Evaluation of Cloud Liquid Water Path Trends Using a Multidecadal Record of Passive Microwave Observations

ANDREW MANASTER and CHRISTOPHER W. O’DELL

Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado

GREGORY ELSAESSER

NASA Goddard Institute for Space Studies, Columbia University, New York, New York

(Manuscript received 21 May 2016, in final form 9 February 2017)

ABSTRACT

In this study, observed cloud liquid water path (LWP) trends from the Multisensor Advanced Climatology of Liquid Water Path (MAC-LWP) dataset (1988–2014) are compared to trends computed from the temporally coincident records of 16 global climate models (GCMs) participating in phase 5 of the Coupled Model Intercomparison Project (CMIP5). For many regions, observed trend magnitudes are several times larger than the corresponding model mean trend magnitudes. Muted model mean trends are thought to be the result of cancellation effects arising from differing interannual variability characteristics and differences in model physics–dynamics. In most regions, the majority of modeled trends were statistically consistent with the observed trends. This was thought to be because of large estimated errors in both the observations and the models due to interannual variability. Over the southern oceans (south of 40°S latitude), general agreement between the observed trend and virtually all GCM trends is also found (about 1–2 g m⁻² decade⁻¹). Observed trends are also compared to those from the Atmospheric Model Intercomparison Project (AMIP). Like the CMIP5 models, the majority of modeled AMIP trends were statistically consistent with the observed trends. It was also found that, in regions where the AMIP model mean time series better captures observed interannual variability, it tends to better capture the magnitude of the observed trends.

1. Introduction

Clouds play an important role in the climate system. They are intimately linked to the global hydrologic cycle and associated condensation and precipitation processes. They also greatly impact the global radiation budget by altering the radiative flux at the top of the atmosphere (TOA) through effects on both the reflected solar shortwave (SW) and outgoing longwave (LW) radiation components. General circulation models (GCMs), including those in phase 5 of the Coupled Model Intercomparison Project (CMIP5), simulate how clouds change in conjunction with a changing climate (i.e., cloud feedbacks), with an aim toward accurately simulating Earth’s future climate and climate sensitivity.

However, it has been found in many studies that cloud feedbacks constitute one of the largest sources of uncertainty in future warming projections (e.g., Stephens 2005; Soden and Held 2006; Randall et al. 2007; Boucher et al. 2013). Therefore, continued development and analysis of cloud feedback-related observational records remains a prerequisite for constraining the spread in simulated cloud–climate feedbacks.

Cloud radiative forcing (CRF), defined as the difference in TOA radiative flux between a cloudy and cloud-free atmosphere, is one important parameter useful for understanding cloud feedbacks and for model validation. Examining long-term CRF changes in a warming climate can add insight into the expected spread in cloud feedbacks (Soden et al. 2004). Several studies (Dessler 2010; Dessler and Loeb 2013; Zhou et al. 2013) have used the CERES-EBAF observational CRF dataset (Loeb et al. 2009) to calculate observed, short-term, global cloud feedbacks in recent years. However, the CERES CRF data only span the past ~15 yr, which makes it difficult to diagnose long-term cloud feedbacks.
due to the time scale dependence of such estimates (Dessler 2010).

The lack of robust, long-term CRF datasets requires that we relate long-term changes in CRF to long-term changes in cloud properties. Among the cloud properties available, cloud fraction is one often-analyzed variable that is retrieved through use of visible–near-infrared (VIS–NIR) satellite sensors [see, e.g., Stubenrauch et al. (2013), and references therein]. Significant efforts have been made to measure trends in cloud fraction, but spurious artifacts can arise in computed trends due to aliasing of satellite drift onto the cloud diurnal cycle (e.g., Waliser and Zhou 1997) and changes in the satellite observing system over time. The latter is a distinct problem when geostationary satellites are involved, especially in the case of ISCCP data (Norris 2000; Campbell 2004; Evan et al. 2007). Thus, generally speaking, trends in the available cloud fraction datasets are viewed skeptically (Stubenrauch et al. 2013). Despite this, many efforts have been made in recent studies to remove artifacts from ISCCP cloud fraction retrievals so that certain aspects of cloud feedbacks can nonetheless be inferred from these observations (Clement et al. 2009; Bellomo et al. 2014; Norris and Evan 2015; Norris et al. 2016).

Another approach involves analyses of cloud water content properties. We have pursued this approach through use of the Multisensor Advanced Climatology of Liquid Water Path (MAC-LWP) dataset (Elsaesser et al. 2017, manuscript submitted to J. Climate) a climate data record that is currently an updated version of the University of Wisconsin cloud liquid water path (LWP) climatology described in O’Dell et al. (2008). This dataset currently provides 27 yr of LWP information over oceans derived from a suite of 12 spaceborne passive microwave sensors (further described in section 2).

This microwave LWP dataset was chosen since it offers several distinct advantages over VIS–NIR cloud property datasets (e.g., ISCCP). First, most VIS–NIR scanners perform cross-track scans and therefore their data are subject to a view angle dependence [as is seen in studies such as Evan et al. (2007)]. Nearly all of the microwave sensors in the MAC-LWP product perform conical scans at approximately 53° and thus their data are not subject to the same view angle dependencies. Second, the VIS–NIR sensors determine cloud properties based on solar reflectance and therefore are only available during the daytime. As discussed, this limitation can result in an aliasing of the diurnal cycle onto long-term measurements for cases where the satellite sensor equatorial crossing time drifts as a function of the year (Waliser and Zhou 1997). In contrast, the passive microwave sensors retrieve cloud properties from observed brightness temperatures, which are available at all hours; thus, a few of the sampling issues that plague VIS–NIR sensors are evaded. The fact that the VIS–NIR sensors determine cloud properties based on solar reflectance measurements also leads to a large solar zenith angle dependence of the cloud properties measurements. In contrast, passive microwave retrievals are insensitive to the solar zenith angle because they measure brightness temperature as opposed to solar reflectance.

We use the new multidecadal LWP dataset to assess the fidelity of multidecadal changes in LWP (i.e., LWP trends that are caused by changes in cloud fraction and cloud optical thickness over time) in CMIP5 models by comparing simulated trends to observed trends both regionally and globally. This approach adds insight toward determining which models may be more likely to correctly replicate long-term changes in CRF (since CRF and LWP are closely related) and thus cloud feedbacks. The rest of this paper is organized as follows: section 2 discusses the observational and model datasets and data record creation, section 3 provides an analysis of regional and global LWP trend results, and section 4 provides further discussion and conclusions.

2. Dataset and methods

a. MAC-LWP dataset

The MAC-LWP dataset (Elsaesser et al. 2017, manuscript submitted to J. Climate) is an updated version of University of Wisconsin (UWisc) cloud liquid water path climatology (O’Dell et al. 2008). It currently contains 27 yr of LWP data (1988–2014) seamlessly merged from 12 different passive microwave sensors (see Fig. 1). The product uses intercalibrated, 0.25° ocean-only retrievals from the Remote Sensing Systems (RSS)
version 7 algorithm (Wentz 1997; Wentz and Spencer 1998; Hilburn and Wentz 2008; Wentz 2013). The dataset provides a monthly mean value of LWP for each 1° × 1° grid box over the ocean for every month. It should be noted that the LWP provided for each grid box represents the average LWP over the entire grid box; thus, sections of clear sky where LWP is theoretically zero are included in the grid box averaged LWP (the CMIP5 output was averaged in the same manner).

MAC-LWP monthly mean LWP values were corrected for the diurnal cycle by fitting for the mean monthly diurnal cycle and monthly mean simultaneously. O’Dell et al. (2008) and Elsaesser et al. (2017, manuscript submitted to J. Climate) provide more information on the diurnal cycle fit process. Systematic errors present in the dataset can be as large as 30% (O’Dell et al. 2008) and include, but are not limited to, cross-talk errors, cloud–rain partitioning, aliasing of scattering signals from ice into the LWP retrieval, cloud-top temperature errors, and clear-sky biases [for more information, see O’Dell et al. (2008)]. We are not able to quantify to what extent such errors affect trends; however, our avoidance of heavily precipitating regions (explained below) does mitigate negative impacts on trend computations.

With respect to cloud–rainwater partitioning using microwave radiometric signals, it is difficult to separate the signal of rainwater from that of cloud water. The RSS retrieval algorithm assumes that clouds precipitate above a total water path threshold of 180 g m⁻². If the retrieved total water path is below this threshold, the cloud is assumed to be nonraining and LWP is equal to the total water path. If the retrieved total water path is above this threshold, the LWP is calculated via the following equation given in Wentz and Spencer (1998):

\[
LWP = 0.18(1 + \sqrt{HR}),
\]

where 0.18 kg m⁻² is the threshold value, \(H\) is the height of the rain column (km), parameterized as a function of the Reynolds SST, and \(R\) is the column averaged rain rate (mm h⁻¹). The freezing level height parameterization is derived from a regression of freezing heights on SSTs, and the SSTs used were from the National Centers for Environmental Prediction (NCEP)–National Center for Atmospheric Research (NCAR) reanalysis (Hilburn and Wentz 2008). Since the threshold for this calculation is likely variable, LWP retrievals may be overestimated (underestimated) if clouds are precipitating at values below (above) this threshold. In addition to the cloud–rain LWP threshold uncertainty, since the microwave signal is proportional to total water path, the potential exists for errors in the LWP retrieval due to an inaccurate (and possibly variable or regime dependent) regression of \(H\) on SST.

The extent to which these issues would affect the computation of trends is unclear. However, we have the ability to partially evade the cloud–rain partitioning problem by analyzing regions where precipitation is more minimal. Figure 2 shows a map of the ratio of 1988–2014 average LWP (taken from the MAC-LWP dataset) to total water path (TWP; also available in the MAC-LWP dataset) over the same period. A ratio of 1.0 indicates that the retrieved total water path, more proportional to the true “observable” as explained, is equivalent to the LWP. As the ratio decreases, an increasing fraction of TWP is forced into the rainwater (associated with nonzero rainfall rates) category. Regions for which rainfall is more substantial, and thus uncertainty in LWP is higher, are associated with ratios of less than 0.6. Therefore, we limit our trend analyses to regions where the 27-yr mean ratio of cloud liquid water path to cloud total water path (LWP/TWP) is greater than 0.6. Quantifying the effect systematic errors in cloud rain-partitioning have on the MAC-LWP product, and impacts on the calculation of LWP trends in higher-precipitation regions, is an important topic for future study.

b. CMIP5 models

For this study, the LWP trends from the MAC-LWP dataset were compared to LWP trends from 16 different models in the CMIP5 collection (Taylor et al. 2012). These models are listed in Table 1. The data were obtained from the Program for Climate Model Diagnosis and Intercomparison (PCMDI) archive (https://pcmdi.llnl.gov/projects/esgf-llnl/; first accessed August 2014).

For this analysis, only data from the r1i1p1 ensemble run for every model were used. A brief analysis for determining the magnitude of the spread in LWP trends
Table 1. CMIP5 models used in this study. An asterisk by the model name indicates an AMIP experiment was available for this given model. (Expansions of acronyms are available online at http://www.ametsoc.org/PubsAcronymList.)

<table>
<thead>
<tr>
<th>Model</th>
<th>Center</th>
<th>Country</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCC_CSM1.1*</td>
<td>Beijing Climate Center</td>
<td>China</td>
<td>Wu et al. (2010)</td>
</tr>
<tr>
<td>CanESM2</td>
<td>Canadian Centre for Climate Modelling and Analysis</td>
<td>Canada</td>
<td>Arora et al. (2011)</td>
</tr>
<tr>
<td>CCSM4*</td>
<td>National Center for Atmospheric Research</td>
<td>United States</td>
<td>Gent et al. (2011)</td>
</tr>
<tr>
<td>CESM1(BGC)</td>
<td>National Center for Atmospheric Research</td>
<td>United States</td>
<td>Lindsay et al. (2014)</td>
</tr>
<tr>
<td>CESM1(CAM5)</td>
<td>National Center for Atmospheric Research</td>
<td>United States</td>
<td>Neale et al. (2012)</td>
</tr>
<tr>
<td>CMCC-CM*</td>
<td>Centro Euro-Mediterraneo per I Cambiamenti Climatici</td>
<td>Italy</td>
<td>Scoccimarro et al. (2011)</td>
</tr>
<tr>
<td>CNRM-CM*</td>
<td>Centre National de Recherches Météorologiques</td>
<td>France</td>
<td>Voldoire et al. (2013)</td>
</tr>
<tr>
<td>CSIRO Mk3.6.0*</td>
<td>CSIRO–Queensland Climate Change Centre of Excellence</td>
<td>Australia</td>
<td>Rotstyn et al. (2010)</td>
</tr>
<tr>
<td>GFDL-CM3</td>
<td>NOAA/Geophysical Fluid Dynamics Laboratory</td>
<td>United States</td>
<td>Donner et al. (2011)</td>
</tr>
<tr>
<td>GFD-ESM2g</td>
<td>NOAA/Geophysical Fluid Dynamics Laboratory</td>
<td>United States</td>
<td>Dunne et al. (2012)</td>
</tr>
<tr>
<td>GISS-E2-H</td>
<td>NASA Goddard Institute for Space Studies</td>
<td>United States</td>
<td>Schmidt et al. (2014)</td>
</tr>
<tr>
<td>GISS-E2-R*</td>
<td>NASA Goddard Institute for Space Studies</td>
<td>United States</td>
<td>Schmidt et al. (2014)</td>
</tr>
<tr>
<td>INM-CM4*</td>
<td>Institute of Numerical Mathematics</td>
<td>Russia</td>
<td>Volodin et al. (2010)</td>
</tr>
<tr>
<td>MIROC5*</td>
<td>AORI, NIES, and JAMSTEC</td>
<td>Japan</td>
<td>Watanabe et al. (2010)</td>
</tr>
<tr>
<td>NorESM*</td>
<td>Norwegian Climate Centre</td>
<td>Norway</td>
<td>Kirkevåg et al. (2013)</td>
</tr>
</tbody>
</table>

across ensemble runs was performed. It was found that the spread in trends between various ensemble members was much smaller than the intermodel differences, therefore affirming the use of only one ensemble. For each model the “historical” and “rcp4.5” experiments were combined to create a time series that spanned the exact same length as the MAC-LWP dataset (27 yr from 1988–2014). The historical experiments are simulations of past climate from 1850–2005 that include changing environments consistent with observations. These can include anthropogenic and volcanic influences on atmospheric composition, solar forcing, emissions or concentrations of short-lived species and natural and anthropogenic aerosols, and land use (Taylor et al. 2012). The rcp4.5 experiments are simulations of future climates (2006–2100) that are forced by representative concentration pathway 4.5 (i.e., these experiments result in a radiative forcing of 4.5 W m⁻² by 2100 relative to preindustrial conditions; Taylor et al. 2012).

To facilitate a fairer comparison, all model data were regridded to the same 1° × 1° grid spacing as the observations, using a simple bilinear interpolation. Additionally, grid boxes for which observations were missing in a given month were also set to missing in the model output. Grid boxes in a given month that had LWP/TWP ratios of below 0.6 in the observations were masked in both the observations and the models, again in accordance with our goal of avoiding potentially large biases arising from cloud–rain partitioning problems. The observed and model records were deseasonalized by removing the mean seasonal cycle in each dataset. Deseasonalizing removes potential errors in trends caused by the seasonal cycle in LWP, although this effect is typically small in such a long dataset. These errors arise from potential nonuniform sampling of the seasonal cycle in different regions of Earth.

Since LWP does not correspond to a specific output variable in the CMIP5 collection, we compute it by subtracting the ice water path variable (“clivi”) from the total water path variable (“clwvi”). For some models, however, clwvi was liquid water path (e.g., Jiang et al. 2012). This was evident when subtracting clivi from clwvi resulted in large negative values for LWP. These models included CCSM4, CMCC-CM, CESM1(CAM5), and IPSL-CM5A-MR. A list of other models where this error is present can be found on the PCMDI CMIP5 errata web page (http://cmip-pcmdi.llnl.gov/cmip5/errata/cmip5errata.html).¹

C. AMIP experiments

Observed LWP trends were also compared against LWP trends from the Atmospheric Model Intercomparison Project (AMIP) experiments. While CMIP and AMIP experiments use the same atmospheric model, in the AMIP experiments prescribed SSTs are used in substitution of an ocean model, and thus deficiencies in the atmospheric model may be more easily revealed (along with interannual variability differences with respect to observations being further minimized).

In this study, only 11 of 16 models we examined performed the AMIP experiment. The modeling centers

¹ The CMCC was notified of this problem since their model did not show up as one containing this error on the errata web page. It also should be noted that in recent studies (Jiang et al. 2012; Lauer and Hamilton 2013), the CSIRO Mk3.6.0 model was found to contain this error as well; however, for this analysis, the output appears to have been corrected.
that ran the AMIP experiments are indicated by an asterisk in Table 1. Different models had various ending years for their corresponding AMIP experiments, with some ending as early as 2008 while others ended in 2012. For consistency and in order to examine as much output as possible, all AMIP trends were computed from 1988–2008 data. The AMIP output was otherwise handled in the exact same manner as the CMIP5 model data.

d. Calculation of trends, errors, and statistical significance

LWP trends were calculated in the observed and model datasets, and then compared in order to assess the realism of the simulated LWP trends. Four different regions were chosen for comparison: the North American stratocumulus deck (15.5°–35.5°N, 144.5°–124.5°W), the South American stratocumulus deck (29.5°–9.5°S, 89.5°–69.5°W), the Namibian stratocumulus deck (20.5°–35.5°S, 0.5°–10.5°E), and the southern oceans (59.5°–44.5°S, all longitudes). These regions were chosen according to the predominant cloud regime, which was generally uniform and not heavily precipitating (i.e., there were relatively few grid boxes containing LWP/TWP ratios < 0.6). These regions are highlighted in Fig. 3.

Trends were calculated in this study by taking the area-weighted average of LWP for each month in a given region, deseasonalizing the resulting time series, and computing a best fit line using an equal-weighting linear least squares fit. Only months for which 85% (or greater) of the region contained valid LWP data (typically due to coastal proximity and/or persistent sea ice).

Once trends were calculated, trend errors were computed based on the method outlined in Santer et al. (2000) given by the following equation:

$$\sigma = \sqrt{\frac{1}{\text{DoF}} \sum_{i=1}^{N} (y_i - a - b t_i)^2 / \sum_{i=1}^{N} (t_i - \bar{t})^2},$$

where $y_i$ are the deseasonalized monthly LWP values at each month, $t_i$, $a$, and $b$ are the linear fit parameters (intercept and slope), and DoF is the effective number of degrees of freedom in the fit (which takes into account autocorrelation in the deseasonalized LWP time series). Specifically, $\text{DoF} = N[(1 - c)(1 + c)] - 2$, where $N$ is the number of months in the time series, and $c$ is the lag-1 autocorrelation coefficient of the fit residuals. Typically, $c$ was relatively small, so DoF was often close to $N - 2$. This trend error computation is rather conservative, and any deviations in the deseasonalized LWP time series apart from a linear change remain unquantified and folded into the trend error. Trend errors can be reduced by regressing interannual variability [e.g., El Niño–Southern Oscillation (ENSO)] out of the deseasonalized time series or by increasing the length ($N$) of the record. Trends are discussed in the following section.

3. Results

a. Spatial patterns of trends

Figure 4 shows the multidecadal trends in observed LWP, the AMIP model mean, observed trends after an ENSO signal is regressed out, and the CMIP5 model mean trends. Regions where the 27-yr mean LWP/TWP ratio is less than 0.6 have been grayed out. First, note that the magnitudes of the trends in the observations (Figs. 4a,c) are larger than those of Figs. 4b and 4d, particularly evident when comparing the CMIP5 model mean trend (e.g., observed LWP trends can be up to 4 times larger in magnitude in regions such as the equatorial Pacific away from the heavily precipitating ITCZ, and south of the Hawaiian Islands).

Figure 5 provides one visual way to understand the trend magnitudes disparities between the model means and observations. Figures 5a–e show 27-yr global LWP trends from different ensemble runs of the MIROC5 CMIP5 model. Figure 5f shows the ensemble mean of Figs. 5a–e. Note that the trends in Figs. 5a–e are, in general, larger than most of the trends depicted in Fig. 5f. This muting of trend magnitudes is likely due to cancellation effects arising from different interannual
(and longer time scale) cycles among many ensemble runs. By analogy, this argument can be applied to the individual model–model mean comparisons. For many locations on Earth, unconstrained by a fixed SST state, each CMIP5 model has its own climate variation (a unique mean trend, and unique modes of interannual variability). An average of disagreeing climate states and trends gives overall muted model-mean trends.

This argument is supported upon analysis of the AMIP model mean trends (Fig. 4b). Since all AMIP models have identical prescribed SSTs, climate cycles are more in phase (i.e., the timing of ocean driven events such as ENSO is the same among the models). Thus, the AMIP model mean trends are not nearly as dampened as the CMIP5 model mean trends. Additional analyses of the AMIP model mean and individual AMIP model trends are provided in section 3c.

In an attempt to assess the impact of one mode of interannual variability on a trend computation, we removed the ENSO signal from the observations (Fig. 4c) by removing (from each grid box LWP time series) the LWP signal associated with the Multivariate ENSO Index (MEI; Wolter and Timlin 1993) using standard linear regression. We might expect that the removal of one major source of interannual variability from the observations will lead to better agreement with the CMIP5 model mean trends in some locations. It can be seen in Fig. 4c that most areas spanning the globe exhibit little to no change in trends once ENSO is regressed out. More moderate changes can be seen in the equatorial Pacific although these are confined to a relatively small area [which noticeably improves the agreement with the AMIP mean in the region between the intertropical convergence zone (ITCZ) and South Pacific convergence zone (SPCZ) around 5° S from 135° E to 165° W]. This result may imply that ENSO, as reflected in the MEI, is not the major source of interannual variability affecting most observed trends globally and, with this in mind, LWP trends are a product of forced trends and one or more other sources of interannual variability. Another possibility is that the forced trends common among CMIP5 models are not being modeled correctly (i.e., they are too weak relative to the observed forced trends). Indeed, other modes of variability potentially impacting LWP trends include the Pacific decadal oscillation (PDO), Atlantic
multidecadal oscillation (AMO), and other flavors of ENSO (i.e., a number of indices exist for quantifying this mode of variability).

Another plausible reason for weakened CMIP5 model mean trends could be related to how each climate model represents certain climatic features relative to one another (e.g., the amplitudes, cycles, and locations of the ITCZ, SPCZ, and subsidence zones of the subtropics). Additionally, model-dependent modes of intraseasonal variability (e.g., models resembling ENSO) might have to be regressed out of each CMIP GCM individually; after this, each model field would need to be averaged to create an average CMIP mean. Overall, differences in locations of features and differing representations of atmospheric variability may lead to trend cancellation effects arising from averaging positive and negative trends over the same geographic regions. Figure 6 shows the zonally averaged trends for the observations (with the MEI ENSO signal regressed out), the CMIP5 model mean, and the 16 individual CMIP5 models used in this study. Grid boxes that had 27-yr mean LWP/TWP ratios of less than 0.6 were not included in these zonal averages. Grid boxes spanning the southern oceans (south of about 45°S) are in agreement with the observations in terms of the sign of the trend;
however, observed trends are larger than the CMIP5 model mean trend. Figure 6 shows that many of the individual models in these regions seem to agree with the observations on the trend sign as well; however, their magnitudes tend to be less than that which is observed. With respect to the large discrepancy in trends from just north of the equator (where observations indicate a large positive trend) and around the latitude band centered near 10°N (notice the large negative observed trend), it is possible that these discrepancies are susceptible to our chosen LWP/TWP threshold, given that these regions straddle the gray zones striped out in Fig. 4. Thus, while it is possible these observed trends, in terms of signs and magnitudes, are reasonable, we limit our discussion of the differences in these regions given that LWP/TWP ratios are between 0.6 and 0.8 [which is not the case for the low cloud and southern oceans grid boxes we highlight (in Fig. 3) for later analyses, where ratios are much larger].

Aside from the southern oceans and regions straddling the ITCZ, other regions exhibit interesting differences and similarities between the CMIP5 model mean and the observations (Figs. 4 and 6). For instance, the model mean, as well as many individual models, appear to oppose the observed negative LWP trends in the North Pacific and North Atlantic storm tracks (40°–60°N; Figs. 4 and 6). It should be noted that the difference over the North Atlantic is not nearly as prominent when comparing the AMIP model mean trends to the observations. This is possibly due to the AMIP model mean simulating observed interannual variability more accurately (a direct result of using prescribed SSTs). In the next section, we quantitatively examine regional trends in LWP, and take care to ensure that our LWP/TWP ratios are even higher (largely greater than 0.8) in these focused regions to avoid rain–cloud partitioning issues as best we can.

b. Regional trends

Figure 7 shows the range of values for trends in the four regions described in section 2c (and highlighted in Fig. 3). Light gray bars represent the number of model trends that fall within a given range of trend values, while the dark gray bars represent the number of those that are statistically consistent with the observations at 95% confidence. To compute statistical consistency between a model and observations, the following equation was used:

\[
(T_1 - T_2) \pm \sqrt{\sigma_1^2 + \sigma_2^2},
\]

where \(T_1\) is the modeled trend, \(T_2\) is the observed trend, \(\sigma_1\) is the error associated with the modeled trend (as calculated from Santer et al. 2000), and \(\sigma_2\) is the error associated with the observed trend. If the difference in trends divided by the errors that were added in quadrature is less than 2 (i.e., the difference in trends is less than two standard deviations away from zero), then the modeled and observed trends are said to agree with one another at 95% confidence. The blue line represents the observed trend with the blue shading indicating the 2σ error associated with the observed trend. Similarly, the orange line indicates the model mean trend for each region.

From Fig. 7, we find that for the North American, Namibian, and South American stratocumulus deck regions (Figs. 7a, 7b, and 7c, respectively) the majority of the models agree with the observed trends at 95% confidence. This is likely due to the fact that errors (in part due to interannual variability) are relatively high in the models and the observations in these regions. In fact, none of the observed trends and no more than two model trends for these three regions were found to be statistically significant (i.e., statistically different from zero at 95% confidence). This indicates that the errors on the observed and most modeled trends in these regions are relatively high compared to the trends themselves. We conclude from this that, although the modeled LWP trends in these three regions generally agreed with the observations (within the allotted errors), these errors were relatively high because interannual variability plays a significant role even in a 27-yr record.

Perhaps the most interesting region is the southern oceans. For this region (Fig. 7d), almost every model trend is positive with nearly the same magnitude as the observed trend. Furthermore, 13 of the 16 model trends were found to be statistically significant. Figure 7d indicates that these trends, with 11 of 16 model trends agreeing with the observed trend at 95% confidence. Also, unlike other regions analyzed, the observed trend in the southern oceans is statistically significant (i.e., the error relative to the observed trend is low enough that it is statistically different from zero at 95% confidence). This suggests that, in addition to most of the modeled and observed errors being small relative to their respective trends, the trends themselves are very similar to one another. No other regions exhibit this measure of agreement. For the southern oceans, we hypothesize that interannual variability has less of an effect on our modeled and observed trends, and/or the components of the trends that are forced outweigh the effects of interannual variability. Unfortunately, given the nature of our dataset, we are unable to determine whether or not these strong forced trends are driven by thermodynamics or dynamics. However, several recent studies (McCoy et al. 2014; Ceppi et al. 2016) suggest that optical thickening of clouds spanning the southern oceans in observations and models could be due to phase
changes brought on by warming or increases in adiabatic water content with a warming climate. If this is the case, it would indicate that the positive southern oceans trends are primarily thermodynamics driven.

c. AMIP regional trends

It has been shown that many CMIP5 models agree with the observed liquid water path time series in various regions globally to within their respective errors; however, these errors were relatively high because interannual variability plays a significant role even in a 27-yr record. A similar analysis was performed with the 11 AMIP model simulations for the same four regions. Model AMIP trends from 1988 to 2008 were compared with observed trends for the same time period and statistical consistency was calculated again using Eq. (3). The results of this analysis can be seen in Fig. 8. Similar to Fig. 7, the majority of trends in the North American, Namibian, and South American stratocumulus decks agree with the observations at 95% confidence. Also like the CMIP5 models, none of the observed trends and no more than two models in any of these regions are statistically significant. Again, this is likely due to the fact that the errors due to interannual variability are relatively high in both the observations and the models in these regions.

As in the CMIP5 models, the observed trend in the southern oceans is statistically significant; however, unlike the CMIP5 models, few models are actually statistically significant (4 of 11 AMIP simulations as

![Figure 7](https://example.com/fig7.png)

**Fig. 7.** The LWP trends in the models and the observations in the (a) North American stratocumulus deck, (b) Namibian stratocumulus deck, (c) South American stratocumulus deck, and (d) southern oceans regions for the past 27 yr. Light gray bars indicate the number of models that fall in a given range of trend values. Dark gray bars indicate the number of those that are statistically consistent. The blue line indicates the observed trend with the light blue shading indicating the uncertainty associated with these observations at 2σ [calculated using the method outlined in Santer et al. (2000)]. The orange line indicates the model-mean trend value.
opposed to the 13 of the 16 in the CMIP5 simulations). So while interannual variability may have less of an effect on some modeled AMIP trends and our observed trends in this region, these results may indicate, even at a time scale of 6 years less than that of our full dataset, that interannual variability can play an even stronger role.

Because of its large role in determining how well both AMIP and CMIP simulations capture observed trends, it seems pertinent to examine how well these simulations capture interannual variability. Figure 9 shows the observed, CMIP5 model mean, and AMIP model mean time series as well as the observed 1988–2008 trend for the four regions described in section 2c. The $R$ values represent the detrended correlation coefficient between the observed and AMIP model mean time series and serves as a metric for how well the AMIP model mean captures the observed interannual variability. The higher the correlation coefficient, the better the AMIP model mean does at replicating the interannual variability of the observed time series. It is worth noting that no heavily raining grid boxes (i.e., grid boxes with an LWP/TWP ratio of less than 0.6) have been removed in the calculation of these time series [although the occurrences of such ratios are minimal, as we mentioned earlier (and also suggested by Fig. 2)]. This ensures that very cloudy (or rainy) grid boxes influencing interannual variability were not artificially removed.

Note that the four AMIP model mean time series are not as dampened as the CMIP5 model mean time series in Fig. 9. Since each CMIP5 model has its own climate variation (i.e., a unique mean trend, and unique modes of interannual variability), an average of these climate states will give an overall muted model-mean trend. Conversely, since all of the AMIP models have identical prescribed SSTs, the timing of ocean driven events, influencing interannual variability (i.e., ENSO), is more similar. Thus, the AMIP model mean trends are not nearly as dampened as the CMIP5 model mean trends in
some regions. These results are consistent with those presented in section 3a.

Although the amplitude of variability in AMIP model mean time series is greater than the CMIP5 counterpart, differences in the detrended correlation coefficient (i.e., the ability of the AMIP model mean to simulate interannual variability) exist between the observed and AMIP model mean time series in these four regions. Specifically, the correlation coefficients for the North American and Namibian stratocumulus deck regions are much higher (approximately 0.6) than those of the South American stratocumulus deck and the southern oceans region (between 0.30 and 0.40). For the two regions with higher correlation coefficients, it can be seen that the resulting AMIP model mean trends are closer to the observed trends than the CMIP5 model mean trends, particularly over the North American stratocumulus deck where the AMIP time series closely tracks the observed time series (and the resulting trend is almost identical to the observed 1988–2008 trend). Conversely, in the two regions with lower correlation coefficients (the South American stratocumulus deck and the southern oceans), the AMIP model mean appears to perform either the same as or slightly worse than the CMIP5 model mean with respect to observations.

4. Summary and discussion

In this work, global and regional trends in the observations and the model means were compared. It was found that observed trends were larger in magnitude (or of different sign) for a number of regions when compared to the model mean (e.g., North Atlantic storm track; southern oceans). In addition to physics/parameterization differences, this may be caused by cancellation effects arising from differing interannual variabilities between models.
(discussed with respect to the model ensembles in Fig. 5, and results in Fig. 9). For regions where observed and model trends were in agreement on sign (e.g., some of the low cloud regions discussed in Fig. 9), differing interannual variabilities may play a greater role. AMIP model mean trends were found to be in better agreement with the observations than those from CMIP5 (perhaps less surprising given the use of prescribed/observed SSTs); the agreement is at least partly the result of AMIP runs constraining interannual variability (by construction) in a way that more closely resembles observation. Regressing the ENSO signature out of the LWP trends was found to have little impact on most regional LWP trends, with the exception of those in the tropical Pacific; thus, trends in many regions remained substantially larger than those in the CMIP5 model mean. From this, it was concluded that either the larger trends in the observations are due to non-ENSO sources of variability not simulated by most models or that forced trends common among CMIP5 models are inadequately simulated.

CMIP5 trends were examined in individual regions. Statistical errors were calculated for each trend, under the assumption that nonlinear behaviors in the time series are due to unknown physics (and are therefore mapped into trend errors). In only one of four regions analyzed were the observational trends statistically different from zero at 95% confidence. This indicates that interannual variability effects in the observations were less impactful or that the time series was of sufficient length relative to the amplitude of cyclic variability expected.

In most regions, few CMIP5 models were found to have trends that were statistically different from zero at 95% confidence (i.e., statistically significant). However, the majority of modeled trends in these regions were statistically consistent with the observed trends (i.e., they agreed with the observations at 95% confidence), although this was typically due to large estimated errors in the observations and/or models, most likely caused by large interannual variability. Without a longer time series or more negligible interannual variability, it remains difficult to quantify regional model trends from single ensemble runs at high confidence. Large errors in trends themselves arise from interannual variability that is substantial enough to impact analyses of 27-yr-long regional records; we conclude that these large errors contribute to the good agreement between CMIP5 LWP trends and observations for a number of regions.

In the southern oceans, the majority of CMIP5 modeled trends were found to be both statistically significant and statistically consistent with the observations indicating that not only are most of the modeled and observed errors small relative to their respective trends, but also the trends themselves are very similar in magnitude/sign. For this region, we hypothesize that interannual variability has less of an effect on our modeled and observed trends.

AMIP trends were examined in individual regions in a similar manner to the CMIP5 trends. It was found that, like CMIP5, very few AMIP models were statistically different from zero in almost every region. However, the majority of modeled trends in these regions were statistically consistent with the observed trends. Unlike the CMIP5 model analysis however, few model trends were found to be statistically significant in the southern oceans, which we hypothesize is due to interannual variability imparting a larger footprint on the time series that is 6 years shorter than the full MAC-LWP record (due to temporal overlap requirements with the AMIP periods and observations). Note that for regions where the AMIP model mean better captured observed interannual variability, simulated trends were also closer to observed trends. We again interpret such results as indicative of trends being dominated by interannual variability in such locations.

As the MAC-LWP record increases in temporal length, observed trends will become more robust and accompanied by reduced errors, thereby making this dataset an even better diagnostic tool for trends. Additionally, the effects that oscillations acting on multi-decadal time scales have on LWP trends will become more apparent. Understanding the impact that interannual variability has on long-term LWP records and in model outputs is of the utmost importance. By averaging together several available model ensembles, we can potentially reduce the effects of interannual variability on trends in model composites, thus leaving us with a clearer signal of forced trends (see, e.g., Fig. 5). Either way, reducing errors tied to interannual variability in observations and models is important for understanding LWP trends. Characterizing possible algorithm-related trend artifacts remains an important area of future research, particularly in regions where heavier precipitation is observed.

Finally, it may be possible to quantify the relationship between LWP and CRF. This would allow us to relate the observed LWP trends to changes in radiative forcing and hence cloud feedbacks. For instance, Bellomo et al. (2014) analyzed trends in observed and simulated cloud cover toward estimating “cloud amount” feedbacks. Their work is consistent with our results in that cloud cover trends to decrease over the northeast Pacific and increase over the southeast Pacific from the beginning of the twentieth century onward. All else being equal, solely increasing SST would likely be associated with decreasing clouds in subsidence.
regions such as the northeast (North American) and southeast (South American) Pacific stratocumulus decks. However, both the Bellomo et al. (2014) and our study find that cloud cover/LWP actually increases (i.e., positive trend) over the South American stratocumulus regions. While the causes are not fully understood, we speculate this may be due to a weakening of the Walker circulation (e.g., Bellomo and Clement 2015). Nonetheless, our work and other analyses (e.g., Zelinka et al. 2012a,b; Bellomo et al. 2014; Bellomo and Clement 2015) suggest that changes in LWP are relevant to cloud feedbacks; therefore, even an approximate quantification would be useful for reducing the spread in modeled cloud feedbacks.

Acknowledgments. This work was funded through the NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) project, via subcontract with the NASA Jet Propulsion Laboratory (JPL). We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Special thanks to Libby Barnes and Mark Zelinka for their expertise and helpful comments regarding this work. We also thank the three anonymous reviewers for their constructive feedback, which greatly helped to improve the paper.

REFERENCES


