September 28, 2023** — vol. XXX — no. XX — 7

25. G Feingold, I Koren, H Wang, H Xue, WA Brewer, Precipitation-generated oscillations in open cellular cloud fields. *Nature* **466**, 849–852 (2010).

26. D Watson-Parris, SA Sutherland, MW Christensen, R Eastman, P Stier, A large-scale analysis of pockets of open cells and their radiative impact. *Geophys. Res. Lett.* **48** (2021).

27. S Klein, D Hartmann, The seasonal cycle of low statiform clouds. *J. Clim.* **6**, 1586–1606 (1993).

28. HS Baker, C Mbengue, T Woollings, Seasonal sensitivity of the Hadley cell and cross-hemispheric responses to diabatic heating in an idealized GCM. *Geophys. Res. Lett.* **45**, 2533–2541 (2018).

29. RC Levy, et al., The Collection 6 MODIS aerosol products over land and ocean. *Atmospheric Meas. Tech.* **6**, 2989–3034 (2013).

30. DT McCoy, et al., Natural aerosols explain seasonal and spatial patterns of Southern Ocean cloud albedo. *Sci. Adv.* **1**, e1500157 (2015).

31. Y Chen, et al., Machine learning reveals climate forcing from aerosols is dominated by increased cloud cover. *Nat. Geosci.* **15**, 609–614 (2022).

32. Y Xin, et al., Evaluation of IMERG and ERA5 precipitation products over the Mongolian Plateau. *Sci. Reports* **12**, 21776 (2022).

33. 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.

34. WK Lau, KM Kim, Robust hadley circulation changes and increasing global dryness due to co2 warming from cmip5 model projections. *Proc. Natl. Acad. Sci.* **112**, 3630–3635 (2015).

35. T Kurihana, J Franke, I Foster, Z Wang, E Moyer, Insight into cloud processes from unsupervised classification with a rotationally invariant autoencoder. *arXiv preprint arXiv:2211.00860* (2022).

36. Hersbach, et al., The ERA5 global reanalysis. *Q. J. Royal Meteorol. Soc.* **146**, 1999–2049 (2020).

37. R Eastman, R Wood, K Ting O, The subtropical stratocumulus-topped planetary boundary layer: A climatology and the Lagrangian evolution. *J. Atmospheric Sci.* **74**, 2633–2656 (2017).

38. E Gryspeerdt, 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) https://dx.doi.org/10.5285/864a46cc65054008857ee5bb772a2a2b.

39. S Khanal, Z Wang, JR French, Improving middle and high latitude cloud liquid water path measurements from MODIS. *Atmospheric Res.* **243**, 105033 (2020).

40. F Pedregosa, et al., Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).