

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

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 CO₂ 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 CO₂ (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: ^aUniversity of Chicago, Department of the Geophysical Sciences, 5734 S Ellis, Chicago, IL 60637; ^bToyota Technological Institute at Chicago, 6045 S Kenwood Ave, Chicago, IL 60637; ^cUniversity of Chicago, Department of Computer Science, 5730 S Ellis, Chicago, IL 60637; ^dArgonne 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.

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

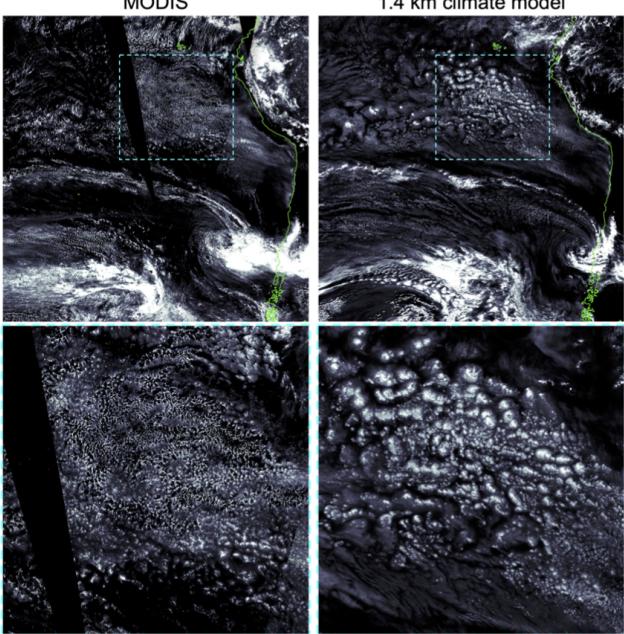


Fig. 1. Cloud textures off the coast of South America: liquid water path from [left] MODIS observations (Nov. 1, 2018) and [right] a 1.4 km horizontal resolution simulation of the ECMWF model, initialized with observations in the morning of Nov. 1, 2018 (11), so that weather patterns should match. Color 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 128×128 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 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.

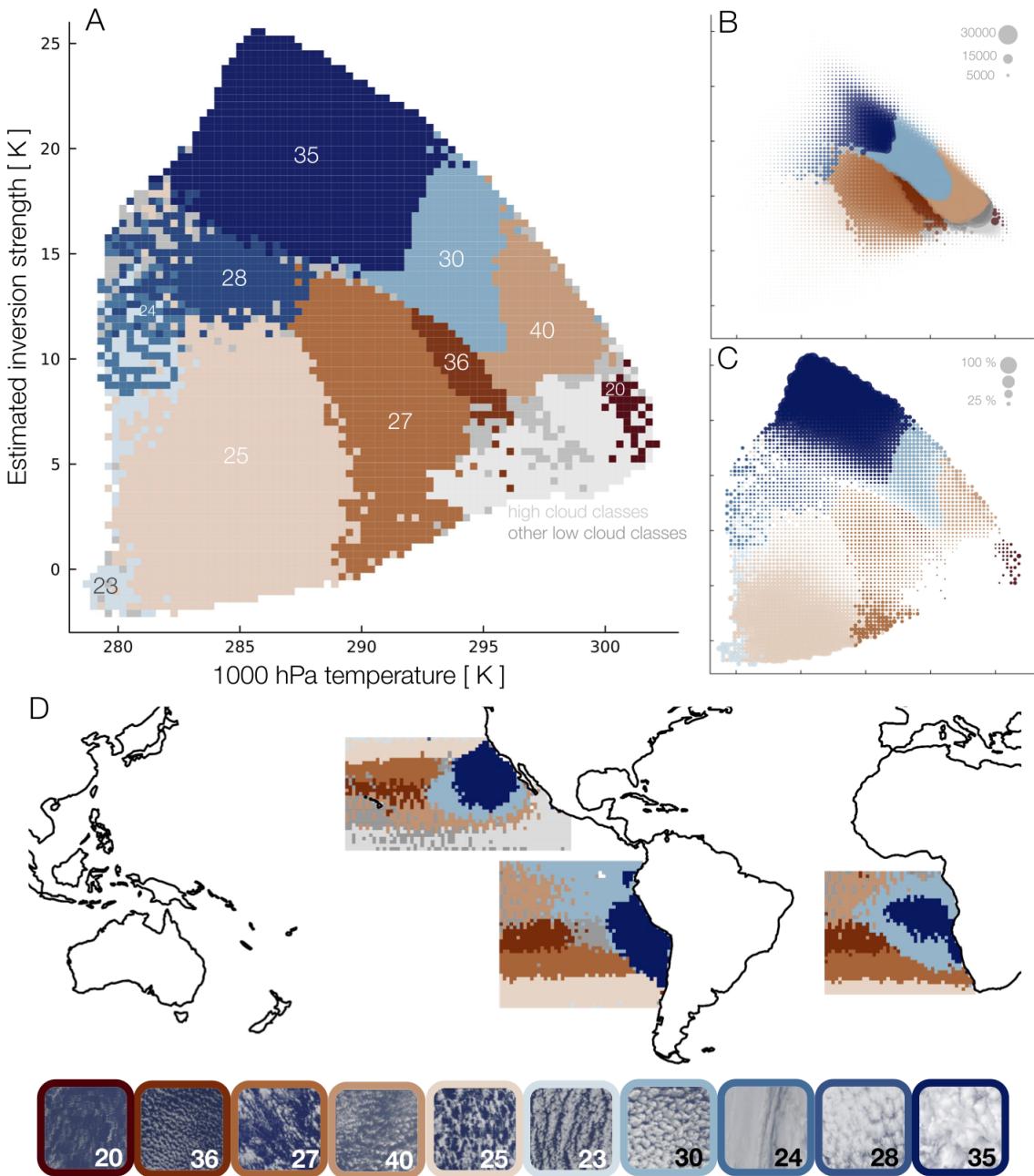


Fig. 2. AICCA cloud classes mapped to a meteorological domain of near-surface temperature and inversion strength, for all patches in the three main subtropical stratocumulus regions. (A) The 10 most dominant classes in meteorological space, color-coded. Thumbnail images of typical class members are shown 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.

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 (28), 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

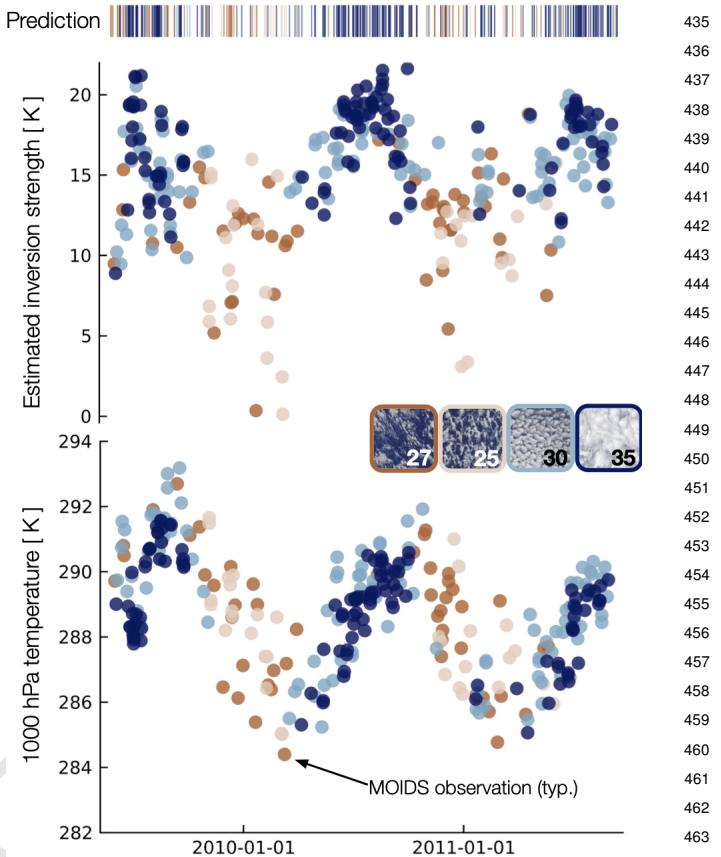


Fig. 3. Timeseries of meteorological variables and observed and predicted cloud classes for a single representative location at 29N, 130W in the North Pacific basin. Top shows inversion strength and bottom near-surface temperature, with observed cloud classes shown as colors for the four main classes 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.

stratocumulus (25, 29). 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 Figure 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

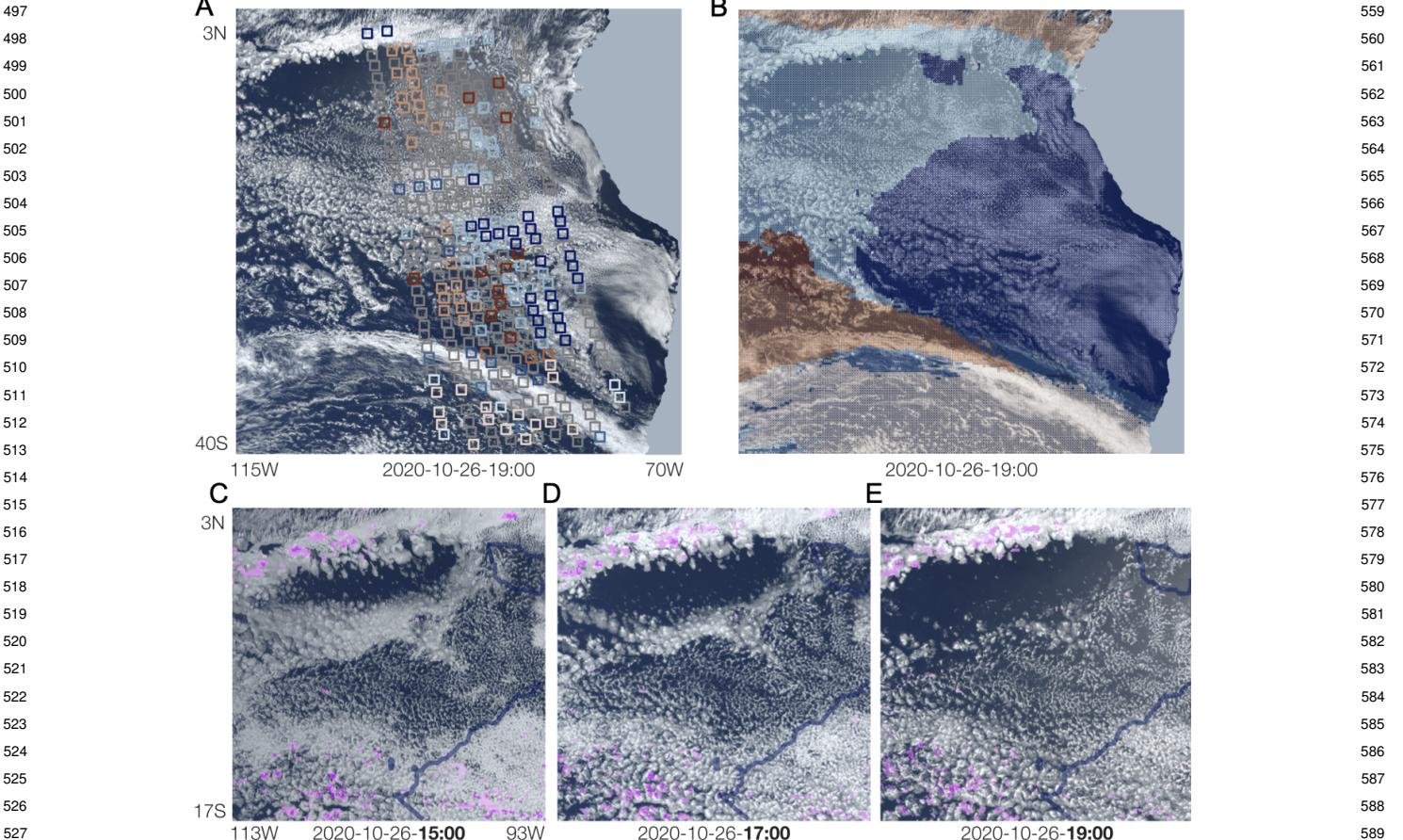


Fig. 4. (A) Visible image from GOES-16 on October 26, 2020 at 19:00 UTC (~3 PM local time) with coincident MODIS-AQUA cloud classes, following the same color code used throughout, and (B) with meteorologically-predicted cloud classes. While the temperature structure predicts large-scale features of the 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.

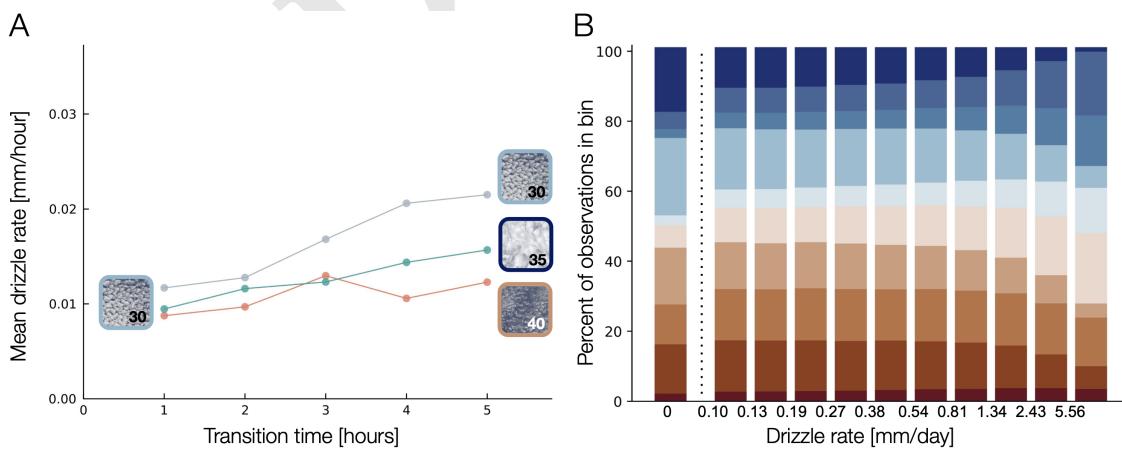


Fig. 5. Population level precipitation impacts. Drizzle does not aid in explaining texture. (A) Short timescale transitions (<6h) from the most populous class # are tracked in the South Pacific basin. 3,000 such transitions are observed via Lagrangian tracking between MODIS TERRA and AQUA passes. Mean 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.

stratocumulus studies (e.g. 24), though more recent work has used IMERG (30). The two estimates differ by a factor of 10 (IMERG is drier) and show essentially no correlation (SI Fig. S13). 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⁻¹ for #30 to #30), increases more substantially for the thickening cases (0.6 to 1.5 mm day⁻¹ for #30 to #35), and actually decreases for the closed- to open-cell transitions (1.0 to 0.4 mm day⁻¹ 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 (23, 24, 26).

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⁻¹ (in IMERGE). Most transitions of stratocumulus texture involve no measurable drizzle. It is therefore unsurprising that 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⁻¹). Cloud types are

two thirds of all samples experience very little precipitation at all. Finally we include 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). 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 mm day⁻¹, Fig. 5b). Classes are nearly evenly-distributed across the precipitation sample space.

Discussion

Self-supervised cloud classification is shown to help reveal processes driving cloud formation and dynamics and the diversity of low cloud morphology in the subtropics is highlighted by the approach. Our automatic cloud classification algorithm identifies over 25 cloud types in what would classically be termed stratocumulus. A large portion of this low cloud diversity in taxonomy is easily explained by the temperature structure of the lower troposphere, confirming previous observational work focused on mean cloud fraction (18, 31–33, 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 textural variation remains unexplained by this simple framework. We identify the following possibilities for improving this discrepancy: additional meteorological drivers, hysteresis or lag effects, regional differences, and within class variation. Additional meteorological variables from reanalysis can only increase prediction about skill 1–2%. Whether this is because these factors are not important to cloud dynamics or because they

are important at scales below 25 km (ERA5 resolution), we cannot readily say with this approach based on ERA5 data. Likewise, we find no regional differences in prediction skill, 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 regimes. 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 necessarily more likely to be located outside their region of temperature space, with a few potential exceptions. See SI Fig. 10.

Remaining unexplained variance in cloud texture is likely due to small scale processes, not reliably represented in ERA5. While we cannot quantity test for these small scale processes with our framework, the different textures of cloud classes might 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 larger stochastic convection/precipitation.

Three additional limitations with our study include precipitation measurements, aerosols and the Lagrangian scheme. A major potential caveat with this work is that the GPM IMERG product may struggle to measure the low levels of drizzle found in these regions. A rigorous evaluation of GPM IMERG over the subtropical ocean is not available at this time, but rain gauge comparisons in other parts of the world have found an *overestimation* of the low end of the precipitation distribution (34). 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). We directly compare the MODIS drizzle proxy (deterministic function of liquid water path and droplet concentration (24)) to the GPM IMERG measurements and find the drizzle proxy shows considerable rain where GPM IMERG does not (strong low end bias, see SI Fig. S13). Others have employed CALIPSO satellite measurements of rain water path. CALIPSO measurements would greatly reduce our MODIS based classification sample size and may show the same missing drizzle as GPM IMERG. More research is needed to evaluate these remote sensing precipitation monitoring methods.

Aerosols—clearly critical to cloud processes—are not included in our framework because sufficient large scale aerosol observational data do not exist. Daily aerosol data observational data from satellites is only possible in cloud-free sky (35)—which clearly biases our sample in an irreconcilable way. Consequently, most satellite-based cloud research uses data products at the monthly timescale (36, 37). Using monthly data in our analysis, we can only report that classes 30 and 35 broadly increase in share with increasing AOD, but cannot say anything about the short term dynamics evident in our sample.

Several limitations with the Lagrangian tracking scheme are also identified. MODIS sampling in this region, further reduced to daytime-only observations, cause return period to run past 24 in most cases. Many unseen transitions may take place in between the two observations. This severely limits our abilities to test shorter term transitions and potential hysteresis. A second issue is the assumption 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 properties are available from [NASA](#) and droplet concentrations are 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, 40). See SI Fig. 14 and (17, 40) 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 Fig. S14 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 (41). 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 the method of (31). We also include daily and hourly precipitation from the GPM IMERG composite satellite product (42). Precipitation 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 (43) and liquid water path data from (44).

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* (45)) 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.

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

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).
- Frank 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**, 177–190 (2016).
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.ncdc.noaa.gov/products/international-satellite-cloud-climatology>.

- 869 16. T Kurihana, E Moyer, R Willett, D Gilton, I Foster, Data-driven cloud clustering via a
870 rotationally invariant autoencoder. *IEEE Transactions on Geosci. Remote. Sens.* **60**, 1–25
(2021). 931
871 17. T Kurihana, EJ Moyer, IT Foster, AICCA: AI-Driven Cloud Classification Atlas. *Remote.
Sens.* **14**, 5690 (2022). 932
872 18. R Eastman, R Wood, Factors controlling low-cloud evolution over the eastern subtropical
873 oceans: A Lagrangian perspective using the A-Train satellites. *J. Atmospheric Sci.* **73**,
874 331–351 (2016). 933
875 19. D Randall, Conditional stability of the first kind upside-down. *Am. Meteorol. Soc.* **37**,
876 125–130 (1980). 934
877 20. R Wood, Stratocumulus clouds. *Mon. Weather. Rev.* **140**, 2373–2423 (2012). 935
878 21. EM Agee, TS Chen, KE Dowell, A review of mesoscale cellular convection. *Bull. Am.
Meteorol. Soc.* **54**, 1004–1012 (1973). 936
879 22. B Stevens, et al., Pockets of open cells and drizzle in marine stratocumulus. *Bull. Am.
Meteorol. Soc.* **86**, 51–58 (2005). 937
880 23. TM Sharon, et al., Aerosol and cloud microphysical characteristics of rifts and gradients in
881 maritime stratocumulus clouds. *J. Atmospheric Sci.* **63**, 983–997 (2006). 938
882 24. R Wood, et al., Open cellular structure in marine stratocumulus sheets. *J. Geophys. Res.*
883 **113**, D12207 (2008). 939
884 25. G Feingold, I Koren, H Wang, H Xue, WA Brewer, Precipitation-generated oscillations in
885 open cellular cloud fields. *Nature* **466**, 849–852 (2010). 940
886 26. D Watson-Parris, SA Sutherland, MW Christensen, R Eastman, P Stier, A large-scale
887 analysis of pockets of open cells and their radiative impact. *Geophys. Res. Lett.* **48** (2021). 941
888 27. S Klein, D Hartmann, The seasonal cycle of low stratiform clouds. *J. Clim.* **6**, 1586–1606
889 (1993). 942
890 28. HS Baker, C Mbengue, T Woollings, Seasonal sensitivity of the Hadley cell and
891 cross-hemispheric responses to diabatic heating in an idealized GCM. *Geophys. Res. Lett.*
892 **45**, 2533–2541 (2018). 943
893 29. I Koren, G Feingold, Aerosol–cloud–precipitation system as a predator-prey problem. *Proc.
Natl. Acad. Sci.* **108**, 12227–12232 (2011). 944
894 30. MW Christensen, WK Jones, P Stier, Aerosols enhance cloud lifetime and brightness along
895 the stratus-to-cumulus transition. *Proc. Natl. Acad. Sci.* **117**, 17591–17598 (2020). 945
896 31. R Wood, CS Bretherton, On the Relationship between Stratiform Low Cloud Cover and
897 Lower-Tropospheric Stability. *J. Clim.* **19**, 6425–6432 (2006). 946
898 32. TA Myers, JR Norris, On the Relationships between Subtropical Clouds and Meteorology in
899 Observations and CMIP3 and CMIP5 Models*. *J. Clim.* **28**, 2945–2967 (2015). 947
900 33. DT McCoy, R Eastman, DL Hartmann, R Wood, The Change in Low Cloud Cover in a
901 Warmed Climate Inferred from AIRS, MODIS, and ERA-Interim. *J. Clim.* **30**, 3609–3620
902 (2017). 948
903 34. Y Xin, et al., Evaluation of IMERG and ERA5 precipitation products over the Mongolian
904 Plateau. *Sci. Reports* **12**, 21776 (2022). 949
905 35. RC Levy, et al., The Collection 6 MODIS aerosol products over land and ocean.
906 *Atmospheric Meas. Tech.* **6**, 2989–3034 (2013). 950
907 36. DT McCoy, et al., Natural aerosols explain seasonal and spatial patterns of Southern Ocean
908 cloud albedo. *Sci. Adv.* **1**, e1500157 (2015). 951
909 37. Y Chen, et al., Machine learning reveals climate forcing from aerosols is dominated by
910 increased cloud cover. *Nat. Geosci.* **15**, 609–614 (2022). 952
911 38. WK Lau, KM Kim, Robust hadley circulation changes and increasing global dryness due to
912 co2 warming from cmip5 model projections. *Proc. Natl. Acad. Sci.* **112**, 3630–3635 (2015). 953
913 39. X Qu, A Hall, SA Klein, PM Caldwell, The strength of the tropical inversion and its response
914 to climate change in 18 CMIP5 models. *Clim. Dyn.* **45**, 375–396 (2015). 954
915 40. T Kurihana, J Franke, I Foster, Z Wang, E Moyer, Insight into cloud processes from
916 unsupervised classification with a rotationally invariant autoencoder. *arXiv preprint*
917 arXiv:2211.00860 (2022). 955
918 41. Hersbach, et al., The ERA5 global reanalysis. *Q. J. Royal Meteorol. Soc.* **146**, 1999–2049
919 (2020). 956
920 42. G Huffman, E Stocker, D Bolvin, E Nelkin, J Tan, GPP IMERG late precipitation L3 1 day 0.1
921 degree x 0.1 degree V06 (2019) Accessed: August 1, 2023. 957
922 43. E Gryspenert, et al., Cloud droplet number concentration, calculated from the MODIS
923 (Moderate resolution imaging spectroradiometer) cloud optical properties retrieval and
924 gridded using different sampling strategies (2022)
925 https://dx.doi.org/10.5285/864a46cc5054008857ee5bb772a2a2b. 958
926 44. S Khanal, Z Wang, JR French, Improving middle and high latitude cloud liquid water path
927 measurements from MODIS. *Atmospheric Res.* **243**, 105033 (2020). 959
928 45. F Pedregosa, et al., Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**,
929 2825–2830 (2011). 960
930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992