

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

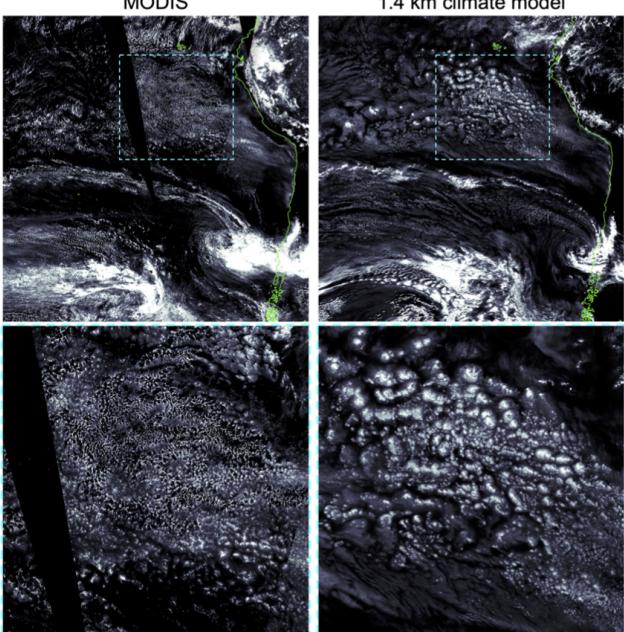


Fig. 1. Cloud textures off the coast of South America: liquid water path from [left] MODIS (22) observations (Nov. 1, 2018) and [right] a 1.4 km horizontal resolution simulation of the ECMWF model (11), initialized with observations on the morning of Nov. 1, 2018, so that weather patterns should match.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

In this work we make use of a new, deep-learning-based classification of ocean clouds, the AI-driven Cloud Classification Atlas (AICCA), in which 25 of 42 classes would clearly be considered stratocumulus in the ISCCP definition. (See Methods.) While supervised deep learning has recently been applied to cloud classification (e.g. 16–18), AICCA is instead generated with *self-supervised* classification 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 (19, 20). While observational studies of marine stratocumulus have largely focused on its *fractional occurrence*, and have shown that this occurrence is predictable from lower tropospheric stability (21), the AICCA classes allow asking instead what variables predict the *texture* of those clouds. Textures in turn provide insight into the processes that govern stratocumulus formation and evolution.

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

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

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

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

Results

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

To diagnose factors controlling these textures, we map the top 15 most populous classes against a battery of 68 meteorological quantities derived from reanalysis, including surface temperature, lower tropospheric stability, vertical motion, wind shear, and humidity. (See SI Table 1.) Variables are tested in pairs and scored by their ability to separate the 15 cloud classes from one another. The two 2-variable combinations with the highest predictive power across classes are inversion strength paired with either near-surface air temperature or lower tropospheric stability. 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

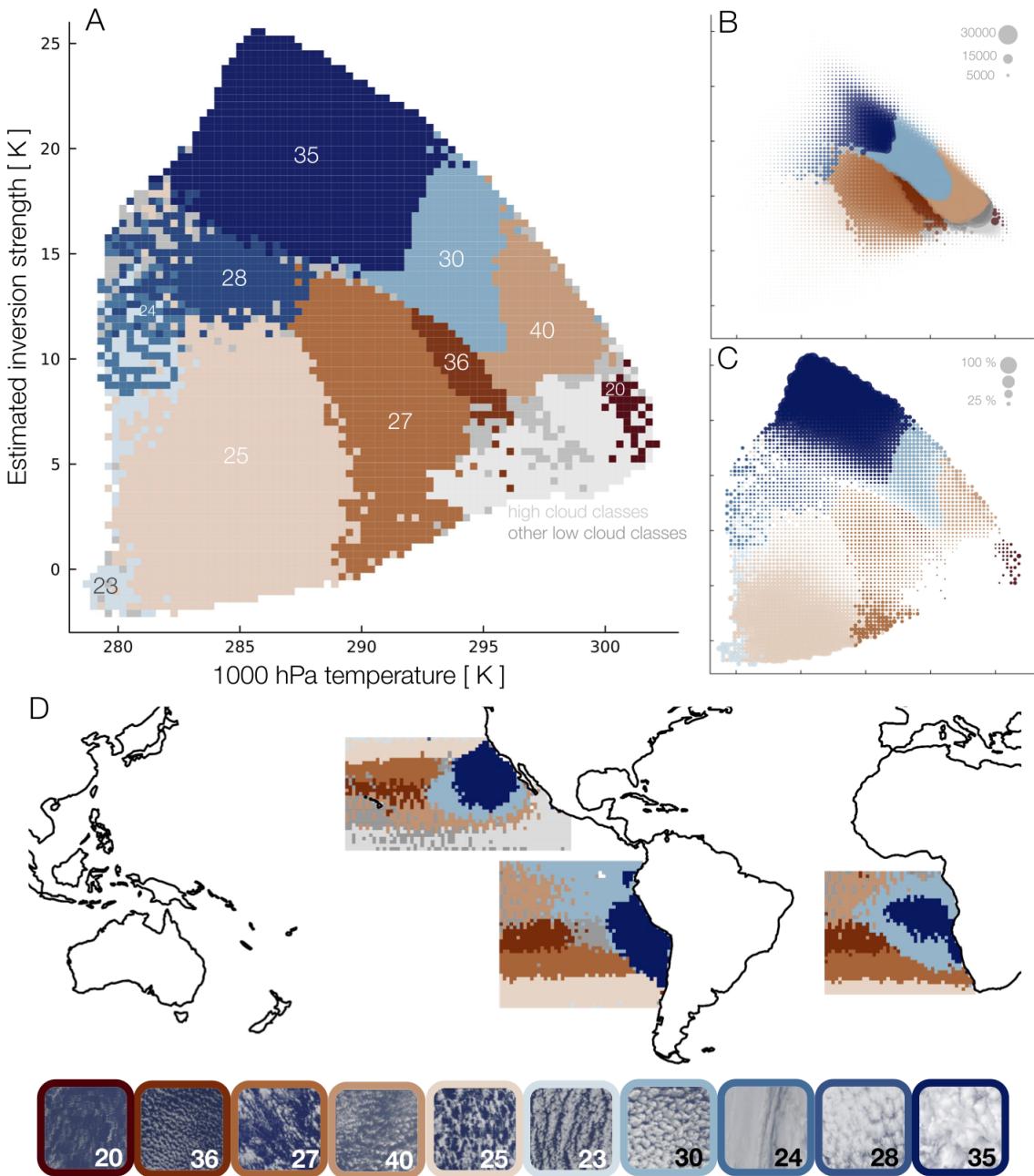


Fig. 2. AICCA cloud classes mapped to a meteorological domain of near-surface temperature and inversion strength, for all patches in the three main subtropical stratocumulus regions. (A) The 10 most dominant classes in meteorological space, color-coded. Thumbnail images of typical class members are shown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

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

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

sparse in texture and then finally breaks up. Cool surface temperatures with low inversion strength are almost never sampled, but when those conditions do occur, they are associated with a well-defined cloud class. The 10 most dominant cloud classes in parameter space are color-coded in Fig. 2. The group includes the two most frequent classes (#30 and #35), and all but one are stratocumulus and are included

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

While each meteorological domain in Fig. 2a is by definition associated with a distinct cloud class, they do not explain all variance. Across all cloud occurrences, the mean “purity” across the 10 domains is 30%, i.e., 30% of cloud occurrences are in fact the dominant class associated with a given meteorological environment. However, nearly half of the “top 10” classes are located in their proper domain. In general, the correlation of cloud texture with environmental conditions is strongest in extreme conditions. Thick stratocumulus decks (class #35, dark blue) tend to occur in very stable conditions, and the purity of the class #35 domain is 90% at the largest inversion strengths (and 10% near the domain edge). Similarly, when low-temperature, low-stability conditions do occur, they produce class #25 with as high as 50% purity.

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

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

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

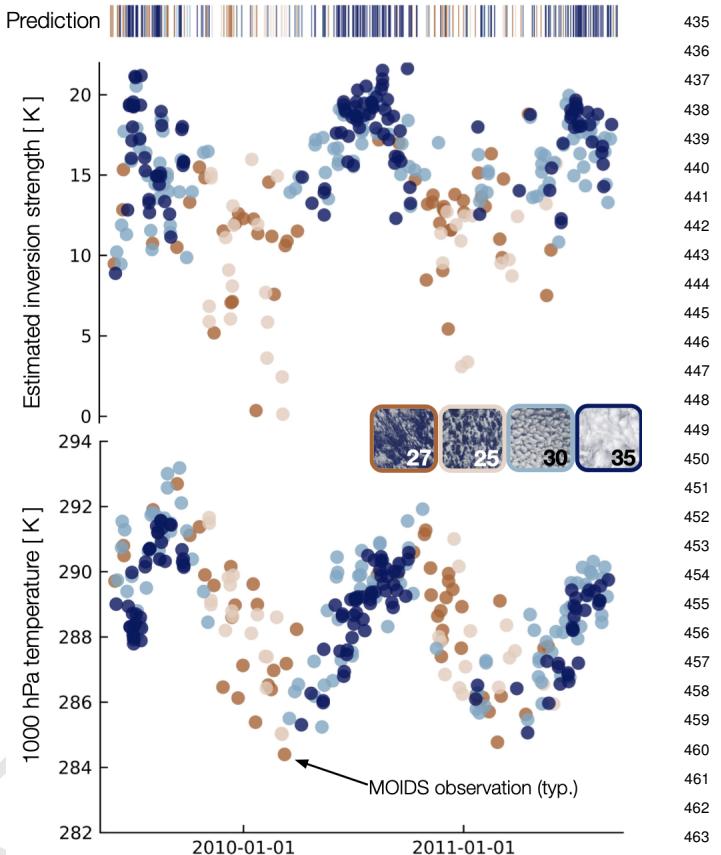


Fig. 3. Timeseries of meteorological variables and observed and predicted cloud classes for a single representative location at 29N, 130W in the North Pacific basin. Top shows inversion strength and bottom near-surface temperature, with observed cloud classes shown as colors for the four main classes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

by precipitation, which occurs only on its northern, upwind edge, while the “hole” expands downwind.

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

We then check this result using a different precipitation estimate. Because the precipitation and cloud class information in Fig. 5a are derived from different satellites, on different orbits, measurements are not coincident and so hourly precipitation at each location is taken as 1/24 of the daily mean, potentially missing short-timescale changes. We therefore

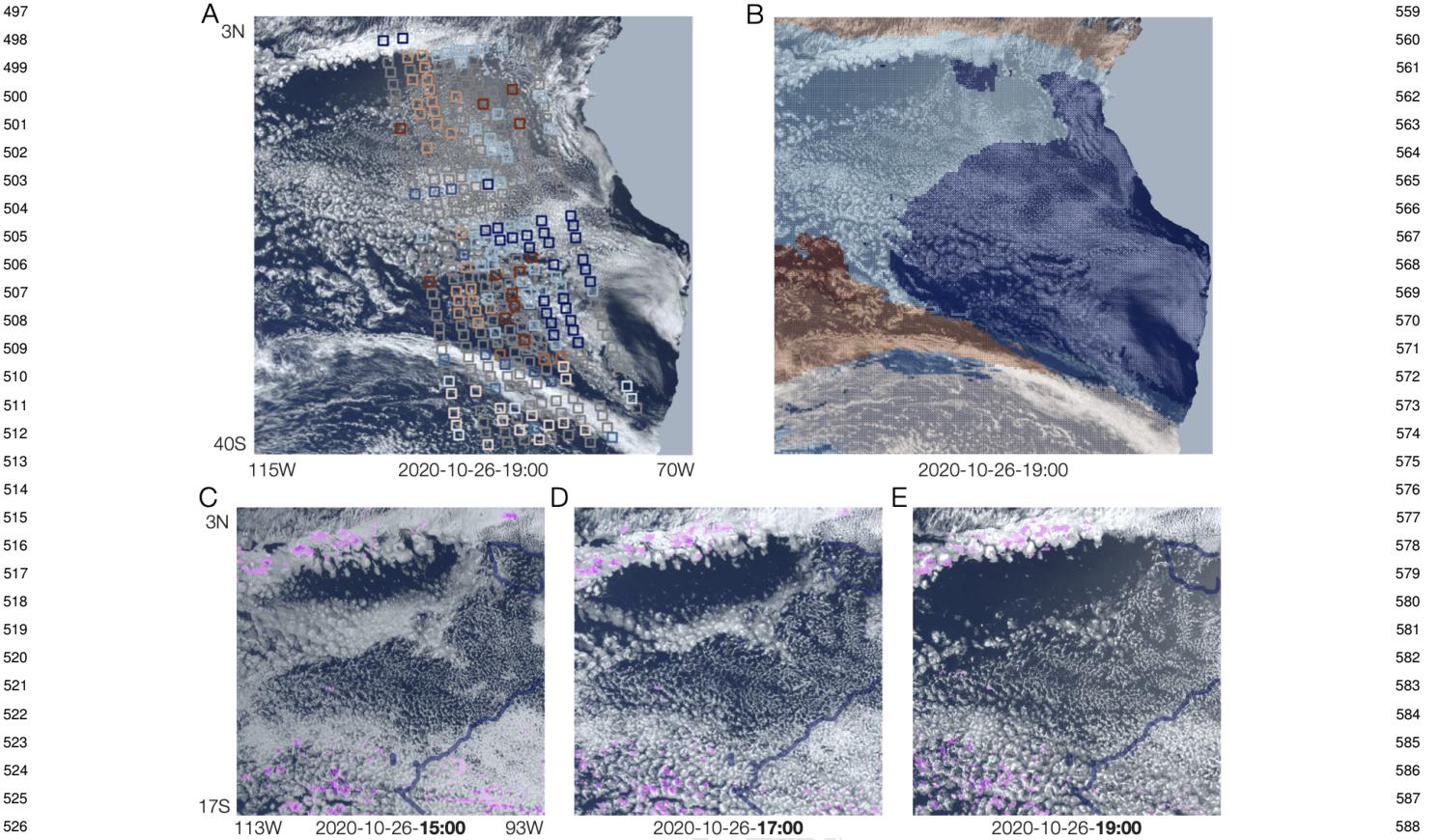


Fig. 4. (A) Visible image from GOES-16 on October 26, 2020 at 19:00 UTC (~3 PM local time) with coincident MODIS-AQUA cloud classes, following the same color code used throughout, and (B) with meteorologically-predicted cloud classes. While the temperature structure predicts large-scale features of the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

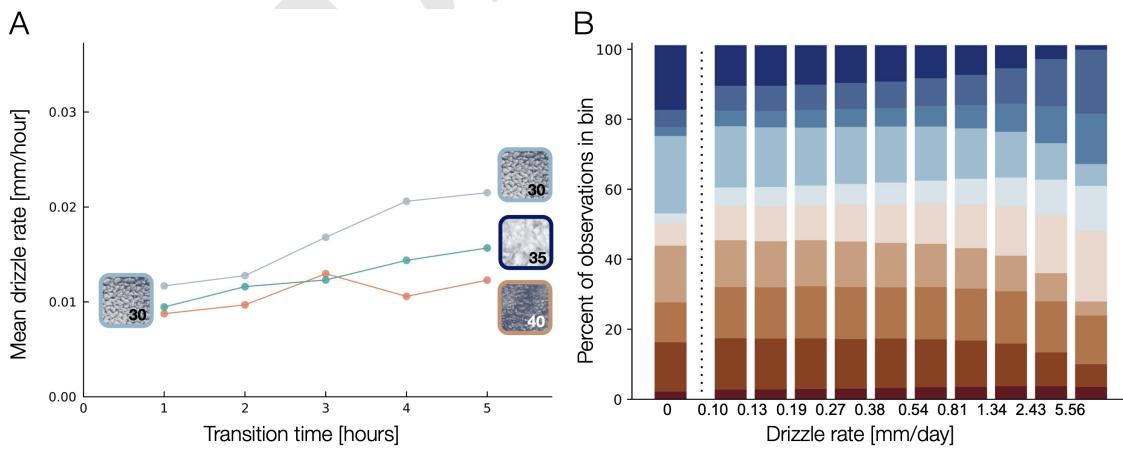


Fig. 5. Population level precipitation impacts. Drizzle does not aid in explaining texture. (A) Short timescale transitions (<6h) from the most populous class # are tracked in the South Pacific basin. 3,000 such transitions are observed via Lagrangian tracking between MODIS TERRA and AQUA passes. Mean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

621 conduct a similar analysis with a more indirect precipitation
622 metric derived from MODIS observations themselves: the
623 “drizzle proxy”, based on estimated droplet concentration
624 number and liquid water path. This proxy was used in early
625 stratocumulus studies (e.g. 28), though more recent work has
626 used IMERG (34). The two estimates differ by a factor of
627 10 (IMERG is drier) and show essentially no correlation
628 (SI Fig. S12). Unsurprisingly, results on stratocumulus
629 transitions using the drizzle proxy are then also different.
630 Median estimated drizzle increases moderately for the static
631 case (0.8 to 1.0 mm day⁻¹ for #30 to #30), increases more
632 substantially for the thickening cases (0.6 to 1.5 mm day⁻¹
633 for #30 to #35), and actually decreases for the closed- to
634 open-cell transitions (1.0 to 0.4 mm day⁻¹ for #30 to open
635 classes). (These values are calculated only for years 2010–
636 2013, reducing the sample size to 650 sequential observations.)
637 While these results could imply a role for precipitation, note
638 that they are opposite from those of prior studies, which
639 found higher precipitation in open cells (27, 28, 30).

640 In general, precipitation is an unlikely candidate to drive
641 observed stratocumulus diversity since most stratocumulus
642 observations are dry. In the dataset over the stratocumulus
643 region used here, clouds are near-ubiquitous but rainfall is
644 scarce: nearly 80% of the 24 M patches are cloudy, but only
645 33% of those cloudy patches exhibit precipitation > 0.1 mm
646 day⁻¹ (in IMERG). Most transitions of stratocumulus texture
647 involve no measurable drizzle. It is therefore unsurprising
648 that precipitation is not particularly predictive for cloud
649 textures (Fig. 5b, which shows the fractional occurrences of
650 the “top 10” stratocumulus classes in 11 precipitation bins
651 (ranging from < 0.1 to > 5.56 mm day⁻¹). Cloud classes
652 are nearly evenly distributed in each bin until the highest
653 rain rates, when some drop sharply (e.g. uniform cloud decks,
654 class #35), but even this change does not produce an overall
655 shift to more open textures.

656 Discussion

657 Self-supervised cloud classification provides a new tool
658 that can aid in understanding the processes that control
659 stratocumulus textures and in turn their radiative impacts.
660 This work demonstrates that AICCA cloud atlas, which
661 divides stratocumulus clouds that would traditionally be
662 placed in a single category into 25 different classes, can be
663 used to generate insight into factors governing their textures.
664 These classes help identify large-scale meteorological drivers
665 that explain at least a portion of stratocumulus diversity, since
666 they map into coherent regions of a 2D variable space of near-
667 surface temperature and inversion strength. These results are
668 consistent with and expand on previous observational work
669 focused on mean cloud fraction (21, 35–37). The remaining
670 unexplained portion of textural diversity is likely driven by
671 small-scale processes. Counter to previous work, results here
672 suggest that precipitation does not play a clear role in driving
673 stratocumulus transitions.

674 Three methodological limitations with our study include
675 precipitation measurements, aerosols, and the Lagrangian
676 scheme. The GPM IMERG microwave-IR product may
677 struggle to measure the low end drizzle found in these low
678 cloud regions. GPM IMERG does report daily accumulated
679 precipitation values below 0.1 mm/day, which is well below
680 the common drizzle-proxy levels reported in previous studies
681

682 (28) and gauge-based evaluations over land have found GPM
683 IMERG actually overestimates the low end (38). Nevertheless,
684 a rigorous evaluation of microwave-IR precipitation measurement
685 in these low cloud regions is needed. Aerosols—clearly
686 critical to cloud processes—are not included in our framework
687 because sufficient large scale aerosol observational data do
688 not exist. Daily aerosol retrievals from passive satellites are
689 only possible in cloud-free sky (39)—which clearly biases
690 our sample in an irreconcilable way. Consequently, most
691 satellite-based cloud research uses data products at the
692 monthly timescale (40, 41) or approximated aerosol products.
693 The short term and small scale variations in cloud textures
694 we find seem unlikely to be driven by differential aerosols.
695 Finally, tracking transitions with MODIS is not ideal. MODIS
696 sampling in this region, further reduced in our sample to
697 daytime-only observations, drives the return period past 24
698 hours in most cases. Many unseen transitions may take
699 place in between the two observations and severely limits
700 our abilities to test shorter term transitions and potential
701 hysteresis effects. A second issue with the transition tracking
702 scheme is the assumption that clouds move perfectly with the
703 prevailing winds. We do observe some apparent transitions
704 ‘up wind’ (see Fig. 4, bottom row).

705 Expected climate change will modify the temperature
706 structure of the lower atmosphere in the subtropics including
707 increasing sea surface temperature and poleward expansion
708 of the Hadley cell (42). EIS is projected to increase moderately
709 (43) under expected climate change but will trade off to
710 some degree with warmer SSTs (37). Observational research
711 carefully coupled with modeled circulation and temperatures
712 changes should be able to inform on future cloud response to
713 warming where model simulated clouds themselves fall short.
714 Future work should employ geostationary satellite images to
715 test the short term dynamics apparent in satellite images of
716 clouds.

717 Future observational work employing deep learning can
718 improve model parameterizations of clouds and therefore
719 constrain CO₂ forcing uncertainty.

720 Materials and Methods

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

730 **Self-Supervised cloud classification.** The AI-Based Cloud Classifica-
731 tion Atlas (AICCA) employs a rotationally-invariant convolutional
732 autoencoder to classify clouds in 2000–2022 MODIS satellite
733 images over the global oceans from both TERRA and AQUA (19)
734 MODIS instruments. The autoencoder is trained without explicit
735 supervision 1 million sample 128×128 pixels patches and consisting
736 of six near-IR and IR radiance bands. Cloud patterns can occur
737 in different orientations. Thus, we develop a rotation-invariant
738 (RI) autoencoder that generates identical latent representations
739 for cloud objects that differ only in their horizontal orientation.
740 Hierarchical agglomerative clustering is then applied to the latent
741 vectors from the autoencoder (See supplemental Fig. S13 and
742 (19, 20) for details) generating an information-maximizing 42 cloud
743 classes. Inference is applied to the remaining ~200M observations
744 with the trained model. MODIS pixels are ~1 km resolution

745 and so the resultant classification is at \sim 1 degree latitude and
746 longitude. Cloud classes are separable, stable, and physically
747 reasonable (20, 44).

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

755 Observed AICCA classes are then mapped against daily mean
756 reanalysis meteorology variables from the ERA5 reanalysis (45).
757 Variables are taken from the hourly product at pressure levels
758 and averaged to daily means. They include temperature, relative
759 humidity, divergence, geopotential, potential vorticity, u, v, w
760 components of wind at different pressure levels between 950 and 700
761 hpa as well as surface quantities including: sea surface temperature,
762 mean sea level pressure, significant wave height, some engineered
763 features including: lower tropospheric stability, estimated inversion
764 strength (EIS), and some calculated wind shears (see supplemental
765 table). AICCA observation patch latitude and longitude is rounded
766 to the nearest 0.25 degree and matched with the ERA-5 value
767 (provided at 0.25 degree spatial resolution) on the same day.
768 Estimated inversion strength is calculated based on the method of
769 (35). We also include daily and hourly precipitation from the GPM
770 IMERG composite satellite product (46). Precipitation values are
771 aggregated to 0.25 degrees (from native 0.1 degrees) to match
772 ERA5 resolution. Finally, MODIS drizzle proxy is calculated from
773 the method in (28) using droplet concentrations from (47) and
774 liquid water path data from (48).

775 Meteorological drivers of low cloud texture are tested in the
776 following fashion. We train a binary logistic regression model with
777 L2 penalty and Limited-memory BFGS solver (in *scikit-learn* (49))
778 for pairs of 5000 randomly-selected samples from each class in the
779 top 15 classes. Every combination of classes is tested across all 68
780 meteorological variables (over 200k combinations in total). Pairs
781

807 of meteorological driver are scored based on their mean prediction
808 skill across all pairs of cloud classes. The process is repeated
809 for combinations of three meteorological variables. Cloud free
810 conditions are ignored in this step.

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

826 **ACKNOWLEDGMENTS.** JAF was supported in part by NSF
827 NRT program (grant DGE-1735359) and by an appointment to the
828 Intelligence Community Postdoctoral Research Fellowship Program
829 at Toyota Technological Institute at Chicago administered by
830 Oak Ridge Institute for Science and Education (ORISE) through
831 an inter agency agreement between the U.S. Department of
832 Energy and the Office of the Director of National Intelligence
833 (ODNI). 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.

- 777 1. GL Stephens, Cloud feedbacks in the climate system: A critical review. *J. Clim.* **18**,
778 237–273 (2005).
- 779 2. M Zhao, et al., Uncertainty in model climate sensitivity traced to representations of cumulus
780 precipitation microphysics. *J. Clim.* **29**, 543–560 (2016).
- 781 3. SC Sherwood, et al., An assessment of Earth’s climate sensitivity using multiple lines of
782 evidence. *Rev. Geophys.* **58** (2020).
- 783 4. MD Zelinka, et al., Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res.
Lett.* **47** (2020).
- 784 5. T Schneider, CM Kaul, KG Pressel, Possible climate transitions from breakup of
785 stratocumulus decks under greenhouse warming. *Nat. Geosci.* **12**, 163–167 (2019).
- 786 6. T Schneider, CM Kaul, KG Pressel, Solar geoengineering may not prevent strong warming
787 from direct effects of CO₂ on stratocumulus cloud cover. *Proc. Natl. Acad. Sci.* **117**,
30179–30185 (2020).
- 788 7. G Bellon, O Geoffroy, Stratocumulus radiative effect, multiple equilibria of the well-mixed
789 boundary layer and transition to shallow convection. *Q. J. Royal Meteorol. Soc.* **142**,
1685–1696 (2016).
- 790 8. P Kuma, FAM Bender, A Schuddeboom, AJ McDonald, Ø Seland, Machine learning of
791 cloud types shows higher climate sensitivity is associated with lower cloud biases.
Atmospheric Chem. Phys. **23**, 523–549 (2023).
- 792 9. L Bock, A Lauer, Cloud properties and their projected changes in CMIP models with
793 low/medium/high climate sensitivity (2023) <https://doi.org/10.5194/egusphere-2023-1086>.
- 794 10. B Stevens, et al., DYAMOND: the DYnamics of the Atmospheric general circulation Modeled
795 On Non-hydrostatic Domains. *Prog. Earth Planet. Sci.* **6**, 61 (2019).
- 796 11. NP Wedi, et al., A baseline for global weather and climate simulations at 1 km resolution. *J.
Adv. Model. Earth Syst.* **12** (2020).
- 797 12. T Schneider, et al., Harnessing AI and computing to advance climate modelling and
798 prediction. *Nat. Clim. Chang.* **13**, 887–889 (2023).
- 799 13. T Schneider, et al., Climate goals and computing the future of clouds. *Nat. Clim. Chang.* **7**,
3–5 (2017).
- 800 14. WB Rossow, RA Schiffer, Advances in understanding clouds from ISCCP. *Bull. Am.
Meteorol. Soc.* **80**, 2261–2287 (1999).
- 801 15. AH Young, KR Knapp, A Inamdar, W Hankins, WB Rossow, The International Satellite
802 Cloud Climatology Project H-Series climate data record product (2018)
<https://www.ncei.noaa.gov/products/international-satellite-cloud-climatology>.
- 803 16. J Zhang, P Liu, F Zhang, Q Song, CloudNet: Ground-Based Cloud Classification With Deep
804 Convolutional Neural Network. *Geophys. Res. Lett.* **45**, 8665–8672 (2018).
- 805 17. V Zantedeschi, et al., Cumulo: A Dataset for Learning Cloud Classes. *arXiv:1911.04227
[physics, stat]* (2020) arXiv: 1911.04227.
- 806 18. T Yuan, et al., Applying deep learning to NASA MODIS data to create a community record of
marine low-cloud mesoscale morphology. *Atmospheric Meas. Tech.* **13**, 6989–6997 (2020).
- 819 19. 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).
- 820 20. T Kurihana, EJ Moyer, IT Foster, AICCA: AI-Driven Cloud Classification Atlas. *Remote.
Sens.* **14**, 5690 (2022).
- 821 21. 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).
- 822 22. S Platnick, et al., The MODIS Cloud Optical and Microphysical Products: Collection 6
Updates and Examples From Terra and Aqua. *IEEE Transactions on Geosci. Remote. Sens.*
55, 502–525 (2017).
- 823 23. D Randall, Conditional stability of the first kind upside-down. *Am. Meteorol. Soc.* **37**,
125–130 (1980).
- 824 24. R Wood, Stratocumulus clouds. *Mon. Weather. Rev.* **140**, 2373–2423 (2012).
- 825 25. EM Agee, TS Chen, KE Dowell, A review of mesoscale cellular convection. *Bull. Am.
Meteorol. Soc.* **54**, 1004–1012 (1973).
- 826 26. B Stevens, et al., Pockets of open cells and drizzle in marine stratocumulus. *Bull. Am.
Meteorol. Soc.* **86**, 51–58 (2005).
- 827 27. TM Sharon, et al., Aerosol and cloud microphysical characteristics of rifts and gradients in
maritime stratocumulus clouds. *J. Atmospheric Sci.* **63**, 983–997 (2006).
- 828 28. R Wood, et al., Open cellular structure in marine stratocumulus sheets. *J. Geophys. Res.*
113, D12207 (2008).
- 829 29. G Feingold, I Koren, H Wang, H Xue, WA Brewer, Precipitation-generated oscillations in
open cellular cloud fields. *Nature* **466**, 849–852 (2010).
- 830 30. 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).
- 831 31. S Klein, D Hartmann, The seasonal cycle of low stratiform clouds. *J. Clim.* **6**, 1586–1606
(1993).
- 832 32. 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).
- 833 33. I Koren, G Feingold, Aerosol-cloud-precipitation system as a predator-prey problem. *Proc.
Natl. Acad. Sci.* **108**, 12227–12232 (2011).
- 834 34. MW Christensen, WK Jones, P Stier, Aerosols enhance cloud lifetime and brightness along
the stratus-to-cumulus transition. *Proc. Natl. Acad. Sci.* **117**, 17591–17598 (2020).
- 835 35. R Wood, CS Bretherton, On the Relationship between Stratiform Low Cloud Cover and
Lower-Tropospheric Stability. *J. Clim.* **19**, 6425–6432 (2006).
- 836 36. TA Myers, JR Norris, On the Relationships between Subtropical Clouds and Meteorology in
Observations and CMIP3 and CMIP5 Models*. *J. Clim.* **28**, 2945–2967 (2015).

869	37. DT McCoy, R Eastman, DL Hartmann, R Wood, The Change in Low Cloud Cover in a Warmed Climate Inferred from AIRS, MODIS, and ERA-Interim. <i>J. Clim.</i> 30 , 3609–3620 (2017).	931
870		932
871	38. Y Xin, et al., Evaluation of IMERG and ERA5 precipitation products over the Mongolian Plateau. <i>Sci. Reports</i> 12 , 21776 (2022).	933
872		934
873	39. RC Levy, et al., The Collection 6 MODIS aerosol products over land and ocean. <i>Atmospheric Meas. Tech.</i> 6 , 2989–3034 (2013).	935
874		936
875	40. DT McCoy, et al., Natural aerosols explain seasonal and spatial patterns of Southern Ocean cloud albedo. <i>Sci. Adv.</i> 1 , e1500157 (2015).	937
876		938
877	41. Y Chen, et al., Machine learning reveals climate forcing from aerosols is dominated by increased cloud cover. <i>Nat. Geosci.</i> 15 , 609–614 (2022).	939
878		939
879	42. WK Lau, KM Kim, Robust hadley circulation changes and increasing global dryness due to co2 warming from cmip5 model projections. <i>Proc. Natl. Acad. Sci.</i> 112 , 3630–3635 (2015).	940
880		941
881	43. X Qu, A Hall, SA Klein, PM Caldwell, The strength of the tropical inversion and its response to climate change in 18 CMIP5 models. <i>Clim. Dyn.</i> 45 , 375–396 (2015).	942
882		943
883	44. T Kurihana, J Franke, I Foster, Z Wang, E Moyer, Insight into cloud processes from unsupervised classification with a rotationally invariant autoencoder. <i>arXiv preprint</i> <i>arXiv:2211.00860</i> (2022).	944
884		945
885	45. Hersbach, et al., The ERA5 global reanalysis. <i>Q. J. Royal Meteorol. Soc.</i> 146 , 1999–2049 (2020).	946
886		946
887	46. 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.	947
888		947
889	47. E Gryspenert, et al., Cloud droplet number concentration, calculated from the MODIS (Moderate resolution imaging spectroradiometer) cloud optical properties retrieval and gridded using different sampling strategies (2022) https://dx.doi.org/10.5285/864a46cc65054008857ee5bb772a2a2b .	948
890		949
891	48. S Khanal, Z Wang, JR French, Improving middle and high latitude cloud liquid water path measurements from MODIS. <i>Atmospheric Res.</i> 243 , 105033 (2020).	950
892		951
893	49. F Pedregosa, et al., Scikit-learn: Machine learning in Python. <i>J. Mach. Learn. Res.</i> 12 , 2825–2830 (2011).	952
894		953
895		954
896		955
897		956
898		957
899		958
900		959
901		960
902		961
903		962
904		963
905		964
906		965
907		966
908		967
909		968
910		969
911		970
912		971
913		972
914		973
915		974
916		975
917		976
918		977
919		978
920		979
921		980
922		981
923		982
924		983
925		984
926		985
927		986
928		987
929		988
930		989
931		990
932		991
933		992