Attribution of Observed Recent Decrease in Low Clouds Over the Northeastern Pacific to Cloud-Controlling Factors

Hendrik Andersen[1,2] ⓘ, Jan Cermak[1,2] ⓘ, Lukas Zipfel[1,2] ⓘ, and Timothy A. Myers[3,4] ⓘ

[1]Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research, Karlsruhe, Germany, [2]Karlsruhe Institute of Technology (KIT), Institute of Photogrammetry and Remote Sensing, Karlsruhe, Germany, [3]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO, USA, [4]Physical Science Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, USA

Abstract Marine low clouds cool the Earth's climate, with their coverage (LCC) being controlled by their environment. Here, an observed significant decrease of LCC in the northeastern Pacific over the past two decades is linked quantitatively to changes in cloud-controlling factors. In a comparison of different statistical and machine learning methods, a decrease in the inversion strength and near-surface winds, and an increase in sea surface temperatures (SSTs) are unanimously shown to be the main causes of the LCC decrease. While the decreased inversion strength leads to more entrainment of dry free-tropospheric air, the increasing SSTs are shown to lead to an increased vertical moisture gradient that enhances evaporation when entrainment takes place. While the LCC trend is likely driven by natural variability, the trend-attribution framework developed here can be used with any method in future analyses. We find the choice of predictors is more important than the method.

Plain Language Summary Marine low clouds efficiently cool the Earth's climate, and their prevalence is controlled by environmental factors. Here, a decrease of the cover of marine low clouds in the northeastern Pacific over the past 20 years is analyzed to attribute the trend to changes in environmental factors known to be important for low clouds. A decrease in the strength of the temperature inversion and an increase in sea surface temperatures (SSTs) are shown to be the main causes of the low-cloud trend. The decreased inversion strength leads to more mixing in of dry air from above the clouds, leading to cloud evaporation. The increasing SSTs increase the atmospheric moisture levels near the surface more than above the cloud, enhancing evaporation when the mixing takes place. While the trend in low clouds is likely driven by natural variability rather than climate change, the analytical framework developed here can be deployed to attribute causes for trends with any statistical or machine learning method in the future. The analysis shows that the choice of environmental factors used for the analysis has a larger impact on the results than the method.

1. Introduction

Marine low clouds cover about 25% of the Earth's surface, and efficiently cool the climate as their albedo effect dominates their warming effect. This is particularly the case for subtropical stratocumulus clouds, as they are horizontally extensive and reflect much more of the incoming solar radiation back to space than the dark ocean underneath (Boucher et al., 2013; Hartmann et al., 1992; Randall et al., 1984). While recent observational studies have provided clear evidence of a positive low-cloud feedback (Ceppi & Nowack, 2021; Cesana & Del Genio, 2021; Klein et al., 2017; Myers & Norris, 2016; Myers et al., 2021; Qu, Hall, Klein, & Caldwell, 2015), low-cloud feedbacks greatly vary across global climate models, limiting the efforts in constraining climate sensitivity, that is, the Earth's warming response to a doubling in atmospheric CO_2 (Bony & Dufresne, 2005; Zelinka et al., 2020).

Stratocumulus clouds frequently occur above the ocean off continental west coasts, as these regions are typically characterized by large-scale subsidence over low sea surface temperatures (SSTs), creating a shallow marine boundary layer (MBL) that is capped by a temperature inversion (Norris & Klein, 2000; Wood & Bretherton, 2006). The variability of these large-scale meteorological conditions (or cloud-controlling factors, CCFs) is therefore linked to the marine low cloud cover (LCC). Two primary CCFs are the (estimated) inversion strength (EIS), which controls the entrainment of dry free tropospheric air and therefore the humidity and depth of the MBL (Wood & Bretherton, 2006), and the SST (Myers & Norris, 2016; McCoy et al., 2017), which may be

Writing – review & editing: Hendrik Andersen, Jan Cermak, Lukas Zipfel, Timothy A. Myers

understood as a surrogate for multiple processes that drive LCC (Qu, Hall, Klein, & Deangelis, 2015). One such process may be that higher SSTs increase the latent heat flux at the surface, enhancing the buoyancy and deepening the MBL, which increases the entrainment of dry free tropospheric air, ultimately depleting LCC (Rieck et al., 2012; Qu, Hall, Klein, & Caldwell, 2015). A second process is that an SST increase leads to a stronger moisture increase in the MBL than in the free troposphere (FT), following the Clausius-Clapeyron equation if the relative humidity in the MBL stays fairly constant. This leads to a larger moisture difference between the MBL and the FT, increasing the evaporation-efficiency of entrained FT air (Qu, Hall, Klein, & Caldwell, 2015; van der Dussen et al., 2015). Additional CCFs are, for example, near surface winds and temperature advection (T_{adv}; Klein, 1997; Scott et al., 2020).

Satellite observations are frequently used to estimate the sensitivities of LCC to changes in CCFs in an effort to provide observational estimates for low-cloud feedbacks (Cesana & Del Genio, 2021; Klein et al., 2017; McCoy et al., 2017; Myers & Norris, 2016; Myers et al., 2021). However, most long-term satellite data sets need to be corrected, for example, for temporal drifts, in order to conduct long-term trend analyses (Norris et al., 2015, 2016; Seethala et al., 2015). As observational records of satellites get longer, trend observations of clouds and their interpretation with respect to changes in CCFs will become even more useful to establish robust estimates on cloud responses to climate change. In this context, methodological innovation in how these sensitivities are quantified, which is traditionally done in region-specific multiple linear regression frameworks, may facilitate the relevant research (Ceppi & Nowack, 2021). As machine learning methods gain popularity in environmental sciences, these methods have also shown good performance when applied for the quantification of cloud sensitivities in recent studies (Andersen et al., 2017; Dadashazar et al., 2020, 2021; Fuchs et al., 2018; Pauli et al., 2020). Nonetheless, their potentials and limitations for CCF analyses have not been fully explored yet.

Here, satellite observations from the past two decades are used to investigate the recent decrease of LCC in the northeast (NE) Pacific (see Figure 1), the most pronounced LCC trend globally during this time period (see Figure S1 in Supporting Information S1). Satellite observations and reanalysis data are used to analyze, which CCFs drive this regional decrease in LCC. In light of the finding that in this region a marine heatwave in 2015 has been linked to a record low in LCC (Myers et al., 2018), and the hypothesis of this study is that an associated increase in SSTs is a main driver of the regional LCC decrease. To test this, various statistical and machine learning models are trained to model LCC on the basis of two different sets of CCFs to (a) evaluate the effect of SST and other CCFs on the LCC trend, and (b) understand which of the SST-related processes outlined above is most important in this context.

2. Data and Methods

2.1. Data

Satellite observations from the polar-orbiting platforms Terra and Aqua are used. LCC is defined as the liquid-water cloud fraction from gridded monthly data from the level 3 MOD08_M3 and MYD08_M3 products of the Moderate Resolution Imaging Spectroradiometer (MODIS; Platnick et al., 2015). While the main part of the analysis is done using Terra observations (the observational period is ∼2 years longer), they are complemented by Aqua data, as the difference in observation times (∼10:30 am and ∼1:30 pm local equatorial crossing times for Terra and Aqua, respectively) allows to consider diurnal changes in LCC. Data from the newest collection (6.1) are used. To exclude influences from higher-level clouds, LCC is considering single layer cloud situations only (Cloud_Retrieval_Fraction_1L_Liquid_FMean). In the study domain, the depth of the MBL is typically below 1.5 km (Lin et al., 2009; Wood & Bretherton, 2004), so this single layer liquid water cloud product is expected to well approximate LCC, with the limitation of missing LCC in multi-layer cloud situations. Short-wave cloud radiative effects at the top of the atmosphere (CRE_{SW}) are analyzed with the gridded monthly Energy Balanced and Filled (EBAF) level 3b products in the newest edition (4.1) from the Clouds and the Earth's Radiant Energy System (CERES), also mounted on Terra (Loeb et al., 2018). All satellite data sets have a spatial resolution of $1° \times 1°$, and data for all complete years of Terra and Aqua observations (2001–2020 and 2003–2020, respectively) are used. One should note that at this spatial scale, clouds may not yet be in equilibrium with their environment (Klein et al., 1995; Mauger & Norris, 2010), which may pose a limitation to the predictive skill of the models used.
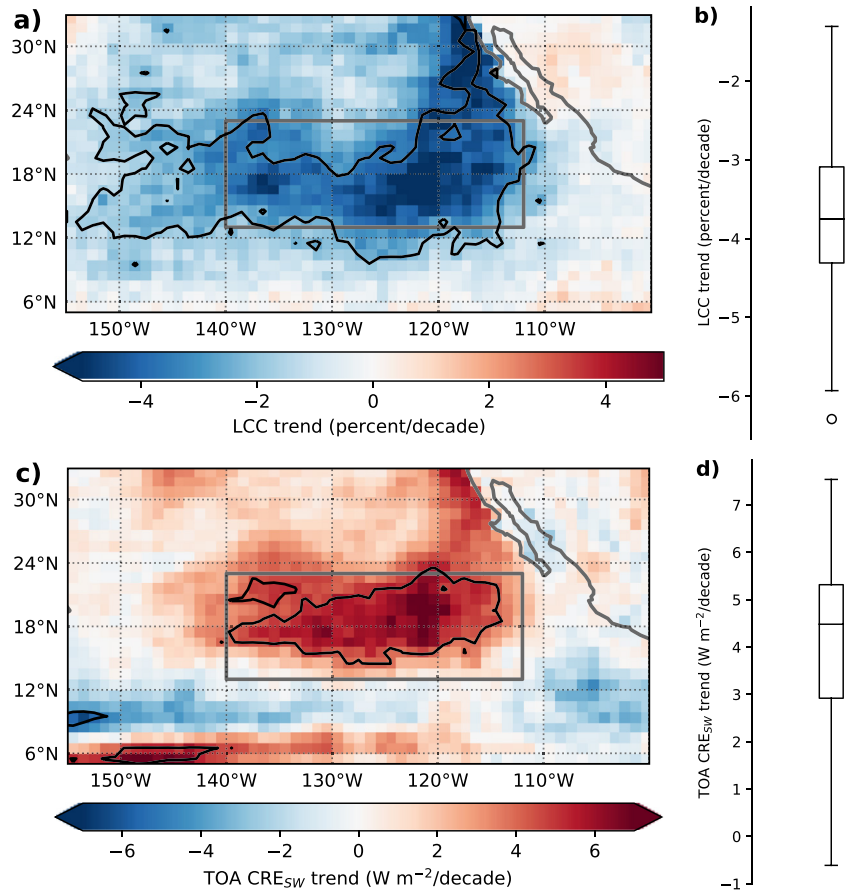
Figure 1. Trend patterns of low cloud cover (LCC) and TOA CRE$_{SW}$ between 2001 and 2020 in the MODIS and CERES data. The black contours mark the regions where the trends in LCC and CRE$_{SW}$ are highly significant (p value < 0.01). The distribution of trends in the study area (gray box) is shown in panels (b) and (d). The median is represented by the horizontal line, framed by the interquartile range (boxes), with the whiskers expanding the boxes by up to 1.5 interquartile ranges.

For information on CCFs, monthly data from ERA5, the most recent reanalysis from the European Center for Medium-Range Weather Forecasts (ECMWF), are used for the observational time period (Hersbach et al., 2019a, 2019b). The ERA5 data come at a native resolution of $0.25° × 0.25°$ but is aggregated to fit the MODIS grid. The CCFs used in this study are: SST, EIS (Wood & Bretherton, 2006), horizontal temperature advection at the surface (Scott et al., 2020, T_{adv}), wind speed at 10 m (WS10), mean surface latent heat flux (MSLHF), relative humidity of the FT (RHFT, defined here as the average relative humidity between 300 and 700 hPa), vertical pressure velocity at 700 hPa (ω700), and the moisture contrast between the boundary layer and FT (Q_{diff}), which is defined as the difference between specific humidity at 1,000 and 700 hPa (Qu, Hall, Klein, & Deangelis, 2015). The authors acknowledge that uncertainties exist for all data sets used, and that their magnitude is likely to vary between the different products and reanalysis parameters. Surface flux products from reanalyses are, for example, expected to contain much larger uncertainties than better observed quantities like the SST (Bentamy et al., 2017; Luo & Minnett, 2020; Martens et al., 2020; Pokhrel et al., 2020). However, all data sets are expected to capture the general patterns and trends adequately for this study.

2.2. Methods

Two different predictor sets are used to model LCC that have been inspired by Myers and Norris (2016) and (Qu, Hall, Klein, & Deangelis, 2015) such as (a) a 6-predictor setup (6-CCF) and (b) a 7-predictor setup (7-CCF). The 6-CCF model uses SST, EIS, WS10, T_{adv}, RHFT, and ω700 to model LCC, while 7-CCF uses the same predictors, but replaces the SST with Q_{diff} and MSLHF to separate the different SST-related processes outlined above.

To compare the traditionally used ordinary least squares regression (OLS) (Cesana & Del Genio, 2021; Klein et al., 2017; McCoy et al., 2017; Myers & Norris, 2016; Myers et al., 2021; Scott et al., 2020) and machine learning techniques that have gained relevance in the field lately, artificial neural networks (ANNs; Andersen et al., 2017) and extreme gradient boosting (XGB; Chen & Guestrin, 2016) are used. XGB is a gradient tree boosting method similar to the popular gradient boosting regression trees (GBRTs) that have been used in many aerosol and cloud related studies recently (Andersen et al., 2021; Dadashazar et al., 2020, 2021; Fuchs et al., 2018; Stirnberg et al., 2020, 2021), with the advantages of a built-in regularization techniques and much shorter run times (Chen & Guestrin, 2016).

For the training of the statistical and machine learning models, the seasonality (region-specific monthly averages) and the linear trend are subtracted from all data sets (TrendOFF). A second version of the data set is created where the trends are not subtracted (TrendON). Both CCF-anomaly data sets are then transformed to z-scores by removing the mean and scaling to unit variance as in Scott et al. (2020), which is necessary for the comparison of the different models and is a standard procedure for the ANNs (Olden & Jackson, 2002). All data are randomly split into separate training (67% of the sample) and test (33% of the sample) data sets. While random sampling may not be ideal in cases where data are temporally autocorrelated as training, and validation data are not fully independent in this case (Roberts et al., 2017); a separation into training and validation time periods is not reasonable here, as independent validation data is needed over the entire time period to analyze the LCC trend. For the two machine learning approaches, hyperparameters of the individual models are tuned during the training phase in a 10-fold cross validation setup (Refaeilzadeh et al., 2009; Stone, 1974) using random grid search. The details of the hyperparameters tested and chosen are provided by Tables S1 and S2 in Supporting Information S1. After the model hyperparameters are set, all models are again trained 10 times on random training and test splits (same split proportions as before) to test the variability in model performance and the quantified contributions to the LCC trend. The contribution of each CCF to the LCC trend is quantified as follows:

$$\Delta LCCtrend_i = \beta_{TrendON_i} - \beta_{TrendOFF}$$

where $\beta_{TrendOFF}$ specifies the model predicted trend based on detrended data of all CCFs and $\beta_{TrendON_i}$ specifies the model predicted trend based on non-detrended data of the ith CCF and detrended data of all other CCFs.

3. Results and Discussion

Figure 1 shows the linear trend of LCC and CRE_{sw} in the NE Pacific over the last two decades (2001–2020) from the MODIS and CERES data. A coherent region of a highly significant LCC decrease is apparent (a), which is associated with a significant decrease of the cloud cooling effect on climate as quantified by the CRE_{sw} trend found in CERES EBAF data (c). Overall, the average LCC trend of all pixels of the study area (black box) is 3.7% decade^{-1} with a standard deviation of 0.9% decade^{-1}, and the associated decreased cooling is on the order of 4.1 Wm^{-2} (median 4.5 Wm^{-2}, standard deviation 1.8 Wm^{-2}). The LCC decrease is even stronger in the Aqua MODIS data (4.5% decade, with a standard deviation of 1.6% decade^{-1}; Figures S2a and S2b in Supporting Information S1). However, this is caused by the different time periods sampled (2001 and 2002 are not observed by Aqua), and the difference in average LCC at the overpass time of both satellites (Figure S3 in Supporting Information S1). The relative trends in the Terra and Aqua MODIS data are similar though when only the years with mutual observations (2003–2020) are considered (Figures S2e and S2f in Supporting Information S1), suggesting that no systematic changes in the diurnal cycle of LCC are apparent. The decrease in LCC is a trend reversal from the LCC increase in the NE Pacific between 1983/1984 and 2009 found in two recent studies (Seethala et al., 2015; Zhong et al., 2021), which was reported to be driven mainly by increasing inversion strength (Seethala et al., 2015).

Figure 2 shows the regional trends between 2001 and 2020 of the eight CCFs considered in this study. The entire study area in the NE Pacific (black box) features a significant increase of SSTs (a), which exceeds 1K decade^{-1} close to the coast of Baja California. EIS (b) is found to decrease over the entire study area. However, this trend is only highly significant (0.01 level) in the northeastern parts of the study area. Q_{diff} (c) features a significant increase over the last two decades, which is thought to be mainly driven by the increase in SSTs as described in the introduction, as relative humidity at 1,000 hPa has been fairly constant during this time (Figure S4 in Supporting Information S1). This is supported by the observation that the specific moisture increase in the MBL
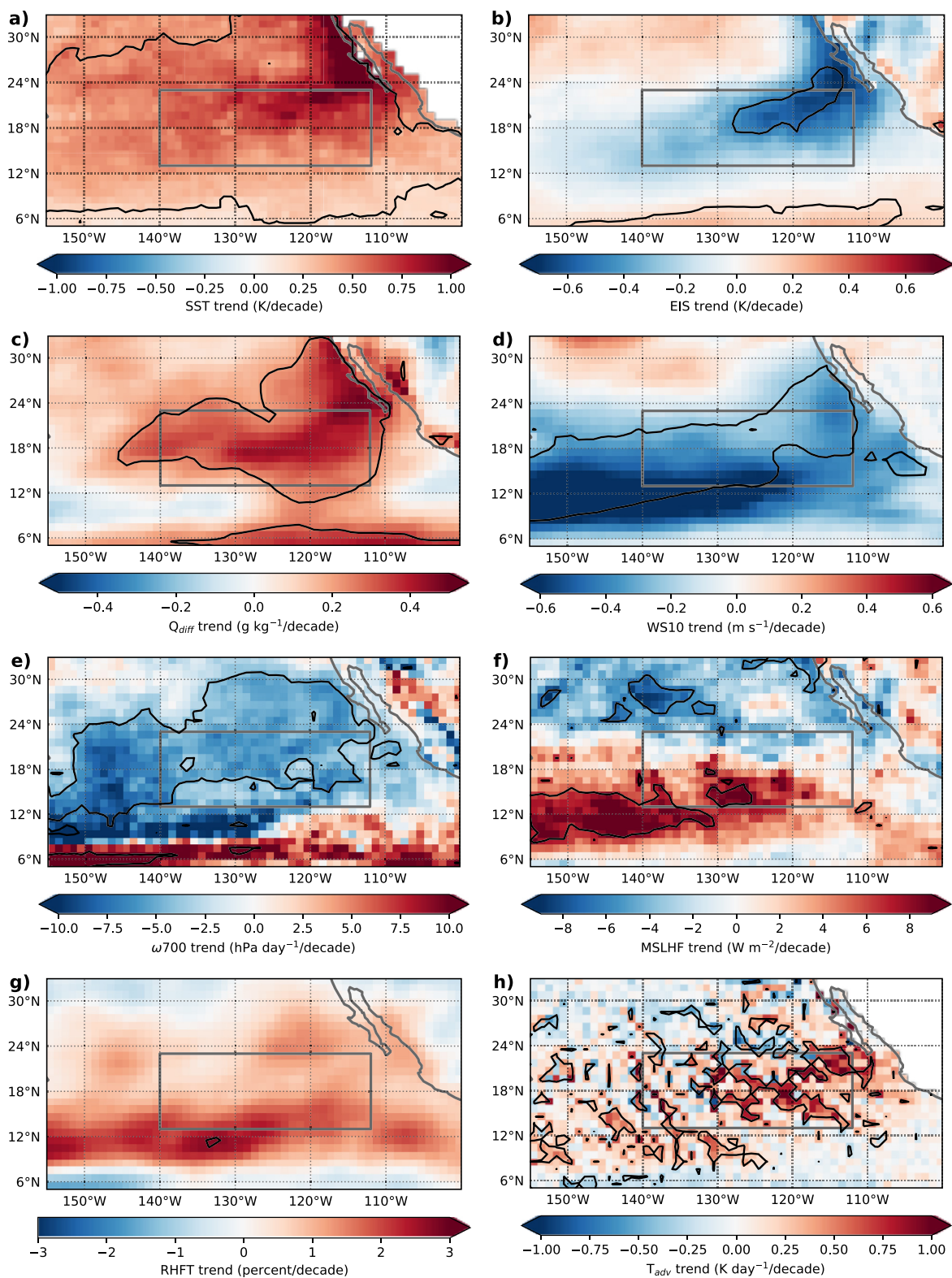
Figure 2. Trend patterns of the analyzed cloud-controlling factors (CCFs) in the ERA5 data between 2001 and 2020, contours as in Figure 1.
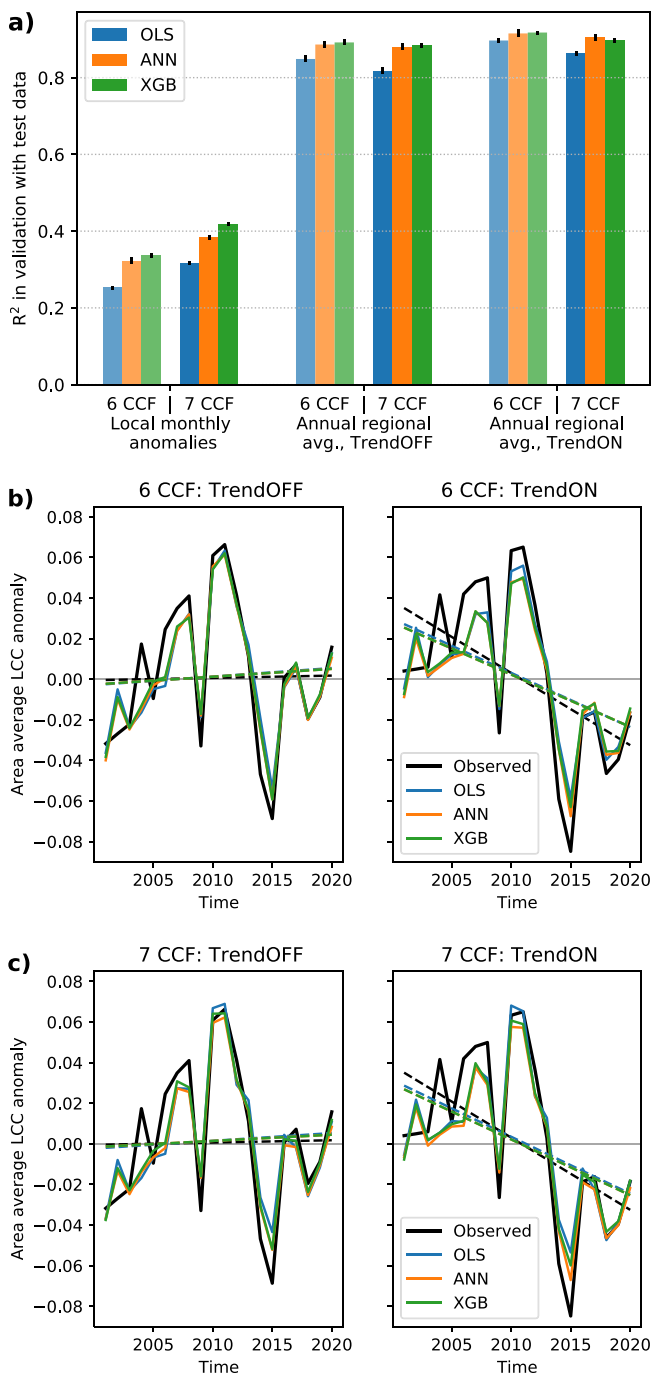
Figure 3. (a) Summary of the model skill (R^2) in predicting LCC in the unseen test data. The lighter-colored bars on the left of each group show the average skill of the 6-CCF models, the darker-colored bars on the right show the average skill of the 7-CCF models. The error bars denote one standard deviation of the model skills. Panels (b) and (c) show the observed (black lines) and predicted (colored lines) yearly area average LCC anomaly (test data) with TrendOFF (left) and TrendON (right) for 6-CCF and 7-CCF, respectively. The dashed lines in (b) and (c) illustrate the observed and average predicted linear trend.

dominates the increase in the FT (Figure S5 in Supporting Information S1). Near-surface surface wind speeds (d) and $\omega 700$ (e) also feature clear, significant trend patterns, with a clear decrease in wind speed and large-scale subsidence. While there is a coherent pattern of increasing RHFT in the NE Pacific, it is not significant at the 0.01 level in the study area. MSLHF and T_{adv} do not show a comparably clear and coherent trend pattern during the considered time period, making it unlikely that they are main drivers of the observed LCC decrease in the NE Pacific. For quantitative estimates on the contributions of each CCF to the LCC trend, statistical and machine learning frameworks are explored in the following.

Figure 3 summarizes the skill of the 6-CCF and 7-CCF setups for the different models (OLS, ANN, XGB). In Figure 3a, it is apparent that for the prediction of the pixel-specific (local) monthly anomalies, the 7-CCF models perform better than the 6-CCF models, with the two machine learning techniques clearly outperforming the OLS. This is notable, as the relationships between LCC and its CCFs are assumed to be linear at these scales (Klein et al., 2017; Klein & Hartmann, 1993; Wood & Bretherton, 2006; Seethala et al., 2015) and may indicate that the capacity for representing more complex nonlinear relationships helps to explain LCC variability. To test to what extent this finding is robust or potentially affected by the way the training and test data sets are sampled, the models are additionally trained with training and test data sets separated into different continuous time periods (models trained 10 times using 18 years each, withholding two consecutive years for testing). In this case, the validation shows that the ANN and XGB models actually produce a slightly but insignificantly higher validation error than the OLS (Figure S6 in Supporting Information S1), showing that the way by which training and test data are separated matters for model validation skill. In fact, the variability of the error produced by the models is much larger when they are trained and tested in different periods, meaning that model skills vary for the different test periods. As the validation skill of the machine learning models seem to benefit from the random sampling (and hence, the fact that training and test data are not fully independent), this indicates that the skill improvement stems from their capacity to exploit the autocorrelation of the CCFs to predict LCC rather than an improved generalization of their relationships. This means that in this stratocumulus-dominated regime and at the scales considered here, nonlinear machine learning models do not better generalize CCF–LCC relationships than an OLS, supporting existing literature (Klein et al., 2017). As such, the most promising methodical way forward for CCF analyses may be the exploitation of spatial information of CCFs as done in Ceppi and Nowack (2021), who used large spatial domains of CCFs in a regularized (ridge) regression framework to predict LCC, therefore capturing non-local CCF effects on LCC (e.g., along the cloud trajectories).

All models/setups have a high skill in predicting regional annual LCC anomalies in the detrended data (Figure 3a, center), and also when the original trends are reintroduced to the CCFs (TrendON for all CCFs, Figure 3a, right). Figures 3b and 3c depict observed and predicted regional annual average LCC anomalies, showing that all models are capable of reproducing the interannual variability well. All models are able to reproduce between 72% and 79% of the observed trend, with the 7-CCF models ~4% closer to the observed trend. In these panels, the clear impact of the LCC anomaly in 2015 on the LCC trend is visible, which is associated with the marine heatwave during this time (Myers et al., 2018). Considering this and the relatively short
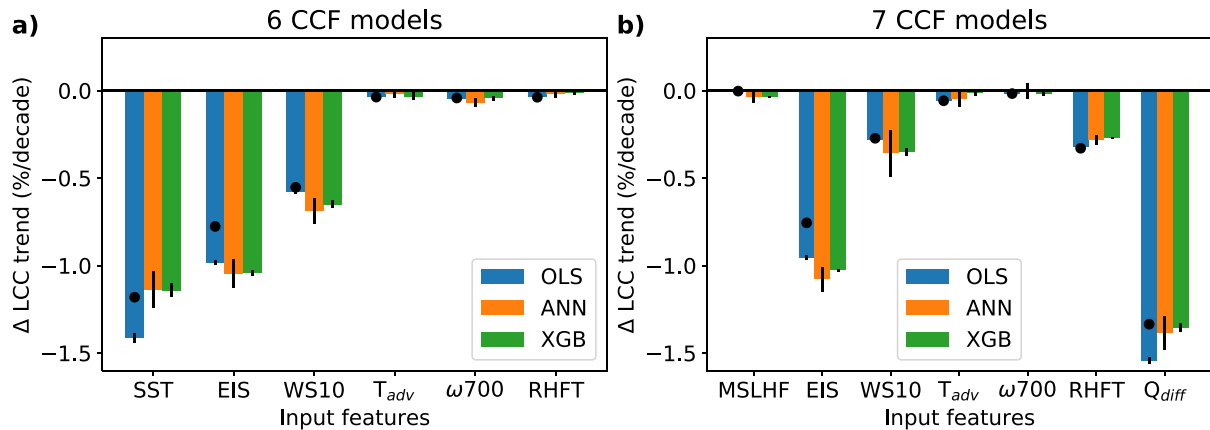
Figure 4. Change in model-predicted northeast (NE) Pacific low cloud cover (LCC) trend (test data) when the trend in a single, specific CCF is introduced for 6-CCF (a), and 7-CCF (b). The black dots are the trend contribution of each CCF calculated directly with the OLS coefficients as in (Myers et al., 2021). The error bars show the standard deviation of the estimated contributions. Error bars are not given for the OLS coefficient estimates for visual clarity, but their lengths are similar to those of the OLS contributions (blue bars).

time period of two decades, the analyzed LCC trend pattern in the NE Pacific is likely to be mainly driven by natural variability rather than climate change.

Figure 4 shows the impact of the different CCFs on the model-predicted LCC trend for the 6-CCF models (a) and the 7-CCF models (b), quantified by Δ LCC trend (bars), and by the OLS coefficients (black dots). The standard deviation in estimated Δ LCC trend for the 10 different model instances is shown as error bars. In the 6-CCF setup (a), all models agree that SST, EIS, and WS10 are the CCFs that drive the LCC decrease in the NE Pacific with their contributions decreasing in that order. A minor difference is that in the OLS approach, the SST contribution is larger than the EIS contribution, while in the ANN and XGB models, their contributions are estimated to be approximately equal. In the 7-CCF setup (b), the EIS decrease is again a major driver of the LCC trend. However, all models agree that the changes in Q_{diff} (driven by the SST) are the most important driver of the LCC decrease.

One should note that the individual estimated CCF contributions to the LCC trend depend on other CCFs included in the statistical/machine learning framework. Here, it is most notable in the case for WS10 and RHFT, where the contribution of WS10 in the 7-CCF setup is estimated to be about half of its contribution in the 6-CCF setup. This is interesting to note, as WS10 and SST are known to be correlated and thereby their correlation may potentially modulate their derived sensitivities (Qu, Hall, Klein, & Deangelis, 2015) and thus the estimated Δ LCC trend contributions. RHFT has no clear Δ LCC trend contribution in the 6-CCF setup but features a marked negative Δ LCC trend contribution in the 7-CCF setup in all models that is comparable to that of WS10. The reasons for the differences in the estimated RHFT contribution to the predicted Δ LCC trend could potentially be explained by RHFT being a proxy for compensating effects in the 6-CCF setup: an increasing RHFT may reduce entrainment drying thereby increasing LCC (Myers & Norris, 2016; van der Dussen et al., 2015). On the other hand, an increasing RHFT may also increase downward longwave radiation (van der Dussen et al., 2015), decreasing cloud-top radiative cooling and thereby limiting the moistening of the boundary layer, ultimately decreasing LCC (Bretherton et al., 2013; Christensen et al., 2013). In the 7-CCF setup, however, entrainment drying is already accounted for by Q_{diff} so that the estimated Δ LCC trend contribution of RHFT in the 7-CCF setup may potentially be associated with the influences of changes in downward longwave radiation. For both CCF setups, there is good agreement between the Δ LCC trend and the estimations of trend contributions using OLS coefficients, with the latter suggesting slightly weaker trend contributions of SST, EIS, and Q_{diff}. While the overall uniform picture across the different models suggests that model choice is much less important than the specific CCF setup chosen for CCF-LCC studies, the spread between the Δ LCC trend quantifications of the different ANNs is comparatively high, while their skills show little variation (Figure 3a). This suggests that ANNs can find multiple solutions of representing the LCC system (i.e., its sensitivities) that work similarly well, and should thus be used with caution in CCF analyses.

4. Conclusions

In this study, the significant LCC decrease in the NE Pacific over the past two decades is investigated with respect to changes in CCFs. It was hypothesized that increasing SSTs are a main driver of the LCC decrease. To analyze this, a multiple linear regression and two machine learning methods were trained to predict LCC anomalies on the basis of CCFs with detrended and deseasonalized data sets. Two predictor sets were used to analyze and attribute the LCC trend to CCFs, namely a 6-CCF setup that includes SST, EIS, and four other CCFs and a 7-CCF setup where SST is replaced with the SST-related process proxies Q_{diff} and MSLHF.

The main findings of this analysis are

1. The 7-CCF models perform slightly better than the 6-CCF models in predicting local monthly anomalies, but both setups are able to accurately reproduce the observed LCC trend
2. The LCC trend is found to be driven by an increase in SSTs and a decrease in EIS and WS10. While the EIS decrease leads to an increase in dry air entrainment, the SST increase causes an increase in the vertical moisture gradient (here: Q_{diff}), leading to increased evaporation and LCC dissipation when the relatively drier free tropospheric air is mixed into the MBL
3. Differences exist between the trend contributions of WS10 and RHFT in the 6-CCF and 7-CCF framework, with the negative WS10 trend contribution stronger when the SST is known, and the negative RHFT trend contribution only present when Q_{diff} is known, highlighting that sensitivities and attributed trends in CCF analyses can be co-dependent
4. In the comparison of the traditional multiple linear regression (OLS) and two machine learning techniques (ANN, XGB), a good agreement across all models is found in terms of CCF contributions to the observed LCC trend. While the ANN and XGB models feature a slightly higher skill in predicting LCC anomalies, this seems to be caused by exploiting the autocorrelation of the CCFs rather than a better generalization of CCF-LCC relationships. The good agreement across the different models suggests that model choice is less important than the CCFs used
5. The trend-attribution framework used here agrees well with traditional OLS-coefficient estimates and can be applied to all statistical/machine learning frameworks in the future

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The ERA5 meteorological reanalysis data (https://doi.org/10.24381/cds.f17050d7; https://doi.org/10.24381/cds.6860a573) are freely available at the Copernicus Climate Change Service (C3S) Climate Date Store: https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset (last access: 26 March 2021). MODIS data (https://dx.doi.org/10.5067/MODIS/MOD08_M3.061 and http://dx.doi.org/10.5067/MODIS/MYD08_M3.061) were downloaded (last access: 11 November 2021) in the Level-1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC) https://ladsweb.modaps.eosdis.nasa.gov/search/. CERES data (https://doi.org/10.5067/TERRA-AQUA/CERES/EBAF-TOA_L3B004.1) are freely available and were obtained from the NASA Langley Research Center CERES ordering tool at https://ceres.larc.nasa.gov/ (last access: 17 May 2021). Code to reproduce the results is available on Zenodo (https://doi.org/10.5281/zenodo.5747221).

References

Andersen, H., Cermak, J., Fuchs, J., Knutti, R., & Lohmann, U. (2017). Understanding the drivers of marine liquid-water cloud occurrence and properties with global observations using neural networks. *Atmospheric Chemistry and Physics*, *17*(15), 9535–9546. https://doi.org/10.5194/acp-17-9535-2017

Andersen, H., Cermak, J., Stirnberg, R., Fuchs, J., Kim, M., & Pauli, E. (2021). Assessment of COVID-19 effects on satellite-observed aerosol loading over China with machine learning. *Tellus B: Chemical and Physical Meteorology*, *73*(1), 1–14. https://doi.org/10.1080/16000889.2021.1971925

Bentamy, A., Piollé, J. F., Grouazel, A., Danielson, R., Gulev, S., Paul, F., et al. (2017). Review and assessment of latent and sensible heat flux accuracy over the global oceans. *Remote Sensing of Environment*, *201*, 196–218. https://doi.org/10.1016/j.rse.2017.08.016

Bony, S., & Dufresne, J. L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, *32*(20), 1–4. https://doi.org/10.1029/2005GL023851

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., et al. (2013). Clouds and aerosols. In T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), *Climate change 2013: The physical science basis, contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.

Bretherton, C. S., Blossey, P. N., & Jones, C. R. (2013). Mechanisms of marine low cloud sensitivity to idealized climate perturbations: A single-LES exploration extending the CGILS cases. *Journal of Advances in Modeling Earth Systems*, *5*(2), 316–337. https://doi.org/10.1002/jame.20019

Ceppi, P., & Nowack, P. (2021). Observational evidence that cloud feedback amplifies global warming. *Proceedings of the National Academy of Sciences*, *118*(30). https://doi.org/10.1073/pnas.2026290118

Cesana, G. V., & Del Genio, A. D. (2021). Observational constraint on cloud feedbacks suggests moderate climate sensitivity. *Nature Climate Change*, *11*, 213–220. https://doi.org/10.1038/s41558-020-00970-y

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *arxiv*, 785–794. https://doi.org/10.1145/2939672.2939785

Christensen, M. W., Carrió, G. G., Stephens, G. L., & Cotton, W. R. (2013). Radiative impacts of free-tropospheric clouds on the properties of marine stratocumulus. *Journal of the Atmospheric Sciences*, *70*(10), 3102–3118. https://doi.org/10.1175/JAS-D-12-0287.1

Dadashazar, H., Crosbie, E., Majdi, M. S., Panahi, M., Moghaddam, M. A., Behrangi, A., et al. (2020). Stratocumulus cloud clearings: Statistics from satellites, reanalysis models, and airborne measurements. *Atmospheric Chemistry and Physics*, *20*(8), 4637–4665. https://doi.org/10.5194/acp-20-4637-2020

Dadashazar, H., Painemal, D., Alipanah, M., Brunke, M., Seethala, C., Corral, A. F., et al. (2021). Cloud drop number concentrations over the western North Atlantic ocean: Seasonal cycle, aerosol interrelationships, and other influential factors. *Atmospheric Chemistry and Physics, Discussions*, 1–39. https://doi.org/10.5194/acp-2021-153

Fuchs, J., Cermak, J., & Andersen, H. (2018). Building a cloud in the Southeast Atlantic: Understanding low-cloud controls based on satellite observations with machine learning. *Atmospheric Chemistry and Physics*, *18*, 16537–16552. https://doi.org/10.5194/acp-18-16537-2018

Hartmann, D. L., Ockert-Bell, M. E., & Michelsen, M. L. (1992). The effect of cloud type on Earth's energy balance: Global analysis. *Journal of Climate*, *5*, 1281–1304. https://doi.org/10.1175/1520-0442(1992)005<1281:teocto>2.0.co;2

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2019a). *Era5 monthly averaged data on pressure levels from 1979 to present, Copernicus Climate Change Service (C3S) climate data store (CDS)*. https://doi.org/10.24381/cds.6860a573 .

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2019b). *Era5 monthly averaged data on single levels from 1979 to present, Copernicus climate change service (C3S) climate data store (CDS)*. https://doi.org/10.24381/cds.f17050d7 .

Klein, S. A. (1997). Synoptic variability of low-cloud properties and meteorological parameters in the subtropical trade wind boundary layer. *Journal of Climate*, *10*(8), 2018–2039. https://doi.org/10.1175/1520-0442(1997)010<2018:svolcp>2.0.co;2

Klein, S. A., Hall, A., Norris, J. R., & Pincus, R. (2017). Low-cloud feedbacks from cloud-controlling factors: A review. *Surveys in Geophysics*, *38*(6), 1307–1329. https://doi.org/10.1007/s10712-017-9433-3

Klein, S. A., & Hartmann, D. L. (1993). The seasonal cycle of low stratiform clouds. *Journal of Climate*, *6*, 1587–1606. https://doi.org/10.1175/1520-0442(1993)006<1587:tscols>2.0.co;2

Klein, S. A., Hartmann, D. L., & Norris, J. R. (1995). On the relationships among low-cloud structure, sea surface temperature, and atmospheric circulation in the summertime northeast Pacific. *Journal of Climate*, *8*, 1140–1155. https://doi.org/10.1175/1520-0442(1995)008<1140:otralc>2.0.co;2

Lin, W., Zhang, M., & Loeb, N. G. (2009). Seasonal variation of the physical properties of marine boundary layer clouds off the California coast. *Journal of Climate*, *22*(10), 2624–2638. https://doi.org/10.1175/2008JCLI2478.1

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., et al. (2018). Clouds and the Earth'S radiant energy system (CERES) energy balanced and filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product. *Journal of Climate*, *31*(2), 895–918. https://doi.org/10.1175/JCLI-D-17-0208.1

Luo, B., & Minnett, P. J. (2020). Evaluation of the ERA5 sea surface skin temperature with remotely-sensed shipborne marine-atmospheric emitted radiance interferometer data. *Remote Sensing*, *12*(11). https://doi.org/10.3390/rs12111873

Martens, B., Schumacher, D. L., Wouters, H., Muñoz-Sabater, J., Verhoest, N. E., & Miralles, D. G. (2020). Evaluating the land-surface energy partitioning in ERA5. *Geoscientific Model Development*, *13*(9), 4159–4181. https://doi.org/10.5194/gmd-13-4159-2020

Mauger, G. S., & Norris, J. R. (2010). Assessing the impact of meteorological history on subtropical cloud fraction. *Journal of Climate*, *23*(11), 2926–2940. https://doi.org/10.1175/2010JCLI3272.1

McCoy, D. T., Eastman, R., Hartmann, D. L., & Wood, R. (2017). The change in low cloud cover in a warmed climate inferred from AIRS, MODIS, and ERA-interim. *Journal of Climate*, *30*(10), 3609–3620. https://doi.org/10.1175/JCLI-D-15-0734.1

Myers, T. A., Mechoso, C. R., Cesana, G. V., DeFlorio, M. J., & Waliser, D. E. (2018). Cloud feedback key to marine heatwave off Baja California. *Geophysical Research Letters*, *45*(9), 4345–4352. https://doi.org/10.1029/2018GL078242

Myers, T. A., & Norris, J. R. (2016). Reducing the uncertainty in subtropical cloud feedback. *Geophysical Research Letters*, *43*(5), 2144–2148. https://doi.org/10.1002/2015GL067416

Myers, T. A., Scott, R. C., Zelinka, M. D., Klein, S. A., Norris, J. R., & Caldwell, P. M. (2021). Observational constraints on low cloud feedback reduce uncertainty of climate sensitivity. *Nature Climate Change*, *11*(6), 501–507. https://doi.org/10.1038/s41558-021-01039-

Norris, J. R., Allen, R. J., Evan, A. T., Zelinka, M. D., O'Dell, C. W., & Klein, S. A. (2016). Evidence for climate change in the satellite cloud record. *Nature*, *536*(7614), 72–75. https://doi.org/10.1038/nature18273

Norris, J. R., Evan, A. T., Norris, J. R., & Evan, A. T. (2015). Empirical removal of artifacts from the ISCCP and PATMOS-x satellite cloud records. *Journal of Atmospheric and Oceanic Technology*, *32*(4), 691–702. https://doi.org/10.1175/jtech-d-14-00058.1

Norris, J. R., & Klein, S. A. (2000). Low cloud type over the ocean from surface observations. Part III: Relationship to vertical motion and the regional surface synoptic environment. *Journal of Climate*, *13*(1), 245–256. https://doi.org/10.1175/1520-0442(2000)013<0245:lctoto>2.0.co;2

Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": A randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, *154*(1), 135–150. https://doi.org/10.1016/S0304-3800(02)00064-9

Pauli, E., Andersen, H., Bendix, J., Cermak, J., & Egli, S. (2020). Determinants of fog and low stratus occurrence in continental central Europe—A quantitative satellite-based evaluation. *Journal of Hydrology*, *591*, 125451. https://doi.org/10.1016/j.jhydrol.2020.125451

Platnick, S., Hubanks, P., Meyer, K., & King, M. D. (2015). *MODIS Atmosphere L3 Monthly Product (08_L3)*. NASA MODIS Adaptive Processing System. Goddard Space Flight Center.

Pokhrel, S., Dutta, U., Rahaman, H., Chaudhari, H., Hazra, A., Saha, S. K., & Veeranjaneyulu, C. (2020). Evaluation of different heat flux products over the tropical Indian ocean. *Earth and Space Science*, *7*, e2019EA000988. https://doi.org/10.1029/2019EA000988

Qu, X., Hall, A., Klein, S. A., & Caldwell, P. M. (2015). The strength of the tropical inversion and its response to climate change in 18 CMIP5 models. *Climate Dynamics*, *45*, 375–396. https://doi.org/10.1007/s00382-014-2441-9

Qu, X., Hall, A., Klein, S. A., & Deangelis, A. M. (2015). Positive tropical marine low-cloud cover feedback inferred from cloud-controlling factors. *Geophysical Research Letters*, *42*(18), 7767–7775. https://doi.org/10.1002/2015GL065627

Randall, D. A., Coakley, J. A., Jr, Fairall, C. W., Kropfli, R. A., & Lenschow, D. H. (1984). Outlook for research on subtropical marine stratiform clouds. *Bulletin of the American Meteorological Society*, *65*(12), 1290–1301. https://doi.org/10.1175/1520-0477(1984)065<1290:ofrosm>2.0.co;2

Refaeilzadeh, P., Tang, L., Liu, H., Angeles, L., & Scientist, C. D. (2009). Encyclopedia of database systems. *Encyclopedia of Database Systems*, *5*, 532–538. https://doi.org/10.1007/978-0-387-39940-9_565

Rieck, M., Nuijens, L., & Stevens, B. (2012). Marine boundary layer cloud feedbacks in a constant relative humidity atmosphere. *Journal of the Atmospheric Sciences*, *69*(8), 2538–2550. https://doi.org/10.1175/JAS-D-11-0203.1

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Scott, R. C., Myers, T. A., Norris, J. R., Zelinka, M. D., Klein, S. A., Sun, M., & Doelling, D. R. (2020). Observed sensitivity of low-cloud radiative effects to meteorological perturbations over the global oceans. *Journal of Climate*, *33*(18), 7717–7734. https://doi.org/10.1175/JCLI-D-19-1028.1

Seethala, C., Norris, J. R., & Myers, T. A. (2015). How has subtropical stratocumulus and associated meteorology changed since the 1980s? *Journal of Climate*, *28*(21), 8396–8410. https://doi.org/10.1175/JCLI-D-15-0120.1

Stirnberg, R., Cermak, J., Fuchs, J., & Andersen, H. (2020). Mapping and understanding patterns of air quality using satellite data and machine learning. *Journal of Geophysical Research: Atmospheres*, *125*, e2019JD031380. https://doi.org/10.1029/2019JD031380

Stirnberg, R., Cermak, J., Kotthaus, S., Haeffelin, M., Andersen, H., Fuchs, J., et al. (2021). Meteorology-driven variability of air pollution (PM1) revealed with explainable machine learning. *Atmospheric Chemistry and Physics*, *21*, 3919–3948. https://doi.org/10.5194/acp-21-3919-2021

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, *36*(2), 111–147. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x

van der Dussen, J. J., de Roode, S. R., Dal Gesso, S., & Siebesma, A. P. (2015). An LES model study of the influence of the free tropospheric thermodynamic conditions on the stratocumulus response to a climate perturbation. *Journal of Advances in Modeling Earth Systems*, *7*, 670–691. https://doi.org/10.1002/2014MS000380

Wood, R., & Bretherton, C. S. (2004). Boundary layer depth, entrainment, and decoupling in the cloud-capped subtropical and tropical marine boundary layer. *Journal of Climate*, *17*(18), 3576–3588. https://doi.org/10.1175/1520-0442(2004)017<3576:bldead>2.0.co;2

Wood, R., & Bretherton, C. S. (2006). On the relationship between stratiform low cloud cover and lower-tropospheric stability. *Journal of Climate*, *19*(24), 6425–6432. https://doi.org/10.1175/JCLI3988.1

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., & Taylor, K. E. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, *47*(1), e2019GL085782. https://doi.org/10.1029/2019GL085782

Zhong, X., Liu, S. C., Liu, R., Wang, X., Mo, J., & Li, Y. (2021). Observed trends in clouds and precipitation (1983–2009): Implications for their cause(s). *Atmospheric Chemistry and Physics*, *21*, 4899–4913. https://doi.org/10.5194/acp-21-4899-2021