AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2017JD027163

Key Points:
- Clustering defines a set of atmospheric states for the great plains that are valuable tools for evaluation of model cloud occurrence
- The NOAA/GFDL AM3 model does not produce enough cirrus but has counterbalancing errors in thick cloud occurrence
- Increasing model horizontal resolution improves state cloud properties but removes counterbalancing errors and increases the overall bias

Supporting Information:
- Supporting Information S1

Correspondence to:
S. Evans,
stuartev@buffalo.edu

Citation:
Evans, S., Marchand, R., Ackerman, T., Donner, L., Golaz, J.-C., & Seman, C. (2017). Diagnosing cloud biases in the GFDL AM3 model with atmospheric classification. *Journal of Geophysical Research: Atmospheres*, *122*. https://doi.org/10.1002/2017JD027163

Received 18 MAY 2017
Accepted 9 NOV 2017
Accepted article online 16 NOV 2017

Diagnosing Cloud Biases in the GFDL AM3 Model With Atmospheric Classification

Stuart Evans[1,2], Roger Marchand[3,4], Thomas Ackerman[3,4], Leo Donner[2,5], Jean-Christophe Golaz[6], and Charles Seman[2]

[1]Princeton Environmental Institute, Princeton, NJ, USA, [2]Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA, [3]Department of Atmospheric Sciences, University of Washington, Seattle, WA, USA, [4]Joint Institute for the Study of the Atmosphere and Ocean, Seattle, WA, USA, [5]Program in Atmospheric and Oceanic Sciences, Princeton, NJ, USA, [6]Lawrence Livermore National Laboratory, Livermore, CA, USA

Abstract We define a set of 21 atmospheric states, or recurring weather patterns, for a region surrounding the Atmospheric Radiation Measurement Program's Southern Great Plains site using an iterative clustering technique. The states are defined using dynamic and thermodynamic variables from reanalysis, tested for statistical significance with cloud radar data from the Southern Great Plains site, and are determined every 6 h for 14 years, creating a time series of atmospheric state. The states represent the various stages of the progression of synoptic systems through the region (e.g., warm fronts, warm sectors, cold fronts, cold northerly advection, and high-pressure anticyclones) with a subset of states representing summertime conditions with varying degrees of convective activity. We use the states to classify output from the NOAA/Geophysical Fluid Dynamics Laboratory AM3 model to test the model's simulation of the frequency of occurrence of the states and of the cloud occurrence during each state. The model roughly simulates the frequency of occurrence of the states but exhibits systematic cloud occurrence biases. Comparison of observed and model-simulated International Satellite Cloud Climatology Project histograms of cloud top pressure and optical thickness shows that the model lacks high thin cloud under all conditions, but biases in thick cloud are state-dependent. Frontal conditions in the model do not produce enough thick cloud, while fair-weather conditions produce too much. We find that increasing the horizontal resolution of the model improves the representation of thick clouds under all conditions but has little effect on high thin clouds. However, increasing resolution also changes the distribution of states, causing an increase in total cloud occurrence bias.

Plain Language Summary Models generally struggle to simulate clouds. Identifying the processes that cause errors is an important step toward improving them. We define a set of weather patterns for a region in the Great Plains that represent different physical processes and evaluate a climate model's cloud occurrence for each of those patterns. The model underpredicts cirrus clouds for all patterns. For thick clouds, however, the model overpredicts during fair-weather conditions and underpredicts during stormy conditions. These errors tend to balance each other out. When the model resolution is improved, it does a better job of predicting thick clouds for most weather patterns, but now, the errors no longer balance each other. The result is that the model with better resolution has a worse overall prediction of thick clouds, despite better predictions for most individual patterns. Evaluating models by pattern, rather than just the overall total, helps to identify when there are underlying improvements that might be missed otherwise. Doing so may be valuable for future efforts toward improving the simulation of clouds in models.

1. Introduction

The representation of clouds and cloud radiative effects has been a persistent problem for general circulation models (GCMs) for many years. A complicating factor in the model development process is that evaluating the quality of a parameterization and identifying areas for improvement can be difficult as well (Jakob, 2010). In order to diagnose when and where a parameterization needs improvement, one must first identify the particular conditions under which it performs unsatisfactorily. A frequent approach to this problem is to composite cloud properties according to the state of the atmosphere, weather regime, or cloud type. Doing so helps to unmask any offsetting biases, gives insight into which physical conditions are simulated well or poorly, and allows for the separation of errors due to misrepresenting the distribution of inputs to a

parameterization from errors due to the parameterization itself. A wide variety of classification methods have been used in pursuing this approach, including using the 500 mb vertical velocity to identify regions of large-scale ascent and subsidence (Bony & Dufresne, 2005), clustering radiosonde (Pope et al., 2009), or C-band radar data (Caine et al., 2009) to create weather or precipitation regimes, or using reanalysis data in either a clustering (Fereday et al., 2008; Mason et al., 2014) or self-organizing maps (Kennedy et al., 2016) algorithm to produce regional atmospheric states. Use of the states from studies such as these have proven effective at understanding the variability of clouds and precipitation a particular region experiences (Evans et al., 2014; Muhlbauer et al., 2014). A number of studies have applied clustering algorithms to cloud observations, particularly the cloud top pressure-optical depth joint histograms from the International Satellite Cloud Climatology Project (ISCCP; Rossow & Schiffer, 1999) in order to create sets of cloud regimes that have been used to describe cloud and radiative properties in observations (Jakob & Schumacher, 2007; Jakob et al., 2005; Oreopoulos et al., 2014) and to evaluate the performance of GCMs (Jin et al., 2017; Mason et al., 2015; Williams & Webb, 2008). These cloud regimes effectively identify the types of cloud that contribute most to particular model biases, but they do not directly address the physics that contribute to errors in model cloud properties.

In this study we aim to diagnose the conditions that produce cloud occurrence errors in the GFDL AM3 model (Donner et al., 2011). As clouds are more directly the result of atmospheric conditions than vice versa, we choose to define our states using atmospheric rather than cloud variables, and then composite cloud properties based on those states. We define our states using an iterative clustering technique first developed in Marchand et al. (2006), and later refined in Marchand et al. (2009) (hereafter M09) and Evans et al. (2012) (hereafter E12). This technique uses ground-based cloud observations to refine and validate atmospheric states that have been defined by dynamic and thermodynamic variables for a region surrounding the observation site. M09 applied the algorithm to a region surrounding the Atmospheric Radiation Measurement (ARM, Ackerman & Stokes, 2003) program's Southern Great Plains (SGP) site in central Oklahoma. The M09 study used a 2 year time series of data to define the states, of which it found 12, and compared the observed vertical cloud occurrence profiles to those produced by the Multiscale Modeling Framework climate model.

We revisit the SGP site in this study; however, we use a much longer time series of data, allowing the algorithm to discern more subtle differences between states, and we evaluate cloud occurrence in AM3 based on both cloud top height and optical thickness. Under climate scenarios with increased CO_2, studies suggest that the land surfaces will warm, and much of the southern United States, including the SGP region, are likely to experience drier conditions with less total precipitation (Shafer et al., 2014; Trenberth, 2011), while individual storms will have shorter duration and larger average intensity (Jiang et al., 2016). It is therefore critical to test climate models for this region. The SGP site also carries with it the benefit of experiencing a broad variety of weather and atmospheric conditions. The frequent cyclones and anticyclones that characterize the day to day weather at the site, along with the large ranges in temperature and humidity, suggest that examining clouds simulated at the SGP site tests the GCM physics in a wide variety of conditions.

In evaluating a GCM, a benefit of using a classification technique based on atmospheric properties is that we are able to parse the model's occurrence bias of different cloud types into contributions from errors in within-state properties and from errors in the distribution of states. In doing so, we identify the particular physical conditions that are most important to the model's cloud biases. In this study, we perform this error decomposition for two versions of the AM3 model that have different horizontal resolution, having approximately $2° \times 2.5°$ and $0.5° \times 0.5°$ horizontal grid spacing but are otherwise identical. Evaluating the model at multiple resolutions allows us to quantify whether the cloud occurrence biases are sensitive to model resolution and thus whether they are likely improve or degrade as GCMs inevitably move to higher resolution. This is of particular interest at a site like SGP, where many of the weather patterns experienced (e.g., fronts) have sharply defined features that GCMs have insufficient resolution to properly represent. While neither experiment explicitly resolves convection or fronts, determining whether higher resolution nonetheless improves the representation of these features, or whether improvements to the parameterizations is also necessary is potentially valuable information for future model development.

In section 2 we summarize the iterative clustering process from M09 and E12 that we use to create the atmospheric states and how we later use those states to classify model output. Section 3 describes the occurrence

and meteorology of the set of atmospheric states produced by our algorithm, with more detailed information on the states in the supporting information. In section 4 we use the states to evaluate cloud occurrence in two AM3 simulations with differing horizontal resolution and identify sources of bias. In section 5 we summarize and discuss our findings.

2. Classification Technique

We evaluate cloud occurrence in the GFDL AM3 using a set of atmospheric states as a basis for compositing. The states that we use for compositing are defined by an iterative cluster analysis algorithm that uses reanalysis data as inputs (described in further detail in M09 and E12). In this study, we use the Interim Re-analysis product from the European Centre for Medium-Range Weather Forecasts (ERA-Interim; Dee et al., 2011) as the source of input data. We sample temperature, the zonal and meridional winds, relative humidity, and surface pressure on a 9×9 horizontal grid with $1.5° \times 1.5°$ spacing centered on the SGP site. The variables other than surface pressure are also sampled on seven pressure levels ranging from 1,000 mb to 250 mb. Both the horizontal locations and pressure levels are those used in the ERA-Interim product. Collectively, these inputs comprise a three-dimensional snapshot of the regional state of the atmosphere over the Great Plains at each time step. These snapshots are collected four times daily for a period from 1996 to 2010, amounting to nearly 20,000 snapshots of the region.

These snapshots of regional meteorology are input to a neural network classifier, an algorithm that finds commonly occurring patterns in the data. The classifier outputs a predetermined number of clusters, representing an initial estimate of the atmospheric states for the region. We used a large number, 40, but testing showed the final product to be largely insensitive to this number. An issue common to all classification studies is how to determine the proper number of classes. In our case, we have no a priori reason to choose a particular number of states, so we undertake a second, iterative stage to our classification to determine the proper number of atmospheric states for the region. To do this, we introduce cloud occurrence data from the Active Remote Sensing of Clouds product (a combination of vertically pointed millimeter cloud radar, micropulse lidar, and laser ceilometers) at the ARM SGP site (Clothiaux et al., 2000), and using the initial time series of state we composite vertical profiles of cloud occurrence for each state. We then apply two statistical tests to these profiles—first a temporal stability test, then a distinctness test. For the stability test, we compare the average cloud profile from the first half of the data set to the average profile from the second half. If these two profiles are not statistically similar, as determined by a bootstrap resampling test (Marchand et al., 2006), we consider the relationship between the state's large-scale environment and its cloud properties to be temporally unstable and thus not suitable for this study. For the distinctness test we compare the average cloud occurrence profile from each state to those from all the other states. If the average cloud profile of one state is not statistically distinct from the profile of any other state, we consider the two states to be indistinct from each other, and again not suitable for this study. Cloud occurrence data were chosen for the role of determining state quality because it was not a part of the initial classification, making it an independent test of the reanalysis-defined states. Further, E12 demonstrated that these single location radar data are representative of cloud occurrence over a 5° region, including for individual states.

States that fail either of the statistical tests are then treated in one of two ways. States that represent a relatively large number of snapshots, defined as greater than 3% of the total input snapshots, are divided in two using a k-means clustering algorithm, increasing by one the total number of state definitions for the region. States that represent a small number of snapshots are simply removed from the list of state definitions, reducing by one the number of states. In both cases, all snapshots are then resorted into the new set of states and the statistical tests are applied again. This process is repeated until all states pass both the temporal stability and distinctness tests. The exact number and definition of the states is somewhat arbitrary, as they depend on the strictness of the statistical tests, but we consider this final set of states to be a good representation of the variety of weather experienced by the region and suitable for compositing studies of the region. Once a final set of states has been created, every snapshot has been assigned to the atmospheric state it most closely resembles. Doing so creates a time series of atmospheric state for the duration of the observed period. Other data sets, both ground- and satellite-based, observed during this time can now be composited according to atmospheric state. In this study, we do so for ISCCP cloud measurements over the SGP site.

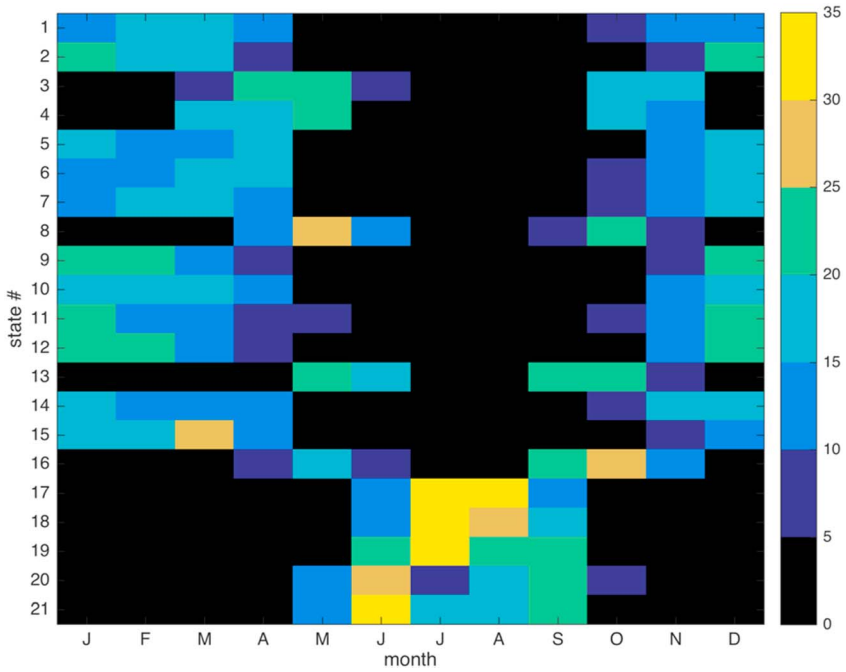

Figure 1. Percentage of each state that occurs during each month of the year. Each row sums to 100.

We can use the observed atmospheric states to classify output from GCM simulations. We sample model output on the same grid as the reanalysis, creating model snapshots for the region. Model snapshots are classified as belonging to one of the observed states by calculating the Euclidian distance from the snapshot to each of the observed state centroids and assigning the snapshot to the closest state. The centroids of each state are defined as the number of standard deviations each state variable is from the observed mean. By using standardized values, all input variables are weighted equally in the sorting, regardless of their units. In order to calculate the distance a model snapshot is from a centroid, the model snapshot must be normalized to make it comparable. We normalize model snapshots by subtracting the observed mean values for each variable and dividing by the observed standard deviations. This creates a description of the model snapshot that is directly comparable to the observed snapshots, that is, a model snapshot that has the same standardized value for a particular variable as an observed snapshot would have for the same raw value for that variable. We also tested other methods of normalizing model snapshots, in order to account for the possible effects of model biases. We removed the model bias at both annual and seasonal time scales before normalizing and also performed a histogram renormalization of the model output but found that none of these more complex methods significantly changed the composite properties of the states. We therefore chose to use the simplest method: using the observed means and standard deviations in normalizing both the observations and model output.

3. Atmospheric States at SGP

The procedure described in section 2 applied to the ERA-Interim reanalysis of the SGP region produces 21 atmospheric states. The states each have an identifiable and coherent meteorology that we discuss here. The states organize themselves by a combination of dynamics and thermodynamics, which produces seasonally dependent occurrences (Figure 1). Eleven of the states occur primarily between October and April and represent wintertime weather patterns. These states are all various stages of the progression of low and high pressure systems that frequently cross the region during winter. Five more states occur during the spring and autumn, primarily April to June and September to November. These states are dynamically quite similar to many of the winter states, but with the warmer temperatures that occur during the transition seasons. The final five states occur almost entirely from June to September and represent summer weather patterns for the region. The summer patterns include anticyclonic days that are clear and hot (States 17 and 18), humid

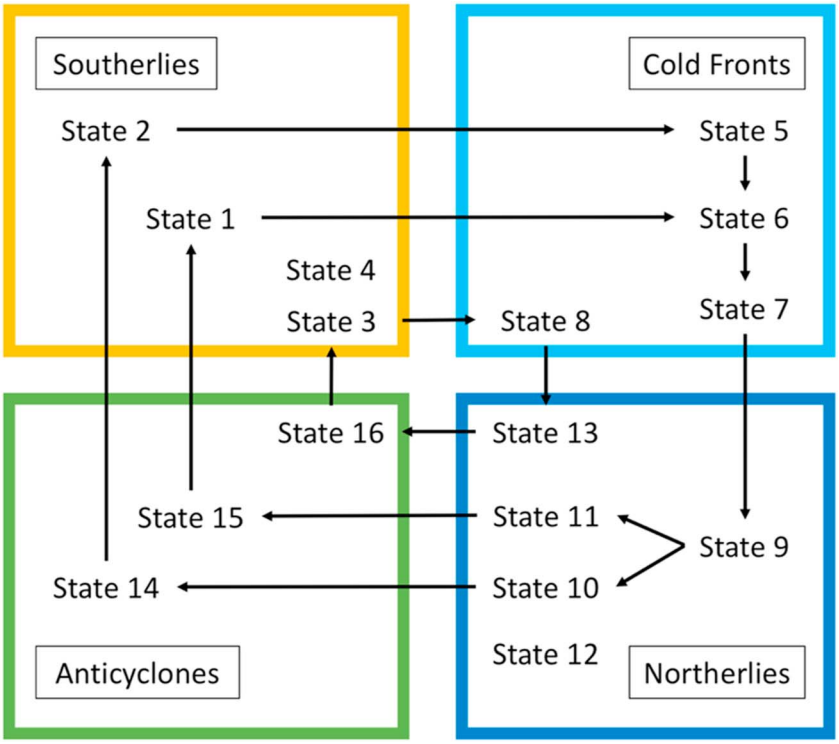

Figure 2. Conceptual flowchart of the progression of atmospheric states. The outer loop is composed of states that occur primarily during winter. The inner loop represents states the occur primarily in spring and fall. Summer states are not shown.

days more prone to deep convection (States 19 and 20), and the occasional weak front (State 21). Further description of these states and the differences between them can be found in the supporting information.

We further organize the nonsummer states by circulation features associated with synoptic forcing. We group the states into categories we call Southerlies, Cold Fronts, Northerlies, and Anticyclones (Figure 2). These categories are based on the near surface winds in the region, as we find this to be a convenient way of separating the atmospheric states into different stages of the synoptic-scale weather systems that predominate in the Great Plains. We also use the time series of state to identify the most frequent predecessor and successor to each state (indicated by the arrows in Figure 2), which we use to number the states such that the numerical progression of states is approximately the sequence of weather patterns the region experiences during the passage of a synoptic event, beginning with warm fronts. Nearly every nonsummer state has a clear location in the sequence, with the exception of State 12 of the Northerlies and State 4 of the Southerlies. These states frequently come both before and after the other states in their respective categories, without any strong preference.

We illustrate the sequential nature of the states by examining States 5, 6, 7, and 9, which depict the passage of a cold front through the region (Figure 3). As described in section 2, many meteorological variables comprise the state definitions; here we choose to show the 875 mb dew point, surface pressure anomaly, and 875 and 500 mb winds because they do an effective job of showing the location of the front. Beginning with State 5, we see a strong northwest-to-southeast gradient in the dew point and a low pressure anomaly to the northwest of the SGP site. The 500 mb winds show the leading edge of an upper-level trough, while the 875 mb winds circulate around the low and show a southwest-to-northeast line of convergence. We interpret this line of convergence as a cold front associated with the surface low and trough. In this state, the front is still slightly to the west of the SGP site. State 6 shows the entire pattern shifted, with both the front and the low now slightly to the east of SGP, indicating that the front has passed over the site. By State 7 the low has moved out the region entirely, with the front now barely within the southeast corner of the domain. We now see cold northerly air advecting into the region, with the upper level winds indicating the trough position. Lastly, in

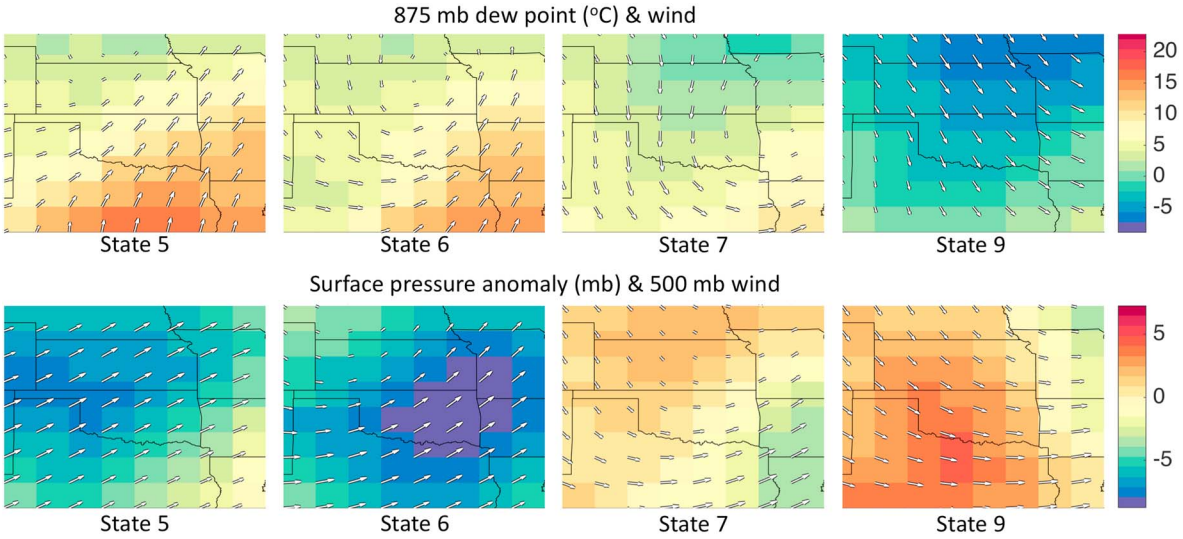

Figure 3. Average meteorological conditions for four states from the SGP classification that often follow each other in sequence. The ARM SGP site is at the center of the domain in Oklahoma. (upper row) 875 mb dew point (°C) and winds; (lower row) surface pressure anomaly (mb) and 500 mb winds. States 5, 6, and 7 show a cold front and low pressure system sweeping across the region, while State 9 shows cold air spilling in from the north as high pressure builds in the wake of the frontal passage. In State 5, the SGP site is in the warm sector, just ahead of the cold front that passes the site in State 6.

State 9, we see a high pressure system move into the region on the leading edge of an upper level ridge, while cold air continues to enter the region from the north. Collectively, this sequence of states shows that the atmospheric states not only have coherent and recognizable meteorology but also relate to each other in sensible and expected ways.

By construction, each state has a statistically distinguishable cloud occurrence profile observed at the SGP site (Figure 4). These cloud profiles help to confirm our interpretations of each state's meteorology. Returning to the sequence of states shown in Figure 3, we start in State 5, which (as shown in Figure 4) displays a deep profile with distinct low and high cloud peaks. This is typical of conditions in the warm sector of approaching cyclones, where the advection of moisture is conducive to cloud formation. Here the bump at approximately 1 km represents a combination of increased low clouds and light precipitation in advance of heavier precipitation to come. Consistent with the passing of the cold front, we see in States 6 and 7 bottom heavy profiles (associated with significant surface precipitation) and a lowering of cloud tops (less cloud above 7 km than in State 5). By the time high pressure arrives with State 9 there is very little cloud occurrence.

The cloud profiles also help to illuminate differences between states in other categories. States 20 and 21, for example, have similar dynamics but very different humidity profiles, which results in the State 20 having cloud approximately twice as often as State 21. In other cases, different cloud profiles represent different dynamics, as in State 1, which has the most cloud occurrence of any state by virtue of containing a stronger warm front over the SGP site than the other Southerly states.

In describing the atmospheric states we have chosen to show a selection of states and variables that make clear the meaning of the atmospheric states and that demonstrate their quality. A more detailed description of the meteorological differences of all 21 states can be found in the supporting information (including a summary Table S1 and Figures S1 and S2 in the supporting information showing the Figure 3 variables for all 21 states).

4. Model Evaluation

One application of the atmospheric states is the examination of errors in GCMs. We evaluate here two experiments created with the GFDL AM3 model (Donner et al., 2011) that differ only in their horizontal resolution. One experiment has a horizontal grid spacing of approximately 2°, while the other has a grid spacing of approximately 0.5°. In both cases, the model has 48 sigma levels and is forced by historical sea surface temperatures for the period 2000–2010, during which snapshots of atmospheric variables are

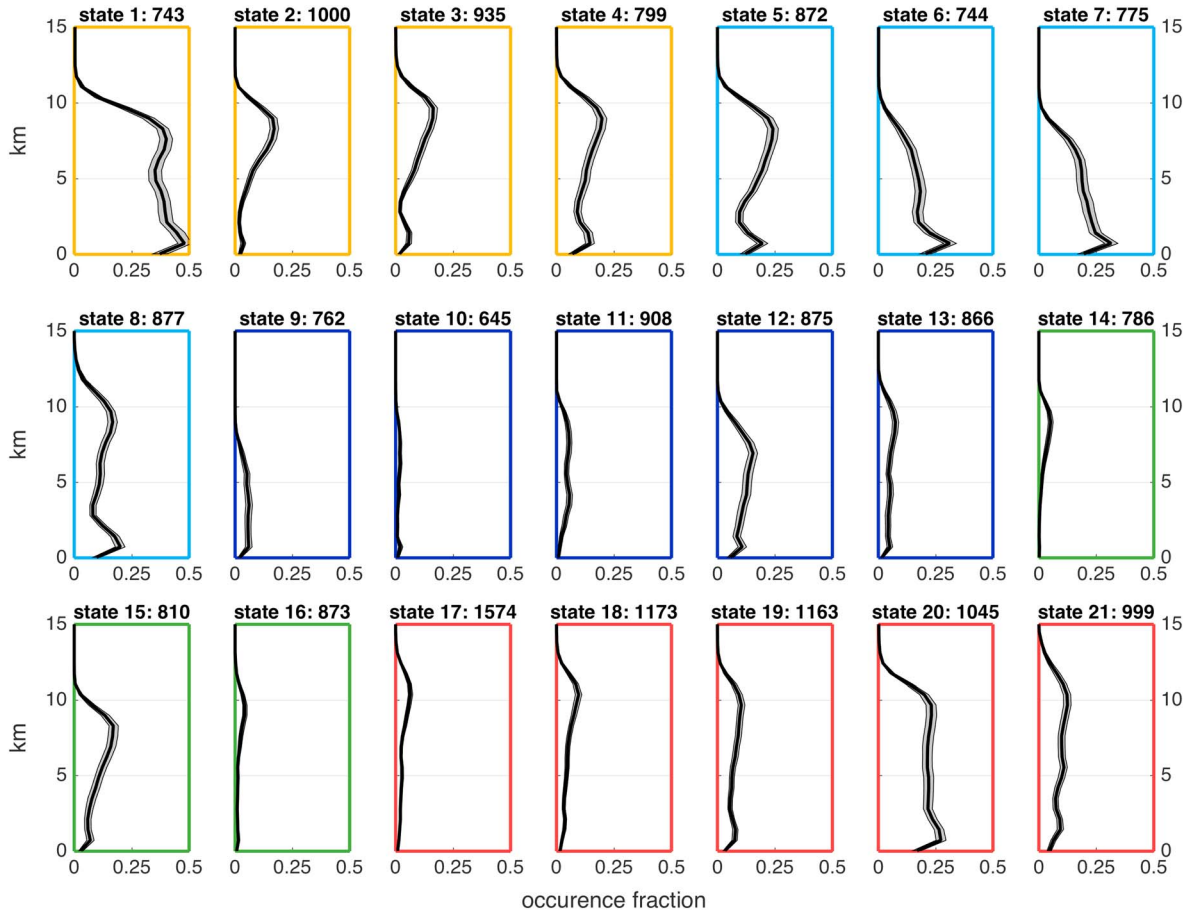

Figure 4. Vertical profiles of cloud occurrence as observed by the vertically pointed millimeter cloud radar at the ARM SGP site for each atmospheric state. Cloud occurrence is defined as the fraction of time a particular altitude has reflectivity greater than −40 dBz. The solid black lines show the mean values, while the grey shaded areas represent the bootstrapped 95% confidence limits. The border color for each panel indicates the group to which each state belongs: Yellow for southerlies, cyan for cold fronts, dark blue for northerlies, green for anticyclones, and red for summer. The number in each panel title is the number of observations of each state.

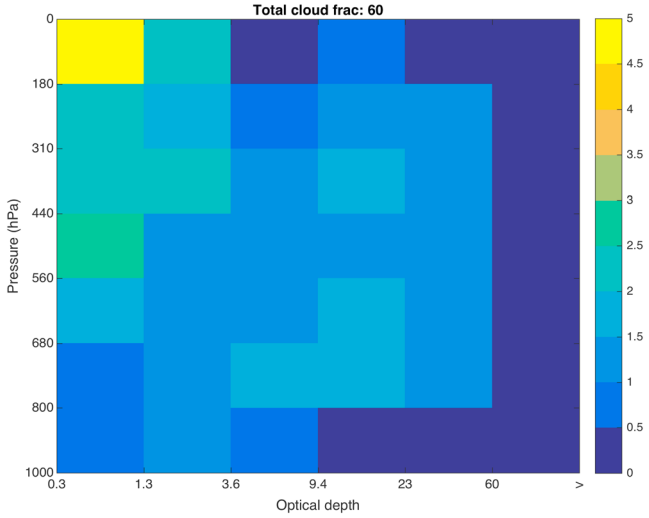

Figure 5. Joint cloud top pressure (vertical axis): optical depth (horizontal axis) histogram of cloud occurrence at the SGP site, as determined by ISCCP. The values are in percentage points, with the sum of the histogram (the total cloud fraction at the site) reported at the top of the figure.

output 4× daily (0, 6, 12, 18Z). Surface pressure, temperature, relative humidity, and horizontal winds are interpolated onto the ERA-Interim grid (1.5° × 1.5°) and the pressure levels at which the reanalysis data were sampled. We normalize the snapshots using the technique described in section 2. This is done in order to make the comparison of model snapshots to observed states a comparison of the same variables at the same locations.

4.1. Evaluating State Cloud Properties at 2° Resolution

We evaluate the within-state cloud properties of the model using data from ISCCP. We use the ISCCP D1 product (Rossow & Schiffer, 1999) that includes joint cloud top pressure—optical depth (τ) histograms of cloud occurrence. The histograms are stored on an approximately 2.5° grid, making them of comparable size to a model grid box. ISCCP joint histograms were generated on a 3-hourly basis (when there is sufficient sunlight for the satellite to make an optical depth retrieval), and we assign each retrieved histogram to the atmospheric state occurring at the time of its observation. We average these joint histograms by state to produce an observed joint histogram of mean cloud occurrence for each atmospheric state. Summing the values of the joint histogram

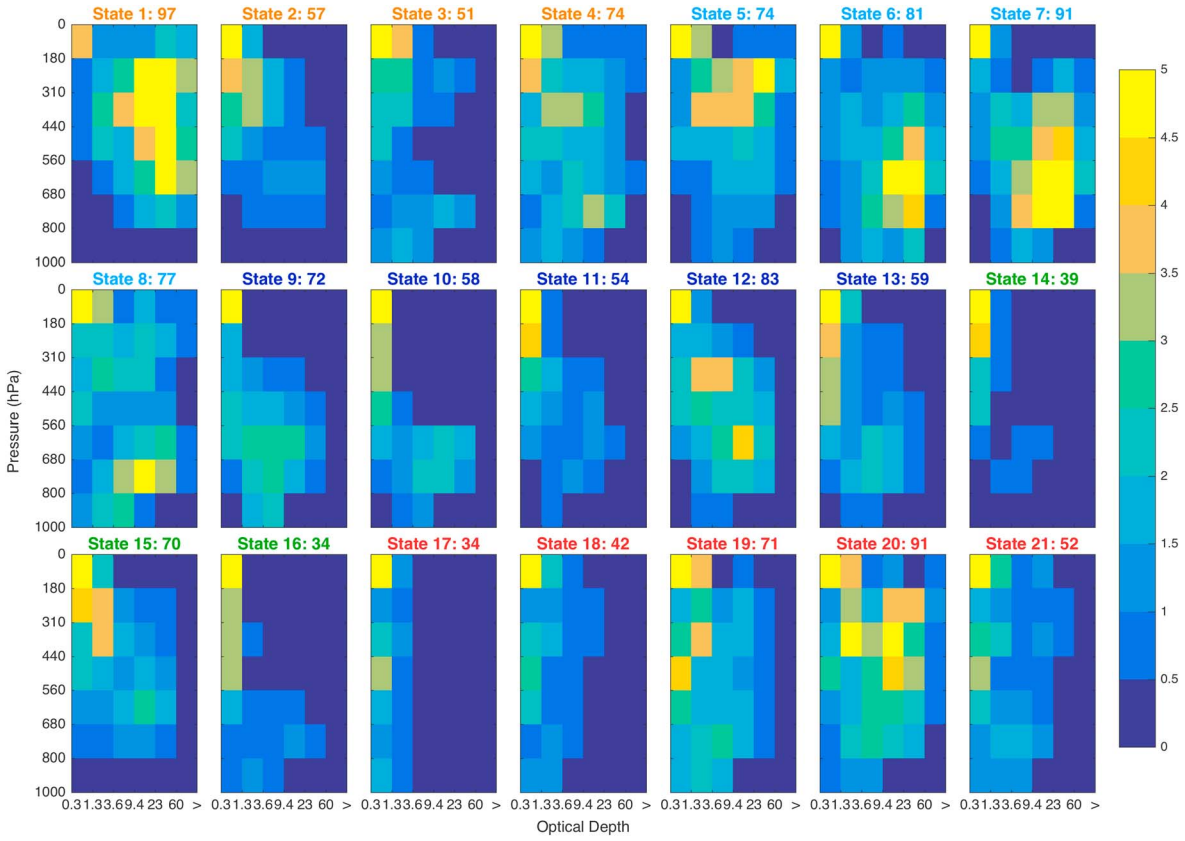

Figure 6. The same as in Figure 5, except for each state. The color of panel title indicates category of state, the same as in Figure 4.

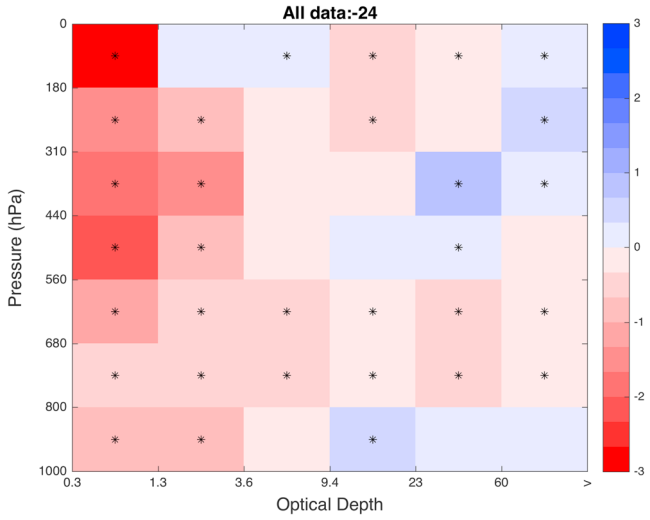

Figure 7. Model bias of cloud occurrence at the SGP site, displayed as a joint histogram of cloud top pressure and optical depth. The values are the difference, in percentage points, between ISCCP simulator output from the 2° run of the AM3 and the observed ISCCP values; positive (blue colors) represent excess cloud in the model. Summing over the bins of the histogram produces the total cloud occurrence bias of the model, −24%, shown at the top of the figure. The asterisks indicate histogram bins whose bias values are significant at 95% confidence, according to a bootstrap resampling test.

provides the total cloud occurrence for the grid box. As shown in Figure 5 the total cloud occurrence observed by ISCCP at the SGP site is 60%, with much of that occurring as high thin cloud (the upper left of the joint histogram). Figure 6 shows that same data composited by atmospheric state. Here the difference in cloud types associated with the states becomes apparent. States with fronts near the ARM site, for example, State 1 (warm front), and States 5, 6, 7, and 8 (cold fronts) show large amounts of thick cloud, consistent with large-scale ascent and deep convection produced by those conditions. The height of the thick cloud varies somewhat between states, with the warm front (State 1), and arriving cold front (State 5) having higher thick cloud and states in which the cold front has already passed the SGP site (States 6 and 7) having somewhat lower thick cloud. In contrast, fair-weather states such as States 14, 16, and 18 (see supporting information) have less cloud overall, most of which is high and thin.

We choose to use ISCCP as the measure of cloud occurrence for the states in order to take advantage of the ISCCP simulator (Bodas-Salcedo et al., 2011) output produced by the model. The simulator estimates the values the satellite would report for the cloud field simulated by the model. In doing so, it creates a model output field that is directly comparable to the observed data, and which attempts to account for some limitations of the satellite product (Marchand et al., 2010). Figure 7 shows the total model bias in cloud occurrence at the SGP site. Overall, the model has a large negative bias in high thin cloud and a

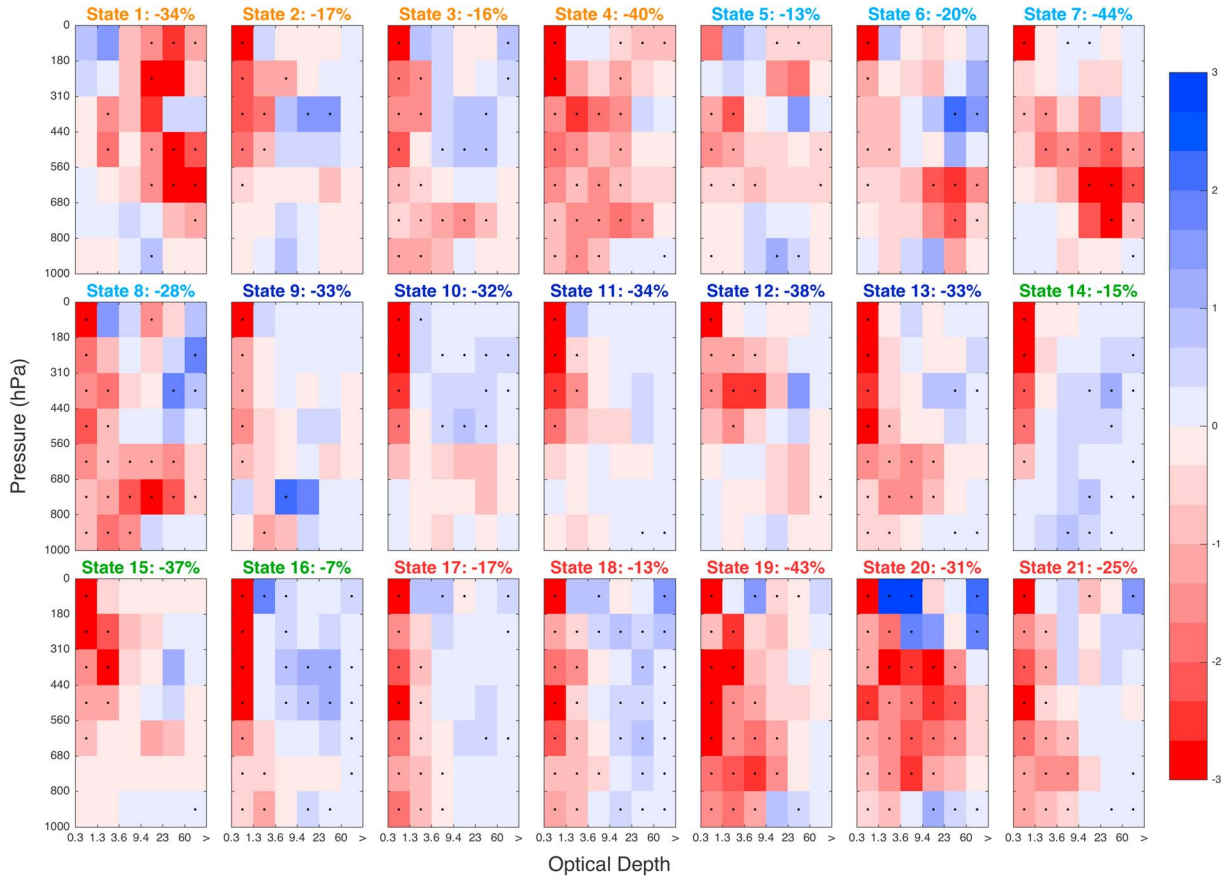

Figure 8. Same as in Figure 7, except for each state. The panel title colors indicate state category, as in Figure 3.

relatively small positive bias in high thick cloud. Low cloud (that with cloud top below 680 hPA) occurs too infrequently in the model, except for the lowest and optically thickest clouds, which occur somewhat more often than they should. The "too few, too bright" problem for low clouds has long been identified as a problem in many GCMs over both land and ocean, with many climate models dramatically underestimating low cloud over the middle latitudes of Eurasia and North America (Klein et al., 2013).

Figure 8 shows that while the lack of high thin clouds is universal across the atmospheric states, some categories of states have larger biases than others. This indicates that the bias is likely due to a variety of problems in the representation of cirrus clouds, an idea we discuss further in section 5. Likewise, the "too few, too bright" problem for low clouds is present in many (though not all) of the states and is particularly prevalent in all of the summer states (17–21) and several of the spring/fall states (4, 8, 13, and 16). Figure 8 also shows that the relatively accurate simulation of thick clouds by the model (Figure 7) is the product of counterbalancing biases among the states. States that have a front within the region (5, 6, 7, 8, and State 1 from the Southerly category, which features a warm front in close proximity to the SGP site) show large negative biases (red colors) in thick cloud. In State 1, the negative bias is due to clouds in all altitudes, while in the frontal states, the reduction is due more clouds below 500 hPa. In States 6 and 8, in particular, there are significant positive biases above 500 hPA, albeit smaller biases than for clouds below 500 hPa, such that there is a net negative bias of total thick cloud amount. This lack of thick cloud can produce substantial shortwave top of atmosphere radiative forcing biases in excess of 40 W/m^2 (Table S2). These are all conditions in which the atmosphere has large-scale ascent at the front. Previous evaluations of climate models have found that models tend to produce clouds that are too optically thick and too high in the atmosphere in frontal conditions, while also producing too little total cloud cover (with larger errors over land than ocean) and with typically too little high-thin, low-level, and midlevel cloud over land (Tselioudis & Jakob, 2002; Gordon et al., 2005; Klein et al., 2013). While this description fits the AM3 model in total (Figure 7) and at least some of the frontal

states (6 and 8), we stress that, for the AM3 model examined here, the total amount of thick cloud (high + low) is too small in all of the frontal states (and cloud radiative forcing is too small; see supporting information).

Fronts are relatively small features compared to a 2° model grid box, making it reasonable to expect the model to struggle to produce a front as sharp as exists in nature. The model cannot resolve subgrid (mesoscale) variability in updrafts and downdrafts and might well have a weak large-scale (grid-box mean) ascent representing neither the stronger updrafts nor compensating subsidence, both limiting the amount of thick cloud in frontal conditions. Naud et al. (2010) found that the GISS GCM did not produce enough high and midlevel cloud during frontal conditions due to weak updrafts in the model and speculated that higher model resolution might improve these errors. Later in this section we test this hypothesis that low model resolution is an important factor in causing a lack of thick cloud in frontal conditions.

In contrast to the frontal conditions, the anticyclonic, northerly, and summer states show an excess of high thick cloud. This excess is most prominent in the drier and higher pressure states within these categories —States 14 and 16 among the anticyclones, States 17 and 18 among the summer states, and State 10 of the northerlies (supporting information). All of these states represent conditions not associated with large-scale ascent, though with some potential for isolated convection to produce high thick cloud. As a result, all of these states have very little high thick cloud in observations (2–9%, Figure 6). That the model produces too much high thick cloud in these conditions (6–13%) suggests that, while convection is not frequent, the deep convective parameterization is triggering too often.

In AM3, deep convection is triggered when the convective available potential energy (CAPE) exceeds a critical threshold of 1,000 J/kg, and the difference in pressure between the level of free convection and the level of neutral buoyancy is at least 500 mb (Benedict et al., 2013; Donner et al., 2011). The CAPE calculation assumes that no entrainment occurs as the parcel rises. This has been shown to produce a relationship between parcel properties and temperature and humidity above the planetary boundary layer that is weaker in the model than in reality (Donner & Phillips, 2003). This may explain how the model can exceed the CAPE threshold as often as it does for these states, despite the relatively dry and stable conditions. Interestingly, this would mean that while the lack of thick cloud in frontal conditions appears to be a result of model resolution, the compensating excess of thick cloud in fair-weather conditions may be the result of the model parameterization. We explore this possibility further in section 4.3.

4.2. Decomposition of Errors

One notable benefit of state-based classification is that it allows the overall bias in a variable of interest to be decomposed into contributions from errors in the frequency of occurrence of states, contributions from errors within each state, and a cross term. To do so, we create metrics for each error source based on the relative frequency of occurrence (RFO) and cloud fraction (CF) of each state *n*:

$$\mathrm{Err\_RFO}_n = \Delta\mathrm{RFO}_n * \left(\mathrm{CF}_{\mathrm{obs},n} - \overline{\mathrm{CF}_{\mathrm{obs}}}\right) \tag{1}$$

$$\mathrm{Err\_within}_n = \mathrm{RFO}_{\mathrm{obs},n} * \Delta\mathrm{CF}_n \tag{2}$$

$$\mathrm{Err\_cross}_n = \Delta\mathrm{RFO}_n * \Delta\mathrm{CF}_n \tag{3}$$

The total change in cloud fraction, that is, the total error, is given by:

$$\mathrm{Err} = \overline{\mathrm{CF}_{\mathrm{model}}} - \overline{\mathrm{CF}_{\mathrm{obs}}} = \sum\nolimits_n \left(\mathrm{Err\_RFO}_n + \mathrm{Err\_within}_n + \mathrm{Err\_cross}_n\right) \tag{4}$$

Here Δ indicates the difference between model and observations, and the overbar, $\overline{\mathrm{CF}}$, is the mean cloud fraction across all states (in the observations or model as indicated by the subscript). Cloud fraction (CF) in these equations can be the *total ISCCP cloud fraction* (sum of all ISSCP histogram bins) or the sum of a subset of ISCCP bins (e.g., the *fraction of ISSCP optically thick cloud*).

These equations are the same as those developed by Williams and Tselioudis (2007) in their state-based analysis, except that (1) we examine cloud fraction rather than cloud radiative forcing and (2) in equation (1), we subtract $\overline{\mathrm{CF}_{\mathrm{obs}}}$. Because the sum of the relative frequency of occurrence across all states must (by definition)

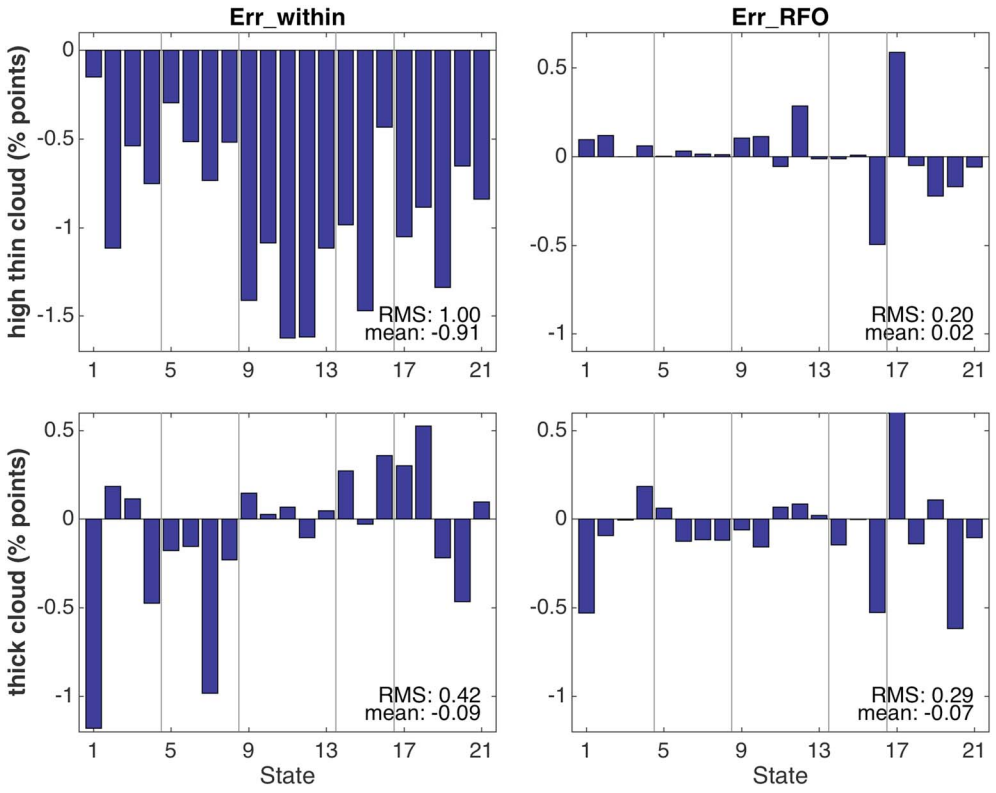

Figure 9. From the 2° run of AM3, state contributions to the bias in (top row) high thin and (bottom row) thick cloud occurrence from (left) within-state cloud errors and (right) relative frequency of occurrence. Sources of error are defined in equations (1) and (2). "High thin" is defined as cloud occurrence in the ISCCP simulator above 560 hPa with less than 9.4 optical depths, and "thick" is defined as all cloud occurrence greater than 9.4 optical depths, respectively. All panels have the same scale size. The values are percentage points of cloud occurrence. The panels are labeled with the RMS and mean values of the plotted error contributions.

be equal to one (i.e., $\Sigma\ \mathrm{RFO}_n = 1$), it is also necessarily true that the sum of changes must be zero (i.e., $\Sigma\ k * \Delta\mathrm{RFO}_n = 0$, where k is any constant). And so one can place any constant in the spot occupied $\overline{\mathrm{CF}_{\mathrm{obs}}}$ in equation (1) and this would not affect the total error, given by equation (4). However, if one wants to compare the magnitude of $\mathrm{Err\_RFO}_n$ and $\mathrm{Err\_within}_n$ for an individual state n, then the value of this constant can matter. Subtracting the mean cloud fraction minimizes the variance of this term and removes this sensitivity. At a practical level, this can be understood because the importance of an error in the frequency of occurrence depends on whether the state in question has an above or below average cloud fraction. For example, imagine that the ith state has a below average cloud fraction $\left(\mathrm{CF}_{\mathrm{obs},i} < \overline{\mathrm{CF}_{\mathrm{obs}}}\right)$. If the model underpredicts the occurrence of the ith state ($\Delta\mathrm{RFO}_i < 0$), this constitutes a positive bias contribution to the total error. Said another way, if the model has too little of a state that has a low-cloud fraction, one would expect the model to be biased high, because the model is (necessarily) spending more time in other states that (on average) have a higher cloud fraction.

In the following analysis we compare the Err_RFO and Err_within for each state in order to determine which source of error dominates for each state and, when summed across all states, which source of error is most important to the overall model error. Err_cross is not dominant for any state, nor is it an important source of overall error, and so for simplicity of presentation we omit it from the following analysis.

To facilitate our decomposition of model bias, we examine two cloud categories: the occurrence of *optically thick clouds* and the occurrence of *high optically thin clouds* by summing over appropriate regions of the ISCCP joint histogram. Summing over regions of the histogram provides a simplified way of interpreting the type of clouds represented by the histogram. For the remainder of the discussion, we refer to high thin clouds as clouds with an optical depth less than 9.4 and cloud tops above 560 mb and thick clouds as

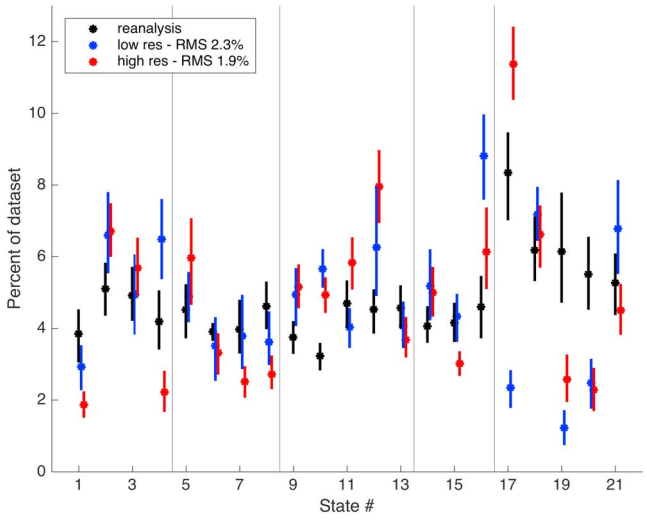

Figure 10. Occurrence rate for each of the 21 atmospheric states in observations (black), the 2° run of AM3 (blue), and the 0.5° run of AM3 (red). The vertical bars indicate the 95% confidence limits from a bootstrap resampling algorithm to estimate sampling uncertainty. The bars have a slight horizontal offset from one another to allow overlap to be visible. The legend indicates the RMS difference between each of the model distributions and the observations.

those with an optical depth larger than 9.4. Clouds in the optically thick portion of the histogram are primarily stratus and physically thick/deep clouds produced by deep convection or fronts, while occurrences in the high thin portion of the joint histogram are primarily cirrus clouds.

The upper panels of Figure 9 show each state's contributions to the model's overall bias in high thin cloud occurrence. It clearly shows that within-state biases are the dominant source of the overall bias. As discussed above, the model has a large negative bias in high thin cloud that is present in all of the states (Figure 8), though here we can see that the bias is somewhat stronger in the Northerly (States 9–13) and Anticyclonic (States 14–16) states. We speculate on possible causes for this in section 5. As all the states have within-state errors of the same sign and roughly similar magnitude, the relative frequency of occurrence of states is not very important to the overall bias. Essentially, if all states produce similar errors, it does not matter which state occurs.

In contrast to this is the total error in thick cloud occurrence (cloud with more than an optical depth of 9.4), shown in the lower panels of Figure 9, that has important contributions from both the distribution of states and the within-state biases. Because some states have positive biases in high thick cloud while other states have negative biases, it is important to the overall bias how frequently each state occurs. The magnitudes of the error contributions from within-state cloud occurrence are generally larger than those from the state distribution (RMS of 0.42 compared to 0.29), but there is substantial cancelation among them due to the differences in thick cloud biases in frontal and fair-weather states described earlier. As a result, the mean error from within state cloud properties (−0.09) is close in magnitude to the error from the distribution (−0.07).

4.3. Evaluation of High-Resolution Experiment

We test the role of model resolution by repeating the analysis from Sections 4.1 and 4.2 on a 0.5° resolution experiment of the AM3. This model has not had its radiative balance retuned from the prior 2° experiment, allowing us to attribute the differences we see to the change in resolution, rather than changes in other model parameters or parameterizations. The 0.5° resolution model remains close to radiative balance, as net radiation at TOA only changes by 1.5 W/m^2. We begin the comparison of the two runs by looking at the frequency of occurrence of the states, as compared to their occurrence in the ERA-Interim data set. Figure 10 shows that the high-resolution run simulates a more accurate distribution of states than the low-resolution run (the frequency of occurrence RMS error shrinks from 2.3 percentage points to 1.9). Notably, States 16, 17, and 19, which were the three states with the largest under- and overpredictions, all have their frequency of occurrence error shrink by large amounts (1.4 to 2.7 percentage points of improvement). This is due to the model producing a more accurate mean state, as the biases in variables such as temperature are generally smaller in the high-resolution run (not shown).

The response of clouds to the increased resolution is more complex. Figure 11 shows the difference in cloud occurrence between the high- and low-resolution runs. The model's ISCCP simulator is run at the resolution of the model grid, so it is possible that some differences may arise as a result of this scale change. Nonetheless, in section 4.1, we hypothesized that the lack of thick cloud in frontal states was likely a result of poor model resolution being unable to simulate fronts well. This is supported by large increases in optically thick cloud in three of the five states discussed in that section, States 1, 6, and 7. Of the other two states, State 8 has a small increase and State 5 has a small decrease. While 0.5° resolution still does not truly resolve fronts, it certainly comes closer to doing so than 2° resolution. A better formed front generates stronger large-scale ascent and thick cloud as a result. More frequent thick cloud in states with fronts is indeed what the high-resolution model produces, and in doing so, the high-resolution run improves the simulation of cloud occurrence in these states. The only state without a front to experience significant increases in thick cloud is State 20, a summertime state representing deep convection across the region. State 20 is by far the cloudiest of the summer states, and we speculate the high-resolution model may be better capturing organized convection.

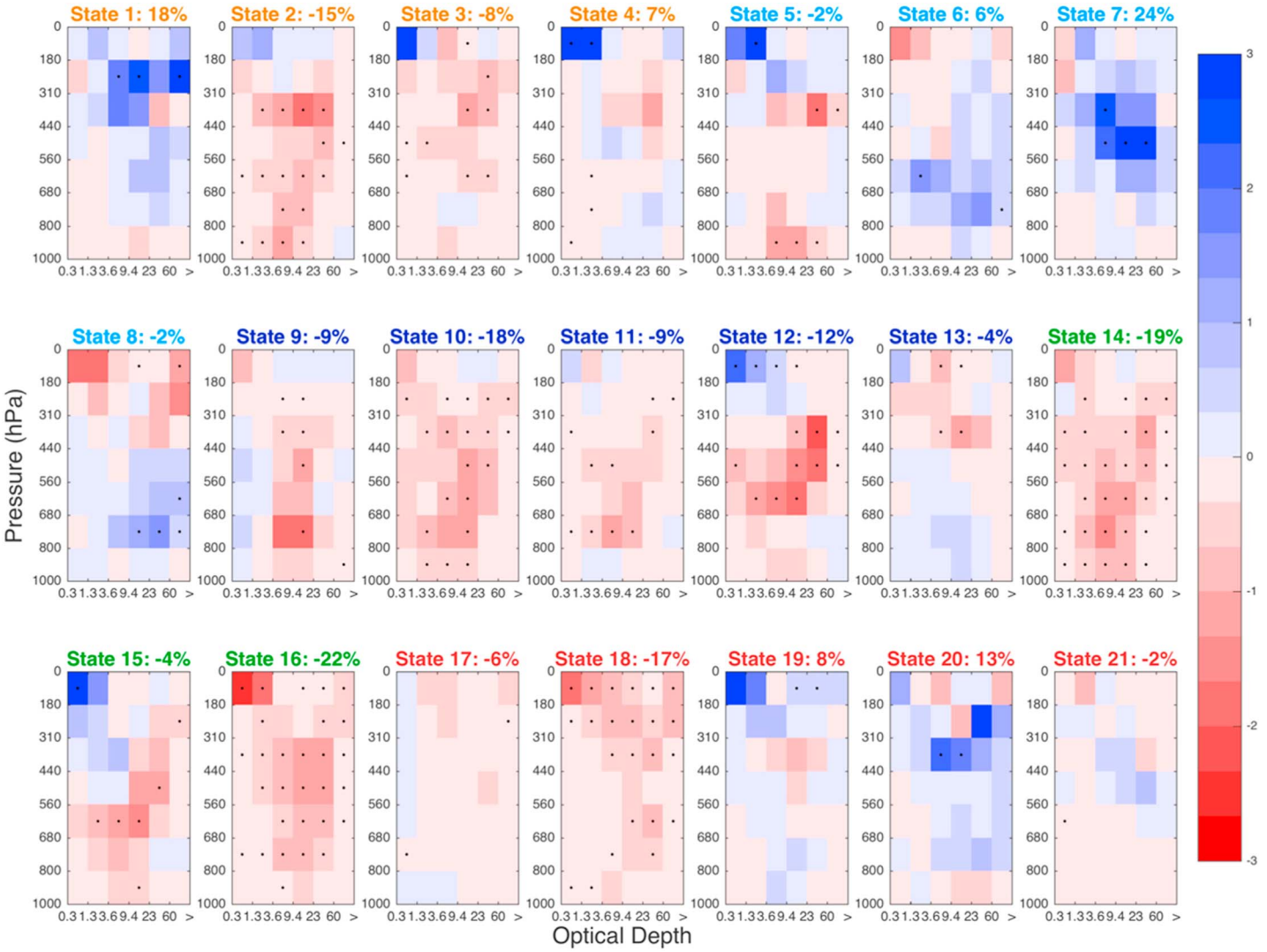

Figure 11. Difference in cloud occurrence, obtained using the ISCCP simulator, between the high- and low-resolution runs of the AM3. The blues indicate greater cloud occurrence in the high-resolution run; the reds indicate less. The asterisks indicate histogram bins whose bias values are significant at 95% confidence, according to a bootstrap resampling test. The sum of each histogram is the total change in cloud occurrence for a given state and is reported in the panel title. The panel title color indicates state category, as in Figure 3.

The warm and dry fair-weather states that had been identified as having excess high thick cloud also experience consistent improvement in the 0.5° run. States 10, 14, 16, 17, and 18 all experience a reduction in high thick cloud. Previously, we hypothesized that the excess high thick cloud was due to the deep convection parameterization triggering too often. This follows from a comparison of CAPE from reanalysis and the model for each atmospheric state, which is shown in Figure 12. We caution here that ERA-Interim reanalysis CAPE is itself a model-based product. Comparison of ERA-Interim CAPE with CAPE based on radiosonde measurements near the ARM SGP site shows considerable scatter with ERA having about 10% fewer events with a CAPE $>$1,000 J/kg and a bias of about $-$200 J/kg, due primarily to errors in the ERA boundary layer temperature and moisture (Gartzke et al., 2017). Nonetheless, we find that at low resolution, high values of CAPE occur too frequently in the AM3 model (by more than 10%) and the model is therefore likely triggering deep convection too often. This is especially true for States 16, 17, and 18, which have a large excess in deep thick cloud. When the model is run at higher resolution, high values of CAPE become rarer in these states, reducing the frequency with which the deep convective parameterization triggers. In conditions with only isolated convection, one would expect to find large areas of subsiding air and small regions of rising air, producing

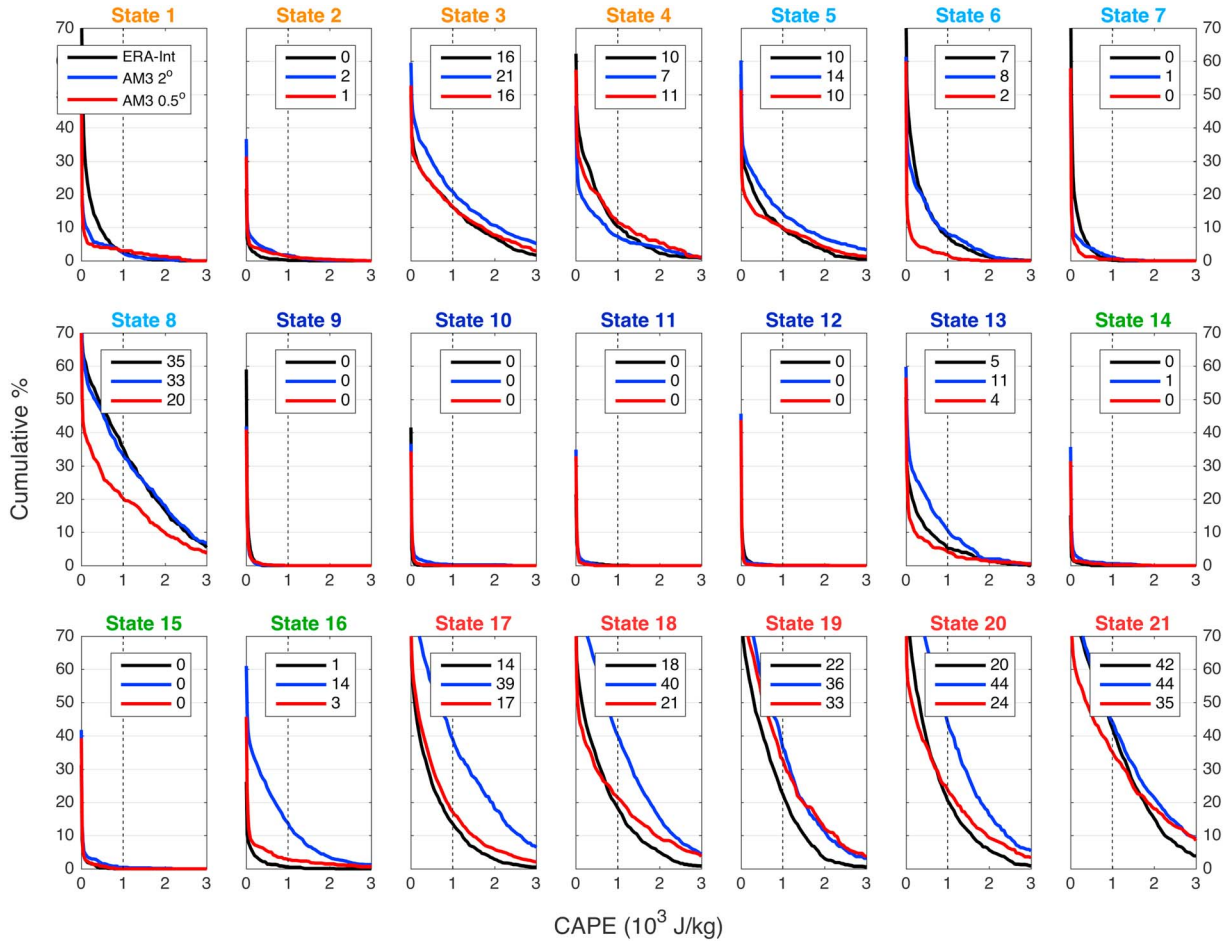

Figure 12. Cumulative distributions of CAPE at the SGP site for each atmospheric state from ERA-Interim (black), the low-resolution AM3 run (blue), and the high-resolution run (red). The dashed vertical black line marks 1,000 J/kg, the cutoff threshold for the deep convective parameterization in the AM3. The values in the legend indicate the percentage of CAPE values greater than this cutoff value. The panel title color indicates state category, as in Figure 3.

a distribution of vertical velocities skewed toward negative velocities with a long tail toward positive velocities. We hypothesize that as the model goes to higher resolution, this distribution becomes more skewed, producing smaller, stronger regions of rising air, that is, regions of high CAPE, and reduces instances of deep convection and high thick cloud in these conditions. As with the frontal states, this improves the simulation of high thick cloud for these states. The same mechanism may be happening for states other than the ones discussed so far, as a reduction in midlevel, midthickness cloud is common to nearly all the nonfrontal states (Figure 11). Unlike the states discussed so far, however, reductions in cloud for these other states, which generally did not have excess of cloud in the low-resolution run, generally make the simulation worse.

The overall improvements to both thick clouds and the frequency of occurrence of states in the high-resolution model have an interesting effect on the overall error from different sources. The lower panels of Figure 13 show the contribution of each state to the total error in high thick cloud in the high-resolution run. The improvements in thick cloud already discussed have substantially reduced the magnitudes of error in most of the individual states (RMS of 0.29 for high resolution compared to 0.42 for low resolution), but they are no longer as effective in compensating for each other, increasing the overall bias (mean of −0.20 at high resolution compared to −0.09 at low resolution). The improved distribution of states has not helped either, actually causing an increase in the error due to frequency of occurrence (mean of −0.17 compared to the previous −0.07). Again, the loss of counterbalancing errors is the cause. As an example, the low-resolution run underpredicted the occurrence of State 17, which has very little high thick cloud in observations, and thus

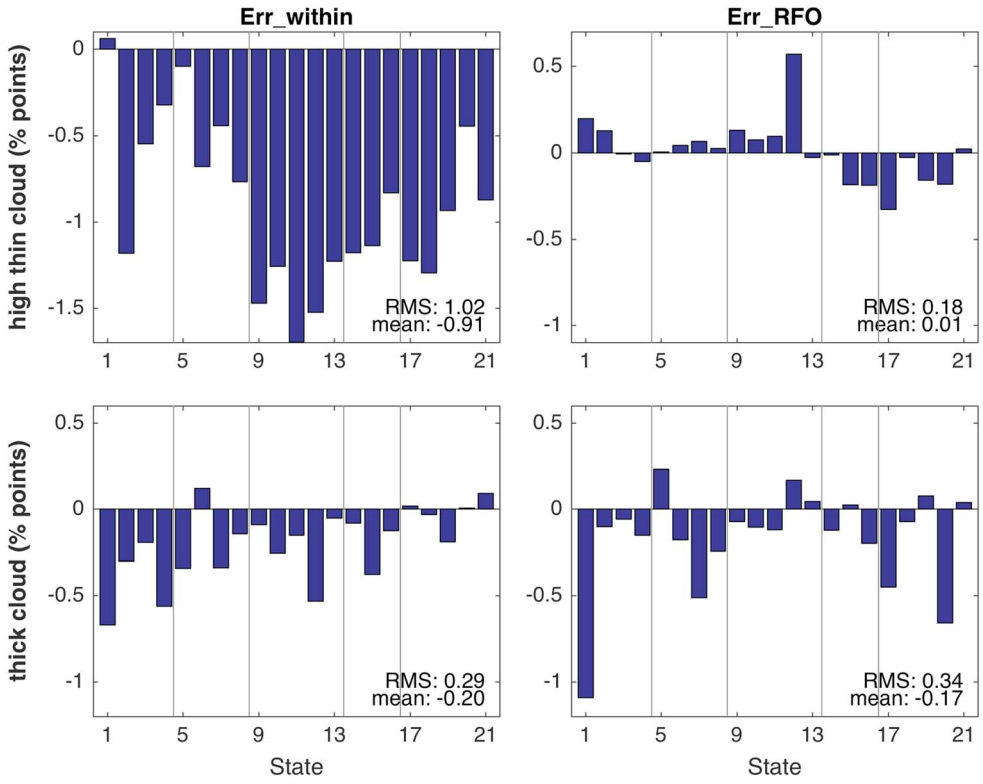

Figure 13. The same as in Figure 9, but for the 0.5° resolution AM3 experiment.

helped to counteract the underprediction of thick cloud from other states. By improving the distribution of states in the high-resolution run, the model has lost this counterbalancing error. As a result, despite an improved distribution of states, and, in most cases, improved state properties, the overall bias of the model has become worse. Figure 13 also shows the contribution to high thin cloud bias, but these are very similar to the results of the low-resolution run. We interpret this to mean that the low bias in cirrus clouds is more likely due to the model microphysics and other factors discussed further in section 5.

5. Summary and Discussion

Our clustering method uses reanalysis data to define sets of atmospheric states for a region surrounding the ARM program site at Southern Great Plains in Oklahoma. Cloud occurrence data from the vertically pointed cloud radar at the sites are used to test the quality of the states defined by dynamic and thermodynamic variables in an iterative process. The number of states and the state definitions are refined until all states have vertical profiles of cloud occurrence that are both temporally stable and distinct from one another. The result is 21 states for the SGP region that are used to understand the meteorology of the region and diagnose the conditions that produce cloud occurrence biases in GFDL's AM3 model.

The SGP states primarily identify different stages of synoptic systems crossing the Great Plains. We use the time series of state to composite ISCCP joint histograms of observed cloud top height and optical depth for each state. We also sort snapshots of model output from two different runs of the AM3 at low and high horizontal resolutions according to these states. Using output from the model's ISCCP simulator, we composite modeled joint histograms of cloud top height and optical depth for the states. Comparing the state joint histograms in observations and in the model allows us to identify which states and thus what types of weather are associated with the model's overall bias in the occurrence of different types of clouds.

We compare the low-resolution run of the model to observations and find that the model is significantly lacking high thin clouds for all atmospheric states, and low thin clouds in most states, while errors in the occurrence of thick clouds varies greatly (including in sign) depending on the atmospheric state. In particular,

states featuring large-scale ascent (i.e., fronts) lack thick cloud in the model (cloud with optical depth greater than 9.4), while states that should feature little convection (i.e., fairweather states) have an excess of thick cloud. We attribute much of the former to the model's difficulty resolving fronts, and the latter to a deep convection parameterization that is triggered too often by excessive values of CAPE, though we acknowledge that convection may contribute significantly in the frontal states, some of which have too much high thick cloud and too little low thick cloud (especially States 6 and 8). In the high-resolution run of the model, the representation of thick clouds is better for both these groups of states; that is, cloud occurrence increases in the frontal states and decreases in the fairweather ones. The former is likely due to improved resolution of atmospheric fronts, while the latter may be due to a better distribution of CAPE. Lastly, we partition the total model error or bias in cloud occurrence into contributions from the distribution of states in the model and contributions from the mean cloud occurrence of the states (within-state error). We find that in the low-resolution run the error in cloud occurrence from individual states (within state errors) contributes approximately twice as much to the total bias as the error due to the distribution of the states, while in the high-resolution run the two are approximately equal sources of bias.

A logical next step after testing the importance of model resolution is to evaluate the model using different convective parameterizations. Benedict et al. (2013) evaluated tropical cloud properties in the AM3 while modifying the deep convective parameterization in a variety of ways. The modifications they implemented included an activation trigger that requires that the time-integrated low-level ascent be sufficient to lift a parcel to the level of free convection. They also replaced the CAPE-relaxation closure with a closure balancing CAPE fluctuations caused by free-tropospheric large-scale processes with CAPE changes due to cumulus convection (Zhang, 2002). They found that compared to a control run, deep convection in the model was strongly suppressed in the simulations with this alternative trigger and closure. Potentially, these changes could improve the excess of thick clouds in fair-weather states in the low-resolution model by making deep convection less frequent. Whether such an improvement would come at the cost of a reduction of thick cloud in frontal states as well (of which the model does not produce enough) is an interesting question; and as noted earlier, some of the frontal states (e.g., 6 and 8) do produce too much high thick cloud with too little low thick cloud. Frontal states should have the low-level ascent necessary to trigger the convection, so it may be that they would not experience the suppression of deep convection that other states would. Benedict et al. (2013) noted that the modified deep convective parameterizations also degrade the mean state of the model. As a result, it may be that a model experiment using the Zhang (2002) might create a situation where the frequency of occurrence of states becomes more important than the within-state errors.

The finding that high thin clouds are lacking in the model under all circumstances requires more analysis to fully understand. High thin clouds can be generated or maintained in a number of ways in the AM3, making it challenging to identify the cause of the bias in their occurrence with the model runs we have analyzed here. For example, advection, stratiform lifting, shallow convection, and deep convection (through detrainment of condensate into the mesoscale updraft), and parameterizations of particle microphysics including sedimentation can all influence high thin cloud amounts in a grid box. The effects of fine scale motions (e.g., gravity waves) that remain unresolved at 0.5 degree resolution might also be involved. The relative importance of each of these sources or sinks of high thin cloud in different atmospheric conditions may help explain why some states are missing much more thin cloud than others. For example, the two groups of states that have the greatest lack of high thin cloud in the model, the northerlies and anticyclones, are also the two groups that have the least high thick cloud in the model. High thick cloud is primarily created by the deep convective parameterization, which transfers 90% of nonprecipitated condensate into anvil clouds (Donner et al., 2011). The fact that the largest biases occur when this process is rare suggests that without detrainment from deep convection, it may be difficult for the model to have enough upper level moisture to sustain high thin clouds. This could be tested by repeating this analysis on a series of runs that alter the condensate partitioning in the deep convection parameterization. Alternately, these states are also ones that occur during upper-level ridges, when advection of thin cloud should be occurring. The lack of high thin cloud in these states may thus indicate that cirrus lifetime in the model is too short, preventing it from being advected into the region. Again, this is testable through the manipulation of model parameters, such as ice fall speed.

This study raises an interesting question of what it means for one model configuration to represent clouds better or worse than another one. The ISCCP observations for the SGP site have a total cloud occurrence of 60%, the low-resolution run of the AM3 has 34%, and the high-resolution has only 28%. In this sense,

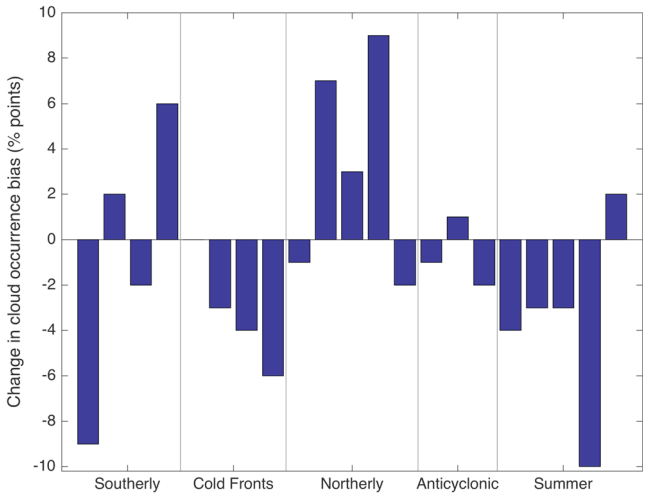

Figure 14. Change in the magnitude of the total cloud occurrence bias for each state in the 0.5° experiment compared to the 2° experiment. The negative values indicate improvement; the positive values indicate worsening.

increasing the resolution of the model has degraded the simulation of clouds at the SGP site. However, as shown in Figure 14, which shows the change in magnitude of the total cloud occurrence bias for each state, cloud occurrence for most atmospheric states improves at higher resolution, especially for the frontal, anticyclonic, and summer states. Furthermore, the frequency of occurrence of states improves at high resolution as well (Figure 10). It is only through changes in the distribution of states and the removal of compensating errors that the total cloud occurrence becomes worse. This makes it a question of priorities whether increasing model resolution improves the representation of clouds in AM3. A similar problem arises in studies of the effect of different parameterizations on model performance. For example, it is a common feature of GCMs for improvements in tropical intraseasonal variability to come at the cost of larger biases in the mean state and that as a result, most GCMs have weak intraseasonal variability (Kim et al., 2011).

That accuracy in the simulation of atmospheric processes has been sacrificed to improve the mean state is understandable, as most GCMs are designed for long-term climate change scenarios, for which an accurate mean state is more important. Toward this purpose, one might be more interested in the radiative effects of the clouds than the clouds themselves. While a detailed examination of the radiative forcing that arises from the cloud biases identified here is beyond the scope of this study, we have included Table S2, which details the shortwave and longwave cloud radiative forcing for each state, as well as the model biases. As would be expected from the counterbalancing cloud biases, there is substantial cancelation of radiative biases between states. For GCM studies designed to test and understand particular processes, these state biases may be important, and nonstandard GCM configurations may be a better choice. Alternately, it is possible that retuning the high-resolution model may prevent degradation of the mean state while preserving the benefits to the individual processes.

The ability of atmospheric classification to discover and characterize compensating errors in a GCM run is potentially a valuable tool in parameterization development. Most directly, it can help to identify the physical situations that are most problematic and require further attention to represent accurately. More broadly, classification can help to evaluate whether a particular parameterization or change to the model configuration is an improvement or not. If the evaluation metric for a particular model run is how well it reproduces the mean state, it would be easy for actual improvements to a parameterization to be wrongly discarded when the mean state fails to improve. Continuing to pursue these advances in parameterization may produce eventual improvements in the mean state that might not be realizable without making it worse first. Using process-oriented metrics to evaluate a GCM, such as how well it performs under particular conditions, in addition to the mean state is one way to identify what otherwise might be overlooked improvements. Given the challenge of producing an accurate parameterization, following such a lead may eventually be very important.

Acknowledgments
We are grateful to Angel Adames-Corraliza and Paul Ginoux for providing helpful commentary on the manuscript. Thank you to the many Atmospheric Radiation Measurement (ARM) instrument mentors and data archivists for all their efforts in producing and maintaining the quality of the ARM data sets, as well as to the ISCCP project. This research has been supported by the Princeton Environmental Institute at Princeton University through the Carbon Mitigation Initiative and the Department of Energy (DOE) Atmospheric Science Research (ASR) program (grant DE-SC0002472). Work at LLNL was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Reanalysis products and cloud observations are available from the ECMWF, ARM, and ISCCP websites. State definitions are available at https://atmos.washington.edu/~roj/nobackup/Classification_Results/SGP/. AM3 data products associated with this paper can be obtained by contacting the first author, Stuart Evans.

References

Ackerman, T. P., & Stokes, G. M. (2003). The Atmospheric Radiation Measurement Program. *Physics Today*, *56*(1), 38–44. https://doi.org/10.1063/1.1554135

Benedict, J. J., Maloney, E. D., Sobel, A. H., Frierson, D. M., & Donner, L. J. (2013). Tropical intraseasonal variability in version 3 of the GFDL atmosphere model. *Journal of Climate*, *26*(2), 426–449. https://doi.org/10.1175/JCLI-D-12-00103.1

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J.-L., Klein, S. A., … John, V. O. (2011). COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, *92*(8), 1023–1043. https://doi.org/10.1175/2011BAMS2856.1

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*, L20806. https://doi.org/10.1029/2005GL023851

Caine, S., Jakob, C., Siems, S., & May, P. (2009). Objective classification of precipitating convective regimes using a weather radar in Darwin, Australia. *Monthly Weather Review*, *137*(5), 1585–1600. https://doi.org/10.1175/2008MWR2532.1

Clothiaux, E. E., Ackerman, T. P., Mace, G. G., Moran, K. P., Marchand, R. T., Miller, M. A., & Martner, B. E. (2000). Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites. *Journal of Applied Meteorology*, *39*(5), 645–665. https://doi.org/10.1175/1520-0450(2000)039%3C0645:ODOCHA%3E2.0.CO;2

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Soli, P., Kobayashi, S., … Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Donner, L. J., & Phillips, V. T. (2003). Boundary layer control on convective available potential energy: Implications for cumulus parameterization. *Journal of Geophysical Research*, *108*(D22), 4701. https://doi.org/10.1029/2003JD003773

Donner, L. J., Wyman, B. L., Hemler, R. S., Horowitz, L. W., Ming, Y., Zhao, M., … Zeng, F. (2011). The dynamical core, physical parameterizations, and basic simulation characteristics of the atmospheric component AM3 of the GFDL global coupled model CM3. *Journal of Climate*, *24*(13), 3484–3519. https://doi.org/10.1175/2011JCLI3955.1

Evans, S. M., Marchand, R. T., Ackerman, T. P., & Beagley, N. (2012). Identification and analysis of atmospheric states and associated cloud properties for Darwin, Australia. *Journal of Geophysical Research*, *117*, D06204. https://doi.org/10.1029/2011JD017010

Evans, S. M., Marchand, R. T., & Ackerman, T. P. (2014). Variability of the Australian monsoon and precipitation trends at Darwin. *Journal of Climate*, *27*(22), 8487–8500. https://doi.org/10.1175/JCLI-D-13-00422.1

Fereday, D. R., Knight, J. R., Scalfe, A. A., & Folland, C. K. (2008). Cluster analysis of North Atlantic-European circulation types and links with tropical Pacific sea surface temperatures. *Journal of Climate*, *21*(15), 3687–3703. https://doi.org/10.1175/2007JCLI1875.1

Gartzke, J., Knuteson, R., Przybyl, G., Ackerman, S., & Revercomb, H. (2017). Comparison of satellite-, model-, and radiosonde-derived convective available potential energy in the southern Great Plains region. *Journal of Applied Meteorology and Climatology*, *56*(5), 1499–1513. https://doi.org/10.1175/JAMC-D-16-0267.1

Gordon, N. D., Norris, J. R., Weaver, C. P., & Klein, S. A. (2005). Cluster analysis of cloud regimes and characteristic dynamics of midlatitude synoptic systems in observations and a model. *Journal of Geophysical Research*, *110*, D15S17. https://doi.org/10.1029/2004JD005027

Jakob, C. (2010). Accelerating progress in global atmospheric model development through improved parameterizations: Challenges, opportunities, and strategies. *Bulletin of the American Meteorological Society*, *91*(7), 869–876. https://doi.org/10.1175/2009BAMS2898.1

Jakob, C., & Schumacher, C. (2007). Precipitation and latent heating characteristics of the major tropical western Pacific cloud regimes. *Journal of Climate*, *21*, 4348–4364.

Jakob, C., Tselioudas, G., & Hume, T. (2005). The radiative, cloud, and thermodynamic properties of the major tropical western Pacific cloud regimes. *Journal of Climate*, *18*(8), 1203–1215. https://doi.org/10.1175/JCLI3326.1

Jiang, P., Yu, Z., Gautam, M., Yuan, F., & Acharya, K. (2016). Changes of storm properties in the United States: Observations and multimodel ensemble projections. *Global and Planetary Change*, *142*, 41–52. https://doi.org/10.1016/j.gloplacha.2016.05.001

Jin, D., Oreopoulos, L., & Lee, D. (2017). Regime-based evaluation of cloudiness in CMIP5 models. *Climate Dynamics*, *48*(1-2), 89–112. https://doi.org/10.1007/s00382-016-3064-0

Kennedy, A. D., Dong, X., & Xi, B. (2016). Cloud fraction at the ARM SGP site: Reducing uncertainty with self-organizing maps. *Theoretical and Applied Climatology*, *124*(1-2), 43–54. https://doi.org/10.1007/s00704-015-1384-3

Klein, S. A., Zhang, Y., Zelinka, M. D., Pincus, R., Boyle, J., & Glecker, P. J. (2013). Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator. *Journal of Geophysical Research: Atmospheres*, *118*, 1329–1342. https://doi.org/10.1002/jgrd.50141

Kim, D., Sobel, A. H., Maloney, E. D., Frierson, D. M. W., & Kang, I.-S. (2011). A systematic relationship between intraseasonal variability and mean state bias in AGCM simulations. *Journal of Climate*, *24*(21), 5506–5520. https://doi.org/10.1175/2011JCLI4177.1

Marchand, R., Beagley, N., Thompson, S. E., Ackerman, T., & Schultz, D. (2006). A bootstrap technique for testing the relationship between local-scale radar observations of cloud occurrence and large-scale atmospheric fields. *Journal of the Atmospheric Sciences*, *63*(11), 2813–2830. https://doi.org/10.1175/JAS3772.1

Marchand, R., Beagley, N., & Ackerman, T. (2009). Evaluation of hydrometeor occurrence profiles in the multiscale modeling framework climate model using atmospheric classification. *Journal of Climate*, *22*(17), 4557–4573. https://doi.org/10.1175/2009JCLI2638.1

Marchand, R., Ackerman, T., Smyth, M., & Rossow, W. B. (2010). A review of cloud top height and optical depth histograms from MISR, ISCCP, and MODIS. *Journal of Geophysical Research*, *115*, D16206. https://doi.org/10.1029/2009JD013422

Mason, S., Jakob, C., Protat, A., & Delanoë, J. (2014). Characterizing observed midtopped cloud regimes associated with Southern Ocean shortwave radiation biases. *Journal of Climate*, *27*(16), 6189–6203. https://doi.org/10.1175/JCLI-D-14-00139.1

Mason, S., Fletcher, J. K., Haynes, J. M., Franklin, C., Protat, A., & Jakob, C. (2015). A hybrid cloud regimes methodology used to evaluate southern ocean cloud and shortwave radiation errors in ACCESS. *Journal of Climate*, *28*(15), 6001–6018. https://doi.org/10.1175/JCLI-D-14-00846.1

Muhlbauer, A., Ackerman, T. P., Comstock, J. M., Diskin, G. S., Evans, S. M., Lawson, R. P., & Marchand, R. T. (2014). Impact of large-scale dynamics on the microphysical properties of midlatitude cirrus. *Journal of Geophysical Research: Atmospheres*, *119*, 3976–3996. https://doi.org/10.1002/2013JD020035

Naud, C. M., Del Genio, A. D., Bauer, M., & Kovari, W. (2010). Cloud vertical distribution across warm and cold fronts in CloudSat-CALIPSO data and a general circulation model. *Journal of Climate*, *23*(12), 3397–3415. https://doi.org/10.1175/2010JCLI3282.1

Oreopoulos, L., Cho, N., Lee, D., Kato, S., & Huffman, G. J. (2014). An examination of the nature of global MODIS cloud regimes. *Journal of Geophysical Research*, *119*, 8362–8383. https://doi.org/10.1002/2013JD021409

Pope, M., Jakob, C., & Reeder, M. J. (2009). Regimes of the north Australian wet season. *Journal of Climate*, *22*(24), 6699–6715. https://doi.org/10.1175/2009JCLI3057.1

Rossow, W. B., & Schiffer, R. A. (1999). Advances in understanding clouds from ISCCP. *Bulletin of the American Meteorological Society*, *80*(11), 2261–2287. https://doi.org/10.1175/1520-0477(1999)080%3C2261:AIUCFI%3E2.0.CO;2

Shafer, M., Ojima, D., Antle, J. M., Kluck, D., McPherson, R. A., Petersen, S., … Sherman, K. (2014). Ch. 19: Great Plains. In J. M. Melillo, T. Richmond, & G. W. Yohe (Eds.), *Climate change impacts in the United States: The Third National Climate Assessment* (pp. 441–461). Washington, DC: U.S. Global Change Research Program.

Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research*, *47*(1), 123–138. https://doi.org/10.3354/cr00953

Tselioudis, G., & Jakob, C. (2002). Evaluation of midlatitude cloud properties in a weather and a climate model: Dependence on dynamic regime and spatial resolution. *Journal of Geophysical Research*, *107*(D24), 4781. https://doi.org/10.1029/2002JD002259

Williams, K. D., & Tselioudis, G. (2007). GCM intercomparison of global cloud regimes: Present-day evaluation and climate change response. *Climate Dynamics*, *29*(2-3), 231–250. https://doi.org/10.1007/s00382-007-0232-2

Williams, K. D., & Webb, M. J. (2008). A quantitative climate performance assessment of cloud regimes in GCMs. *Climate Dynamics*, *29*, 231–250.

Zelinka, M. D., Klein, S. A., & Hartmann, D. L. (2012). Computing and partitioning cloud feedbacks using cloud property histograms. Part 1: Cloud radiative kernels. *Journal of Climate*, *25*(11), 3715–3735. https://doi.org/10.1175/JCLI-D-11-00248.1

Zhang, G. J. (2002). Convective quasi-equilibrium in midlatitude continental environment and its effect on convective parameterization. *Journal of Geophysical Research*, *107*(D14), 4220. https://doi.org/10.1029/2001JD001005