Title

Microphysical Sensitivity of Superparameterized Precipitation Extremes in the Contiguous United States Due to Feedbacks on Large-Scale Circulation

Journal

Earth and Space Science, 7(7)

ISSN

2333-5084

Authors

Charn, Alexander B
Collins, William D
Parishani, Hossein
et al.

Publication Date

2020-07-01

DOI

10.1029/2019ea000731

Peer reviewed

Earth and Space Science

RESEARCH ARTICLE
10.1029/2019EA000731

Key Points:
- The choice of superparameterized Community Atmosphere Model (SPCAM) microphysics parameterization impacts the statistics of precipitation extremes through large-scale circulation feedbacks
- Changes in precipitation extremes from present day to future, warmer climates are insensitive to microphysics parameterization in SPCAM
- Regardless of the microphysics scheme, the precipitation extremes simulated with SPCAM underestimate those observed via rain gauges

Supporting Information:
- Supporting Information S1

Correspondence to:
A. B. Charn,
alexcharn5@berkeley.edu

Citation:
Charn, A. B., Collins, W. D., Parishani, H., Risser, M. D., & O'Brien, T. A. (2020). Microphysical sensitivity of superparameterized precipitation extremes in the contiguous United States due to feedbacks on large-scale circulation. *Earth and Space Science. 7*, e2019EA000731. https://doi.org/10.1029/2019EA000731

Received 28 MAY 2019
Accepted 1 MAR 2020
Accepted article online 4 APR 2020

Microphysical Sensitivity of Superparameterized Precipitation Extremes in the Contiguous United States Due to Feedbacks on Large-Scale Circulation

Alexander B. Charn[1,2], William D. Collins[1,2], Hossein Parishani[3], Mark D. Risser[2], and Travis A. O'Brien[4]

[1]Department of Earth and Planetary Science, University of California, Berkeley, CA, USA, [2]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA, [3]Department of Earth System Science, University of California, Irvine, CA, USA, [4]Department of Earth and Atmospheric Sciences,Indiana University, Bloomington, IN, USA

Abstract Superparameterized (SP) global climate models have been shown to better simulate various features of precipitation relative to conventional models, including its diurnal cycle as well as its extremes. While various studies have focused on the effect of differing microphysics parameterizations on precipitation within limited-area cloud-resolving models, we examine here the effect on contiguous U.S. (CONUS) extremes in a global SP model. We vary the number of predicted moments for hydrometeor distributions, the character of the rimed ice species, and the representation of raindrop self-collection and breakup. Using a likelihood ratio test and accounting for the effects of multiple hypothesis testing, we find that there are some regional differences, particularly during spring and summer in the Southwest and the Midwest, in both the current climate and a warmer climate with uniformly increased sea surface temperatures. These differences are most statistically significant and widespread when the number of moments is changed. To determine whether these results are due to (fast) local effects of the different microphysics or the (slower) ensuing feedback on the large-scale atmospheric circulation, we run a series of short, 5-day simulations initialized from reanalysis data. We find that the differences largely disappear in these runs and therefore infer that the different parameterizations impact precipitation extremes indirectly via the large-scale circulation. Finally, we compare the present-day results with hourly rain gauge data and find that SP underestimates extremes relative to observations regardless of which microphysics scheme is used given a fixed model configuration and resolution.

1. Introduction

Impacts from climate extremes such as heat waves, droughts, and heavy storms include geographic range shifts of plants and animals, food price increases, damage to infrastructure, and increased mortality (IPCC, 2014). In particular, rainfall extremes directly impact the frequency and the severity of floods and the runoff in rivers, which provide over half of the global potable water supply (Barnett et al., 2005). Many global climate models (GCMs) are able to successfully reproduce mean patterns on large scales, but they generally exhibit lower fidelity in simulating regional features and higher order statistics (variance, extremes, etc.) (Flato et al., 2013). One reason is that critical convective processes that generate clouds and precipitation must be parameterized since the characteristic length scales of these processes are much smaller than a conventional GCM grid cell. Typical GCMs often diagnose convection by assuming a quasi-steady equilibrium (e.g., Zhang & McFarlane, 1995), an approximation that is notorious for underestimating higher intensity updrafts and, hence, extreme rainfall (Dai, 2006; Wilcox & Donner, 2007). While the employment of a global cloud-resolving model with horizontal resolution of O(1 km) provides an obvious solution, the disadvantage of this approach is the increased computational cost by a factor of a million.

One solution to this problem that balances resolution and computational cost is the creation and utilization of a superparameterized (SP) model such as the SP version of the Community Atmosphere Model (CAM), known as SPCAM (Khairoutdinov & Randall, 2001). Superparameterization replaces the convective and boundary layer schemes in each GCM grid column with a cloud-resolving model (CRM). Depending on its

resolution, the CRM can resolve deep convective and mesoscale processes in response to large-scale GCM dynamics, and in return it provides subgrid convective heating and moistening tendencies to the larger grid. Hence, cloud-scale interactions between cloud dynamics, microphysics, radiation, and turbulence are more finely resolved. Various studies have documented improved correlation with observations, for example, of the Madden-Julian oscillation (Benedict & Randall, 2009), African easterly waves (McCrary et al., 2014), and the diurnal precipitation cycle (Khairoutdinov et al., 2005). In the case of rainfall extremes, Li et al. (2012) studied the tails of the frequency distribution over the United States and noted that they are much better represented when using SPCAM as opposed to CAM. Globally, Kooperman et al. (2016) found that SPCAM has lower systematic biases relative to satellite-derived products since it simulates higher extreme rain rates, whereas CAM underestimates the heaviest rain rates at coarser (∼2°) resolutions. SPCAM also has a higher, more realistic amount mode, that is, the rain rate that delivers the most accumulated rainfall, at all resolutions. Conversely, Kooperman et al. (2018) showed that the amount mode in CAM is underestimated from 50°S to50°N due to the component from the convective parameterization, which dominates precipitation amounts in low to midlatitudes.

While the structural uncertainties associated with the representation of deep convection have been examined by comparing simulations from CAM to those from SPCAM, less attention has been paid to the parametric uncertainty that arises due to the representation of cloud microphysics in global models. In limited-domain CRM simulations, which are typically run with nonperiodic boundary conditions forced by reanalyses, differing microphysics parameterizations have been shown to impact precipitation structures and statistics. For example, Bryan and Morrison (2012) showed in their case study simulation of a squall line that using hail instead of graupel as the rimed ice species led to faster fall velocities and slower propagation speeds and hence to increased precipitation accumulation. Van Weverberg et al. (2012) also simulated a squall line and found that using hail led to a higher domain-maximum accumulated precipitation but also faster propagation speeds due to more intense cold pools. However, the largest factor contributing to the domain-maximum accumulation in their study is the combined treatment of raindrop self-collection and collisional breakup. Specifically, the representation involves a size threshold that, when exceeded by the number-weighted drop diameter, causes the collection efficiency to decrease exponentially from 1 in order to simulate the effects of increased breakup. We conduct a similar test involving this process, which is described in greater detail in section 2.1.1. Finally, a radiative convective equilibrium experiment (Singh & O'Gorman, 2014) using doubly periodic boundary conditions yielded instantaneous rainfall rates that varied by an order of magnitude as hydrometeor fall speeds fixed to constant values were varied by an order of magnitude.

To the best of our knowledge, one study to date has examined the effects of different parameterizations of microphysics within the SPCAM framework. Elliott et al. (2016) investigated mesoscale convective systems (MCSs) and found that sensitivities in MCS event counts and in precipitation rates were overshadowed by interannual variability. Recently, Parishani et al. (2017) introduced ultraparameterized CAM, in which the embedded CRMs have fine enough resolution to explicitly capture boundary layer turbulence. While not focused on precipitation, Parishani et al. (2018) compared SPCAM and ultraparameterized CAM and found that microphysics, rather than the CRM grid resolution, was mostly responsible for differences in cloud feedback to +4 K sea surface temperature (SST) warming. They attributed this finding to differences in liquid and ice water path in the baseline climatology.

In this study we generalize our analysis (as compared to Elliott et al., 2016) to precipitation rates. We apply extreme value theory to see whether disparate microphysics representations within SPCAM significantly alter the distribution tail, and if so, which representation best matches observations. This question has relevance for the spread in current and future projections of rainfall extremes. As a secondary question, we ask whether any significant differences are due to local effects or feedbacks on the large-scale circulation. The former would encapsulate processes connecting water vapor supersaturation to hydrometeor fallout, while the latter would include radiative processes, latent heat release from condensation and freezing, and evaporative cooling, which influences cold pool development.

Section 2 describes the datasets and the model used, including the microphysics parameterizations analyzed. Section 3 describes how extremes are defined and fit with an extreme value distribution. Results are presented in section 4 and conclusions in section 5.

Table 1
Summary of Microphysics Experiments Performed

Name	SAM scheme	Rimed ice species	Size threshold in raindrop breakup process (μm)
1MG	1MOM	Graupel	n/a
2MG300	M2005	Graupel	300
2MH300	M2005	Hail	300
2MG600	M2005	Graupel	600

2. Data and Methods

2.1. SPCAM

We have employed SPCAM, where CAM is the atmospheric component of the Community Earth System Model (Hurrell et al., 2013), forced by prescribed monthly SSTs and sea ice boundary conditions (a run-time configuration known as the F_2000 compset). The embedded CRM is the System for Atmospheric Modeling (Khairoutdinov & Randall, 2003). In our experiments the CRMs have been configured in a 2D, east-west orientation with periodic boundary conditions with grids of 32 columns at 2-km horizontal resolution and 30 vertical levels corresponding to those in CAM. The large-scale, outer grid is implemented using a finite-volume dynamical core with a $1.9 \times 2.5°$ resolution. Greenhouse gas concentrations are held fixed at year-2000 levels, and (bulk) aerosols are prescribed. Precipitation rates have been output every 3 hr, and to be clear, average rates (equivalent to accumulation over the 3-hr time period), as opposed to instantaneous rates every 3 hr, were output.

2.1.1. Microphysics Parameterizations

System for Atmospheric Modeling is equipped with two fundamentally different microphysics formulations, specifically the original, 1-moment scheme (1MOM in Table 1) described in Khairoutdinov and Randall (2003) and the 2-moment one (M2005 in Table 1) from Morrison et al. (2005). The number of moments refers to the number of degrees of freedom in the size distribution of a precipitating hydrometeor, which is prescribed to be an inverse exponential (Marshall & Palmer, 1948). The 1-moment scheme prognoses the mixing ratio and diagnoses the number concentration, while the 2-moment scheme prognoses both. We note that using 1-moment microphysics was only able to be done by using SPCAM4, while the experiments with 2-moment microphysics were only able to be done using SPCAM5. While CAM5 (Neale et al., 2010b) uses different boundary layer and convection parameterizations from those of CAM4 (Neale et al., 2010a), as mentioned before, the embedded CRM overrides all of them, so most of the difference between the two versions of SPCAM lies in the microphysics and the aerosol processes. 1n 1MOM there is no explicit droplet activation or ice nucleation; rather, the cloud condensate is diagnosed by assuming zero supersaturation at all times. Cloud liquid and cloud ice are then partitioned using a linear relationship that depends only on the temperature. In contrast, M2005 explicitly calculates droplet activation and ice nucleation. One caveat is that CAM4 uses the CAMRT radiation scheme (Collins et al., 2002), while CAM5 uses RRTMG (Iacono et al., 2008), a potentially confounding factor in this analysis.

Two additional experiments have been performed by modifying M2005. In the first, graupel is replaced by hail, which has a higher bulk density as well as a faster fall speed (2MH300). In the second (2MG600), the threshold diameter in the raindrop breakup/self-collection process implemented following Verlinde and Cotton (1993) has been increased from the default of 300 to 600 μm.

Table 1 compares the four representations. By allowing an extra free parameter in a hydrometeor's size distribution, a 2-moment scheme would be expected to generate more realistic cloud dynamics and properties, such as liquid and ice water contents and radiative fluxes, and indeed, it has been shown to do so in several studies (Bryan & Morrison, 2012; Igel et al., 2015). Prognosing both the mixing ratio and the number concentration enables both quantities to evolve independently, thereby admitting size sorting to occur. For instance, a large mixing ratio and low number concentration implies larger particles, which fall faster and are less prone to evaporation than smaller ones. However, excessive sorting has been shown to occur when using 2-moment schemes, and there are various suggestions on how to ameliorate the problem, such as a 3-moment scheme in which the third moment is radar reflectivity (Milbrandt & Yau, 2005).

The 2MH300 and the 2MG600 representations might both be expected to increase rainfall extremes. By falling faster, hail might have less time to melt into raindrops susceptible to evaporation, thus increasing

surface precipitation. This process is complicated, however, by the fact that differing melting and evaporation rates can impact cold pool development and propagation and thus affect the amount of time a storm spends over a particular location. Similarly, the larger raindrops allowed in the 2MG600 experiment should lead to faster fallout that is less impeded by breakup or reevaporation.

The 2MG300 case is considered the baseline, and comparisons will be made between it and the other three in turn. All figures plotting a difference in return values between two microphysics schemes will have that of 2MG300 as the minuend (quantity being subtracted from) and that of the other case as the subtrahend (quantity being subtracted). We omit all maps/comparisons containing fewer than five grid cells showing statistically significant differences, as determined using the method described in section 3.3 below.

2.2. Experiment Design

2.2.1. Climatological Runs

The 7-year climatological runs conducted for this analysis have start dates on 1 June 1991 and end dates on 30 November 1998. The first 6 months have been discarded to remove spinup effects. Two sets of experiments have been performed, one with present-day, annually cyclic, climatological SSTs, and the other with the same SST field uniformly increased by 4 K.

2.2.2. ILIAD

To determine whether the results seen in the (present-day) climatological runs are due to differing immediate, local effects or the microphysics' feedback on the large-scale circulation, we make use of the InitiaLized-ensemble, Analyze, and Develop (ILIAD) framework introduced by O'Brien et al. (2016). This software framework is designed to assess the effects of different parameterizations or model resolutions by conducting repeated short hindcast simulations initialized with reanalysis output and comparing the results against the meteorological conditions observed during the hindcast period. The experimental protocol, which we also use here, consists of one 5-day hindcast initialized at 00Z every day for 5 years. Precipitation output is taken from the fifth simulation day, a timeframe that allows the model to develop a somewhat distinctive dynamical state yet still be constrained to that observed on the verification date. The hindcasts are initialized by Climate Forecast System (CFS) (Saha et al., 2010) output taken from 1 December 2007 to 30 November 2010 and from 1 December 2011 to 30 November 2013, with the missing year reflecting our desire to conduct seasonal analyses amidst the fact that CFS version 2 (Saha et al., 2014) was made operational in March 2011. The CFS data was regridded to the large-scale SPCAM grid and include cloud water but no cloud ice or aerosol outputs. It should also be noted that the CFS data has zero land ice, in contrast to the monthly climatological data used in the free-running experiments. This is a potentially confounding factor but is partially mitigated by the fact that the precipitation probability density functions (PDFs) from the ILIAD runs lie within the interannual spread of those from the free-running ones (not shown).

2.3. Observations

The Climate Prediction Center (CPC) Hourly Precipitation Dataset (HPD Higgins et al., 1996) is used to assess the fidelity of the model output. The HPD is derived from rain gauges and gridded using a modified Cressman scheme to cover a region spanning 20–60N and 220–297.5E. One third of the stations are first-order National Weather service stations, and the remaining two thirds consist of data from cooperative observers. The seven years of observations that are contemporaneous with the simulation period during 1991–1998 have been interpolated to the CAM grid.

3. Identifying and Comparing Extremes

3.1. Background on Extreme Value Distributions

To compare precipitation extremes, extreme value distributions were fitted to 3-hourly data for each season and in each grid cell within the contiguous United States (CONUS). There are two ways in which extremes from a time series are selected and modeled. The first is to divide the time series into regular intervals, say, months, then take the maximum of each period. Building on work done by Fisher and Tippett (1928), von Mises (1954), and Jenkinson (1955) independently showed that the generalized extreme value (GEV) distribution asymptotically approximates the distribution of these block maxima. The primary drawback of this method is that it ignores other extremes that might be present, for example, the second highest rain rate in a block. An alternative approach to fitting extremes is the threshold exceedances method, in which a

distribution is fitted to all values over a chosen threshold. Pickands III (1975) showed that such a distribution can be asymptotically approximated by the generalized Pareto distribution (GPD).

The GPD has a slight disadvantage in that the parameter estimates are a direct function of the threshold u; for example, analytically finding the GPD representation of the standard Fréchet model $F(x) = \exp(-1/x)$ yields a scale parameter $\sigma = u$. Thus, we employ a third approach: the nonhomogeneous Poisson point (NHPP) process (Pickands III, 1971), which is also referred to as peaks over threshold. Like the threshold excess approach, the NHPP method models all extreme values greater than a specified threshold but characterizes the extreme value distribution using statistical parameters that correspond to the GEV approach. An NHPP process is defined by an intensity *function* with the identical analytic expression as that for the GEV PDF, namely,

$$\lambda(t, y) = \frac{1}{\sigma}\left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]^{-1/\xi-1}, \tag{1}$$

where μ is a location parameter, σ a scale parameter, and ξ a shape parameter. This is not a coincidence, as it can be shown that both the GPD and the GEV are special cases of the point process representation. In fact, return values for the NHPP process are calculated here in the same manner as for the GEV distribution: the N-block return value is that which is expected to be exceeded once every N blocks, where blocks here are years. However, λ is not itself a PDF but rather the mean number of exceedances over a given rain rate in one time block. This is analogous to a time rate r for a Poisson distribution, where $\lambda_{\text{Poisson}} = rt$ is the expected number of events in an interval t.

Thus, integrating $\int_0^T \int_u^\infty \lambda(t, y) dt dy$, where T is the total number of time blocks, yields the intensity *measure*

$$\Lambda(T, u) = T\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}, \tag{2}$$

which is the expected number of exceedances of the rain rate over the threshold u over T years.

Given a set of observed extremes $Y_1, \ldots, Y_N$ at times $t_1, \ldots, t_N$, the negative log likelihood of an NHPP process is

$$-\ln L = N \ln \sigma + \left(\frac{1}{\xi} + 1\right) \sum_{k=1}^{N} \ln\left[1 + \xi\left(\frac{Y_k - \mu}{\sigma}\right)\right] + \Lambda(T, u), \tag{3}$$

where L is the likelihood. We use optimization techniques to minimize equation (3) to find the parameters μ, σ, ξ.

3.2. Likelihood Ratio Test

To compare two microphysics schemes, we used the likelihood ratio test to determine whether the fitted NHPP processes are significantly different from each other. Such a test is generally framed with hypotheses

$$H_0 : \theta \in \Theta_0,$$
$$H_a : \theta \in \Theta \setminus \Theta_0,$$

with θ the distribution parameter(s) and Θ_0 a subset of the parameter space Θ.

Wilks (1938) proved that, given the observed data x and assuming H_0 is true, the test statistic

$$z = -2 \ln \frac{\sup\{L(\theta|x) : \theta \in \Theta_0\}}{\sup\{L(\theta|x) : \theta \in \Theta\}}$$

asymptotes to a chi-square distribution with k degrees of freedom, where k is the difference in dimensionality, or number of parameters, between Θ and Θ_0. Thus, a p value can easily be computed.

3.3. Comparison Procedure

A null and an alternative hypothesis can be defined for this study as follows:

$$H_0 : \theta_i = \theta_j,$$
$$H_a : \theta_i \neq \theta_j,$$

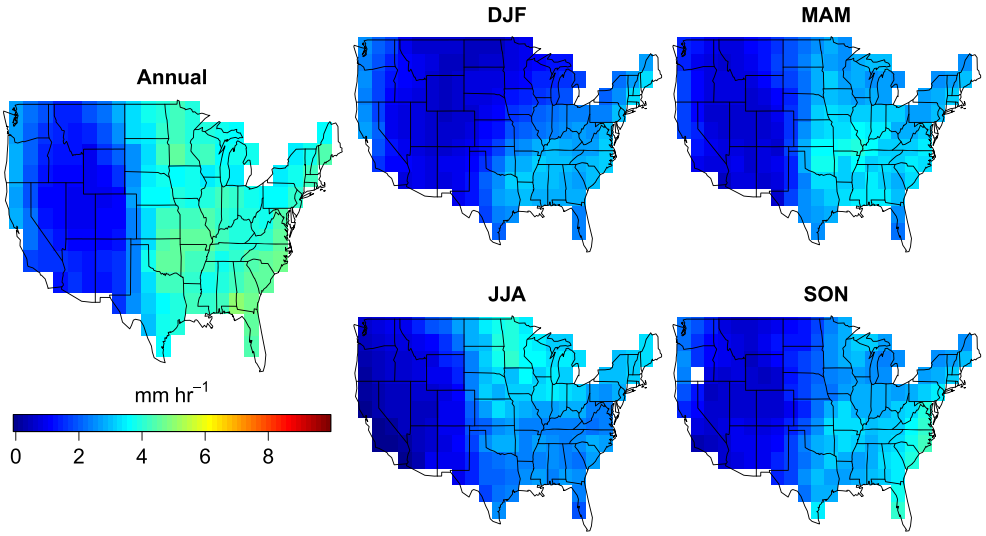

Figure 1. A 2-year return value for 2MG300 (Table 1) in the present-day climatological experiment. The legend axis matches that in Figure 10 for ease of comparison with observations.

where θ_i denotes the parameters for the NHPP fit of microphysics case i and θ_j those of case j. In other words, the null hypothesis states that the two cases being compared can be reasonably represented by one common distribution, while the alternative hypothesis says that the two individual fits are significantly different, that is, the extremes arise from distributions that exhibit meaningful differences. Our test statistic then can be written as

$$z = -2 \ln \frac{\sup\{L(\theta_{12} | x_1, x_2)\}}{\sup\{L(\theta_1 | x_1)\} \sup\{L(\theta_2 | x_2)\}}. \tag{4}$$

Because the test statistic requires that the parameter estimation in the numerator and the denominator be done using the same data, the former is found by analyzing the union of the extremes for each individual microphysics case, rather than extremes determined from the union of the whole datasets. This also necessitates a uniform threshold u. Therefore, when comparing two datasets in a given grid cell and a season, u was defined as the higher of the two individual 98th percentiles, and extremes for both cases were defined as values greater than u. The 98th percentile was chosen because it was found that higher thresholds, for example, the 99th percentile, led to too many failures of the fitting algorithm. The results also depend on the percentile chosen (Figures S1-S3), particularly when comparing the 1- and 2-moment schemes. Thus, we chose the highest integer percentile possible that also mitigated the incidence of convergence failures. This also ensures that the results shown are the most robust. Lastly, to account for temporal clustering, for example, a storm with high rain rates over multiple 3-hr periods, we took only the maximum value in each series of consecutive exceedances.

Once the three sets of extremes were acquired, a p value was calculated using the test statistic with $k = 3$ degrees of freedom since there are six free parameters in the alternative model and only three in the null, where the parameters are the same between the two cases being compared. However, because a p value is the probability of obtaining a test statistic at least as extreme as that observed while assuming H_0 is true, carrying out N tests with a significance level α will, on average, result in αN rejections if the tests are independent, and $> \alpha N$ if there is spatial correlation, simply due to random sampling error. Thus, as noted by Wilks (2016), the effects of multiple hypothesis testing (conducting individual tests at multiple grid points) must be taken into account. We applied the method pioneered by Benjamini and Hochberg (1995) known as controlling the false discovery rate (FDR), which is the expected fraction of rejected local null hypotheses that are actually true. The procedure involves choosing α_{FDR}, the level at which it is desired to limit the fraction of false negatives (type I errors). In this paper we set $\alpha_{\mathrm{FDR}} = 0.05$. In practice, this means that there is then a value p_{FDR} such that an individual p value p must satisfy $p \leq p_{\mathrm{FDR}} \leq \alpha_{\mathrm{FDR}}$ in order to be declared significant to ensure that overall the rate of false discoveries does not exceed the nominal level. This method is robust, and even conservative, when there is strong spatial correlation since the achieved FDR will be smaller than α_{FDR}. For more details, see Appendix A.1.

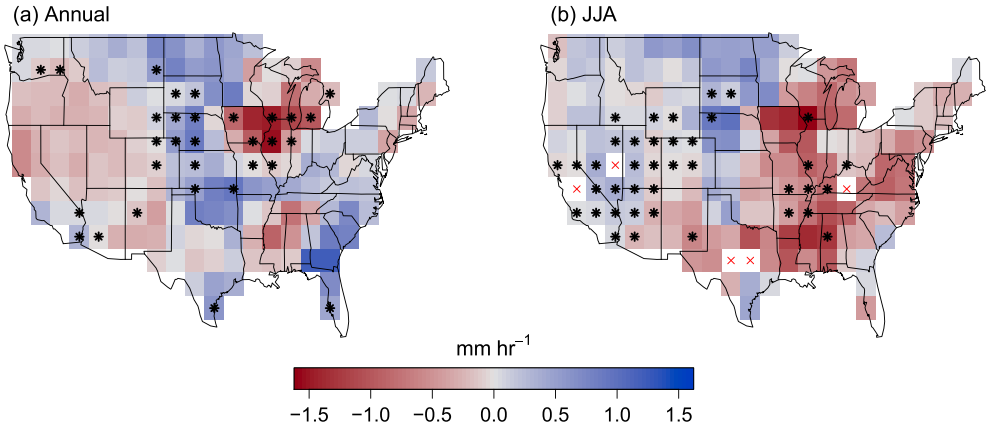

Figure 2. Climatological comparison of 1MG (with 2MG300) (Table 1). Following the convention defined in section 2.1.1, the quantity plotted is the 2MG300 2-year return value minus that of 1MG. Asterisks denote grid cells with statistically significant differences between the two extreme precipitation distributions. Grid cells where parameter estimation failed are denoted by a red X.

4. Results

4.1. Present Day

4.1.1. Climatological Runs

Figure 1 shows the map of 2-year return values for 2MG300. Generally, across all seasons, extremes are lowest near the Rocky Mountains and highest in and between the Midwest and the East Coast, with moderate values on the West Coast except in JJA. When comparing the 1- and 2-moment schemes in the present day (Figure 2), there are numerous significant differences over the full annual cycle in the Midwest and during the summer season in the Southwest and Southeast. In the case of the southwestern United States, the location and the timing might imply interactions with the North American monsoon. Previous work has found that aspects of microphysical processes, for example, a more maritime-like, drop-size distribution and greater liquid and ice mixing ratios, may be important in producing heavier rainfall in organized convection during the North American monsoon (Rowe et al., 2012). We caution, however, that much of this summertime signal may actually be noise from interannual variability (see section 4.1.4). Note also that in this region the diverse distributions have differences in return values that are quite small in magnitude. While this is likely in large part due to the lower rain rates in the region (as compared to the rest of CONUS), it should be noted that disparate distributions can yield identical return values since there are three parameters that describe an NHPP process.

When comparing the two rimed ice species in the 2-moment scheme (Figure 3), the Southwest again displays significant differences, though to a lesser degree than in Figure 2. Graupel leads to higher extremes

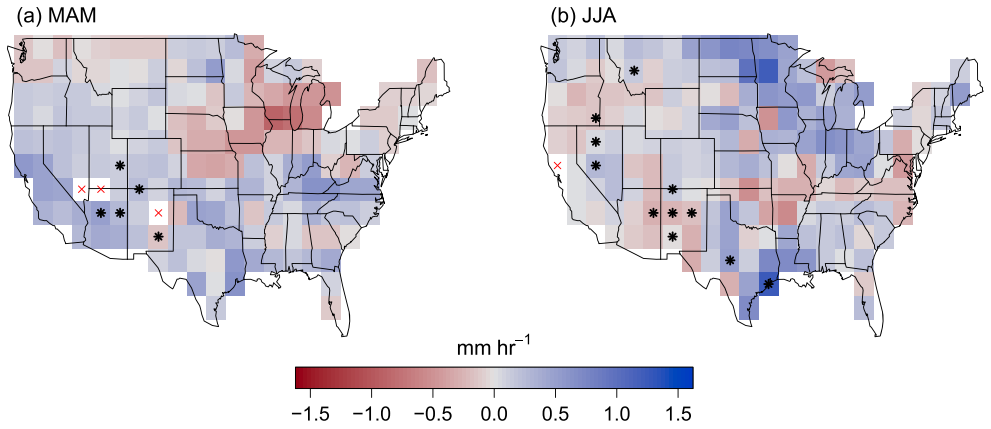

Figure 3. As in Figure 2 but with climatological comparison of 2MH300 and 2MG300 (Table 1).

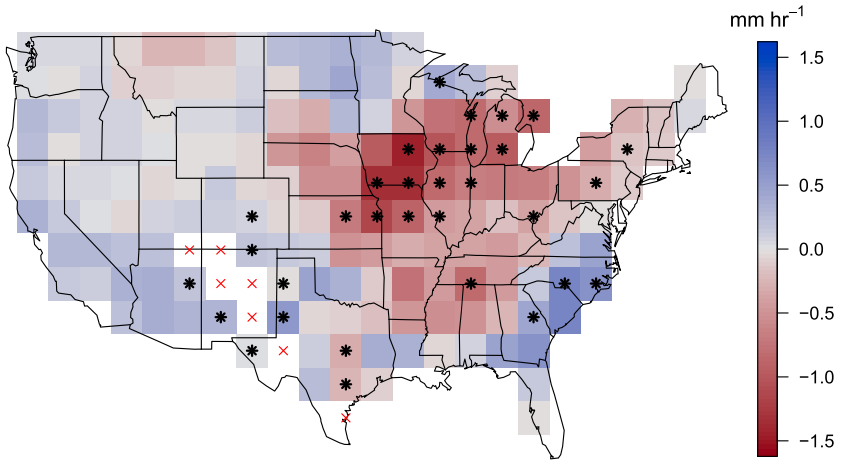

Figure 4. As in Figure 2 but with climatological comparison of 2MG600 and 2MG300 (Table 1) in MAM.

in this area in the spring, while the opposite is true in the summer. The fact that there are so few significant differences in Figure 3 is perhaps a surprising result given other studies (Bryan & Morrison, 2012; Van Weverberg et al., 2012) that have demonstrated impacts on cold pool dynamics and subsequent accumulated precipitation. Finally, increasing the size threshold before increased raindrop breakup intensifies leads to higher 2-year return values in the Midwest spring (Figure 4). Of the present-day findings, this is the most robust when varying the percentile used to demarcate extremes (Figure S3).

The robust signal in the Midwest and the seasonality (occurring mostly in the spring and the summer but never in the winter) points to the choice of microphysics being more important in places and times where storms are dominated by convective processes as opposed to large-scale air motions. This is a similar conclusion reached by Li et al. (2012), who found that CAM and SPCAM yielded similar precipitation extremes in the western United States and during winter but vastly different results in the southeastern United States and during summer. The springtime signal in the Midwest, particularly in Figure 4, could suggest that SPCAM is impacting the transition from predominantly synoptic to convective precipitation, though we do not investigate this further.

4.1.2. 5-Day Runs

The central and western Uniteed States display significant differences in annual extremes when comparing the 1- and 2-moment schemes in the ILIAD framework (Figure 5), though fewer than in the climatological case. These regional disparities are not collocated with those in Figure 2, implying that local microphysical

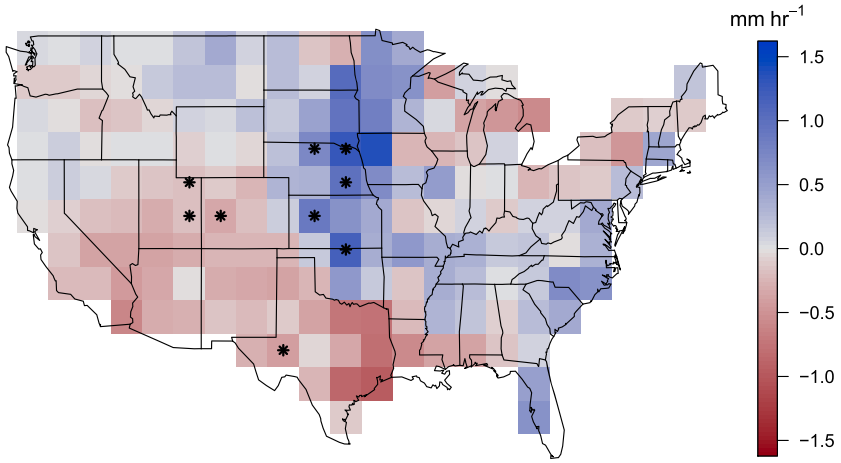

Figure 5. As in Figure 2 but with comparison of 1MG and 2MG300 (Table 1) annual extremes within the ILIAD framework.

effects can cause differences in SPCAM precipitation extremes when varying the number of moments prognosed. Nevertheless, this is the only case where there are significant differences; none occur in the seasonal comparisons with 1MG and 2MG300 nor in any of the comparisons between the 2-moment schemes. Given the general lack of stippling, a conclusion consistent with this evidence is that the microphysics parameterizations alone (i.e., operating under the same meteorological conditions) do not often lead to significantly different extremes in SPCAM simulations within CONUS. Instead, an alternate plausible hypothesis is that precipitation rates are indirectly influenced via feedbacks from the microphysics to the large-scale circulation that operate on longer timescales. These ILIAD runs are more similar to the aforementioned CRM studies forced by reanalyses at the boundaries (e.g., Bryan & Morrison, 2012; Van Weverberg et al., 2012), yet our findings, that is, that the local effect of microphysics do not appear to be significant, seem to differ with those in the literature.

There are several reasons as to why the sensitivities are diminished in SPCAM. One is our use of a 2D domain in the embedded CRM. Various CRM studies (Phillips & Donner, 2006; Wilhelmson, 1974; Zeng et al., 2008) have noted that updrafts, especially the most intense ones associated with deep convection, have faster speeds in 3D simulations. Phillips and Donner (2006) and Zeng et al. (2008) found that this led to higher graupel mixing ratios, and more graupel could potentially make the distinction between graupel and hail more important. Cloud dynamics would also affect the balance between warm rain and cold rain production because more moderate updrafts would favor increased coalescence of cloud water while faster vertical velocities lead to higher penetration, resulting in more ice (e.g., Redelsperger et al., 2000). A second possible cause for the reduced sensitivities is the fact that the data were averaged over the CRM domain within the GCM column, thus perhaps making a comparison to domain-mean (as opposed to domain-maximum) precipitation rates from CRM studies more appropriate. A third confounding factor is the limited vertical resolution (30 levels), which is restricted to match that of CAM. Khairoutdinov et al. (2009) found a dependence of cloud distribution and convection on vertical resolution, so it is possible that future SP models with an increased number of vertical levels could show more realism and/or sensitivity to experiments.

The findings here are robust regardless of the percentile used to denote extreme values (Figures S4-S6), lending support to the idea put forth by Khairoutdinov and Randall (2003), namely, that nonlocal feedbacks can amplify microphysical sensitivities that would otherwise be inoperative or suppressed due to prescription of meteorological boundary conditions in limited-domain CRM simulations. It is also interesting to compare our finding with that of Li et al. (2012), who found significant disparities between CAM and SPCAM even when the model's large-scale fields were replaced with NCEP reanalysis data at each time step. Here, the model is simply initialized with reanalysis fields for short, 5-day runs, yet little difference between the sundry microphysics schemes is observed.

4.1.3. Evidence of Large-Scale Circulation Changes

The results in sections 4.1.1 and 4.1.2 support seeking evidence of large-scale circulation changes that could lead to the hypothesized feedbacks on extreme precipitation. Feng et al. (2018) compared two microphysical schemes within a convection-permitting model and found that one parameterization simulated more stratiform rainfall within MCSs, resulting in heating profiles that were top heavier and had a sharper heating maximum. This in turn led to the strengthening of midlevel mesoscale convective vortices brought on by increased potential vorticity generation in this region, helping deepen the synoptic scale trough within which the MCSs were embedded. Dry air intrusion, followed by increased evaporative cooling, allowed the systems to last longer and affect the large-scale circulation. This change in synoptic flow described by Feng et al. (2018) is a somewhat local phenomenon, with differences in heating profiles ultimately feeding back on the same storm, with the caveat that MCSs travel O(1,000 km) over their lifetime. More importantly, however, MCSs have a lifetime of about 1 day, so we would expect to see effects from this process in the ILIAD runs. Indeed, it may partially explain the results in Figure 5, especially since most of the signal is in the Great Plains, where MCSs are frequent and produce much of their rainfall (Feng et al., 2018). However, as stated above, effects that manifest in the climatological, but not ILIAD, runs imply nonlocal feedbacks affecting extreme precipitation. Figure 6a compares 1MG and 2MG300 heating rates at their respective midtropospheric (400–800 hPa) peaks in each grid column, and Figure 6b shows differences in the zonal mean. We see that the largest differences occur from about 20S–20N. While we do not attempt to further connect heating rates with the large-scale circulation, we note that this is consistent with the idea of a nonlocal feedback. We also looked at changes in convective available potential energy but found it larger nearly everywhere in CONUS when using 1-moment as opposed to 2-moment microphysics (Figure S7).

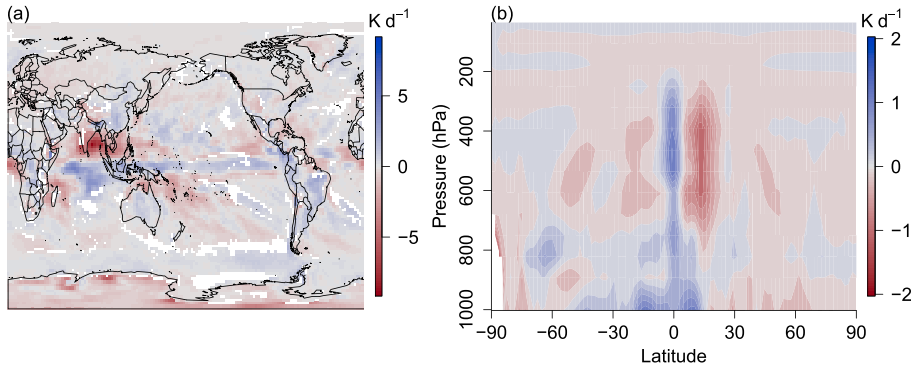

Figure 6. Climatological 2MG300 (Table 1) heating rates minus those of 1MG, averaged across times with columns with $\omega_{500\ \mathrm{hPa}} < 0$ hPa d^{-1}. (a) Difference in the midtropospheric (400–800 hPa) peaks in the climatological heating profile; only columns with peaks (between 400 and 800 hPa) less than 100 hPa apart are colored. (b) Zonal mean of difference in heating rates.

The top row of Figure 7 reprints some of the climatological results from section 4.1.1, while the bottom row shows the respective differences in daily mean pressure velocity at the 500-hPa level (ω_{500}). In the cells/regions with statistically significant differences, the response in precipitation extremes generally follows that expected from the differences in ω_{500} (which can be on the order of 30%), particularly between Panels 7a and 7d and between 7c and 7f. There is less similarity when comparing 1MG and 2MG300 in the summer (between Panels 7b and 7e), especially in the southwestern United States. Low humidity, or relatively modest "extremes" (leading to small return values, as seen in Figure 1), could serve as an explanation. Note that an exact correspondence between the two rows of Figure 7 is not expected for multiple reasons. The differing output frequencies of precipitation and ω_{500} is one problem, which leads to another: it is not possible to pull ω_{500} values from the same time steps as the rainfall extremes. Thus, the bottom row of Figure 7 shows differences averaged over the entirety of the corresponding season and therefore might be more telling of shifts in mean precipitation. Similarities between the top and bottom rows, then, hint at shifts in the precipitation rate distribution that move both the mean and extremes in the same direction, while disparities imply changes in the shape of the distribution. Lastly, as described further below, one would expect strong precipitation events to be proportional not only to the upward motion but to the moisture content (which was unfortunately not output) as well.

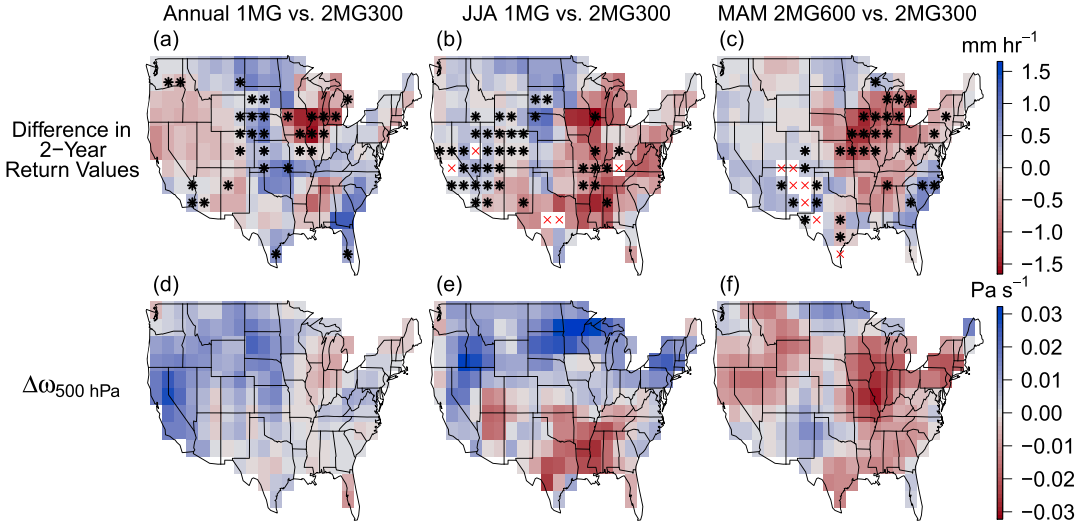

Figure 7. Climatological 2MG300 (Table 1) 2-year return value minus that of (a) 1MG in the annual, (b) 1MG in JJA, and (c) 2MG600 in MAM, with asterisks denoting grid cells with statistically significant differences. 2MG300 daily mean ω_{500} subtracted from that of (d) 1MG in the annual, (e) 1MG in JJA, and (f) 2MG600 in MAM.

Table 2
Number of CONUS Cells with Statistically Significant Differences

Season	1MG[a]	2MH300[a]	2MG600[a]	Reshuffling median
Annual	31	1	0	1
DJF	1	0	0	0
MAM	1	5	32	0
JJA	39	11	2	9
SON	0	3	0	0

[a]In climatological comparison with 2MG300.

4.1.4. Estimating Natural Variability of SPCAM Return Values

It is possible that 7 years is too short an integration period, such that, despite the statistical lengths we have gone to try to ensure a robust signal, the differences we see are artificially significant, and attempts at interpretation are moot. To test this, we grouped the climatological output into 28 years (consisting of the four microphysics experiments of 7 years each). We then randomly reshuffled the 28 years and redid our analysis, with the idea that this gives a sense of the signal in a dataset where none is expected. We did this reordering 600 times, and the median number of CONUS cells flagged each season as statistically significant by our algorithm is listed in the last column of Table 2. Note that in calculating the median, for each reshuffling and season, the three values generated (from comparing the "four" microphysics schemes) are taken as individual data points (as opposed to being summed), so that a fair comparison can be made to the actual climatological results (the middle three columns in Table 2). We see that our statistical procedure generally reports no significant differences when reshuffling the data, except during JJA. This suggests that, at least for extreme precipitation, signal detection over interannual variability is more difficult in JJA, which is in line with the findings of Elliott et al. (2016) regarding summertime MCSs.

4.1.5. Comparing with Observations

While there are significant differences in the climatological runs (Figures 1–4), the disparities in return values are small in comparison to those when measuring against observations. We show here only the comparison of the 1MG case to observations (Figure 8); the maps of the other three microphysics schemes (Figures S8-S10) look very similar. SPCAM systematically underestimates the magnitude of precipitation rates when compared to the CPC HPD data, particularly when looking at annual, summer, and fall extremes, with the closest (spatial) agreement in the winter. While this might seem to be at odds with the conclusions

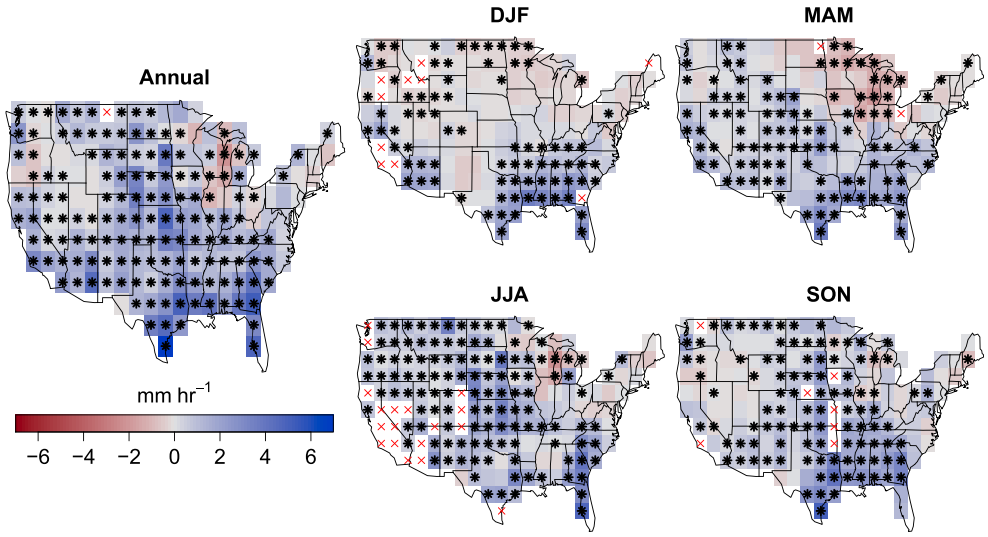

Figure 8. As in Figure 2 but the 1MG (Table 1) 2-year return value is subtracted from that of the CPC HPD observations.

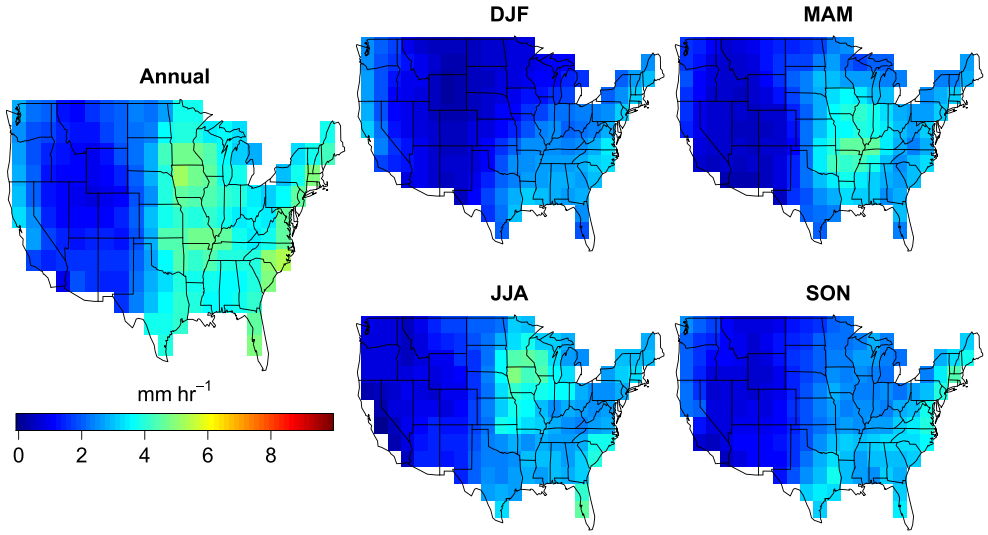

Figure 9. A 2-year return value for 2MG300 (Table 1) in the ILIAD framework. The legend axis matches that in Figure 10 for ease of comparison with observations.

of Li et al. (2012), the conclusions are in fact quite similar; notably, the model underestimates in the southeastern United States while showing better agreement, or even overestimation, in the north, particularly in the 2MG600 case.

Figures 9 and 10 show the maps of return values for the 2MG300 ILIAD case and the CPC observations, respectively. One might expect that the ILIAD framework, with its short, 5-day runs initialized by reanalysis data, better captures the location, if not the magnitude, of precipitation events. Comparing with Figure 1, this appears to be the case at least in the summer, with the largest extremes placed further south in the Midwest, closer to Iowa. Thus, as in Li et al. (2012), it seems that winter precipitation events dominated by synoptic systems are better represented in terms of both magnitude and location, whereas summer events dominated by convective processes are potentially subject to greater errors, even in SPCAM.

Several studies (Fildier et al., 2018; O'Brien et al., 2016; Rauscher et al., 2016) have shown empirical evidence for the idea that, for strong events, precipitation rates are approximately equal to the cloud-base upward moisture flux, which is the product of vertical velocity (in pressure coordinates) and specific humidity. As mentioned above, 3D CRM simulations generally exhibit larger updraft speeds, particularly at the highest

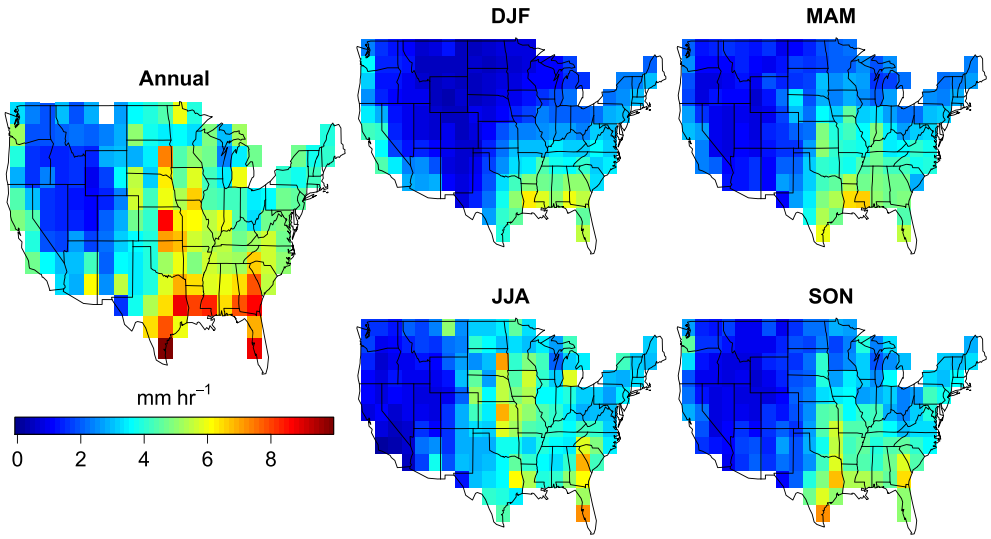

Figure 10. A 2-year return value for the CPC HPD observations.

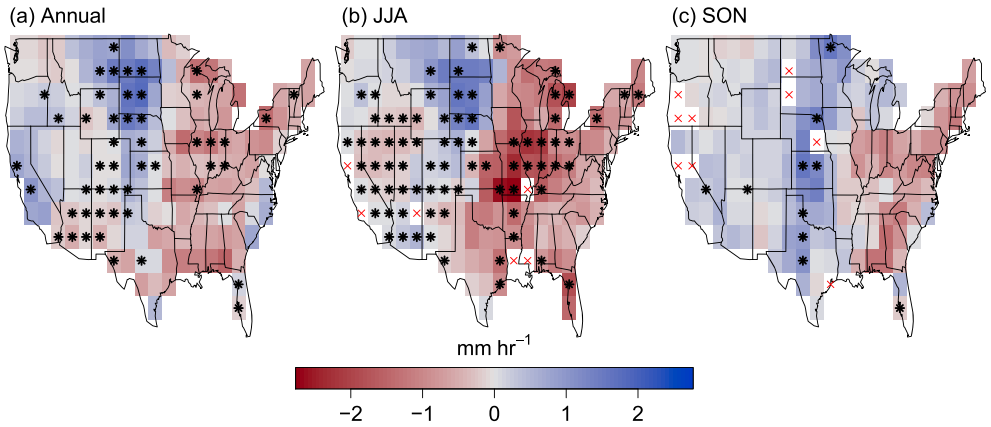

Figure 11. As in Figure 2 but with comparison of 1MG and 2MG300 (Table 1) in experiment with +4 K SSTs.

percentiles: Phillips and Donner (2006) found that the most vigorous 1% of updrafts were 20–40% faster, and Wilhelmson (1974) found that the updraft in a 3D thunderstorm simulation was twice as fast. Indeed, the latter study found that the maximum rainfall intensity in 3D was about three times that of the 2D thunderstorm. Thus, it is possible that embedding a 3D CRM domain (instead of a 2D one) within each GCM column would result in a closer match to observations. Increasing the horizontal resolution of the CRM might also help: Ooyama (2001) showed in 2D simulations of a single-cell cloud that 1-km resolution resulted in higher rainfall intensities than when using 2 km, which in turn yielded a peak precipitation rate twice that in a 4-km experiment. Nevertheless, this effect is not a surety, especially in a 3D setup: Bryan and Morrison (2012) found that domain accumulated precipitation—which, as noted above, might be the more relevant quantity in SPCAM—was relatively similar between a 4- and a 1-km simulation of a squall line. And in fact, precipitation *decreased* with 0.25-km resolution due to increased entrainment which weakened the system intensity.

4.2. Uniformly Increased SSTs

Finally, we analyzed identical free-running simulations but with SSTs uniformly increased by 4 K. Again, there are substantial regions with differences when comparing the 1- and the 2-moment schemes over the whole year as well as summertime (Figure 11). In this case, however, the stippling is not confined to the southwestern United States in the summer; parts of the Midwest show significant differences as well, with 1MG showing more extreme precipitation rates further east, 2MG300 further west. As when using present-day SSTs, there is no detectable signal when comparing 2MG300 and 2MH300. The springtime signal when comparing 2MG300 and 2MG600 (Figure 4) also disappears.

We also investigated changes in return values when going from the present climate to the +4 K SST climate. We would generally expect the most intense precipitation events to increase in magnitude with warming but does the response depend on the microphysics scheme used within SPCAM? Figure 12 suggests the answer is no for annual extremes within CONUS: the return values significantly increase almost everywhere, but there is not much distinction between the four cases. There is some slight variation in the spatial structure, but generally, the largest increases (in an absolute sense) are in the Midwest, with secondary increases on the West and East Coasts.

A potentially interesting and similar simulation would be a pseudo-global warming experiment, as done by Pall et al. (2011), within the ILIAD framework. That is, first gather results from two (ensembles of) climatological (SP)CAM runs, one forced with observed historical time series of

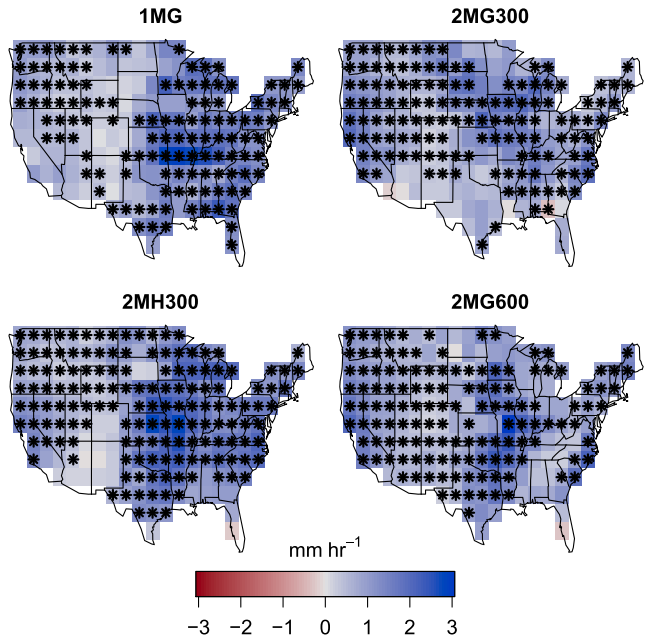

Figure 12. Present-day 2-year return value subtracted from that of the +4 K simulation. Asterisks denote grid cells with statistically significant differences between the two extreme precipitation distributions.

radiative and surface boundary conditions, the other with 4xCO2 (alternatively, the increased amount of CO2 necessary to produce an average of 4 K global warming) and the corresponding surface boundary conditions, as estimated by a fully coupled atmosphere-ocean model, for example, (SP)CESM. Take the difference in atmospheric variables between the future and present climates in, say, the month of the ILIAD simulation and add it to the corresponding reanalysis fields used as initial conditions. This set of integrations would be like the "counterfactual" case in event attribution studies, for example, Pall et al. (2011, 2017), except that instead of being a scenario without already occurred anthropogenic emissions, it represents a future world with continuing emissions. The local effect of changing microphysics schemes on return values in a warmer climate could then be evaluated in an ILIAD framework as in section 4.1.2.

5. Conclusions

Earlier studies have found that superparameterization, which is a mechanism to introduce convection-permitting processes into GCMs, can increase the fidelity of the precipitation extremes simulated with these models. However, even as increasing numbers of climate models transition to convection-permitting or convection-resolving atmospheric dynamics, microphysical processes will still require parameterization. Experiments with limited-domain CRMs have shown that rainfall statistics are sensitive to the choice of microphysical representations. However, it was previously unclear whether nonlocal feedbacks in a global model would amplify or diminish such a response.

In this study we compare precipitation extremes generated by SPCAM for various microphysical configurations. In the climatological simulations, we find significant differences in both present-day and warmer climates, particularly between the extreme statistics generated with the 1- and the 2-moment schemes during summer and the whole year. Short, reanalysis-initialized runs performed via the ILIAD framework suggest that this response is mostly due to effects on the large-scale circulation that then feed back into rainfall extremes. This conclusion is supported by comparing mean 500-hPa vertical velocity maps, whose differences we hypothesize stem from variations in vertical heating profiles caused by the disparate microphysics representations. Despite the disparities in rainfall statistics, we also find that none of the four analyzed microphysical parameterizations yield good agreement with rain gauge data, as SPCAM consistently underestimates the observations. And in fact, if the intensity extremes are in turn underestimated in the observational data sets (both the one used here and others produced in the past), the discrepancies between SPCAM and the data would be further amplified. Recent work done by Risser et al. (2018) showed that current gridded precipitation products have diminished variability and extreme values due to the interpolation of a fractal field in precipitation. They argue that an improved product would first estimate the extreme value distribution locally and only afterwards interpolate the climatology.

Our results are perhaps contrary to what might be expected based on the results of Elliott et al. (2016), namely, that global feedbacks lead to internal variability of storms that dwarfs sensitivity to microphysics configurations. Although, like Elliott et al. (2016), we found significant interannual noise in the summertime, we also found that nonlocal feedbacks magnify and highlight differences that might otherwise have gone undetected in shorter experiments. However, at least for the configuration and resolution used, none of the microphysical schemes tested are clearly preferable in terms of biases in return values of extreme precipitation. Helpful improvements in microphysics parameterizations might address both frozen and liquid processes. In the case of the former, better understanding of ice formation and evolution during aggregation and riming—which are poorly constrained processes and made even more difficult by the nonphysical partitioning of hydrometeors into discrete snow/ice/graupel/hail categories—is needed (Grabowski et al., 2019). For the latter, implementing stochasticity into the collision-coalescence process, possibly using a Lagrangian, particle-based approach (Grabowski et al., 2019; Shima et al., 2009) might lead to a faster onset of rainfall due to some droplets growing faster than predicted in the usual continuous, deterministic formulation (Telford, 1955). In addition to broader work improving microphysics parameterizations, future work should involve seeing whether increasing the resolution, particularly that of the internal CRM, helps SPCAM align more closely with observations in its representation of rainfall extremes.

Appendix A

A.1. Controlling the False Discovery Rate

Once α_{FDR}, the level at which it is desired to limit the false discovery rate (section 3.3) is chosen, the procedure for rejecting null hypotheses is as follows. For each comparison between two datasets in a given season,

and given N finite p values p_i with $i = 1, \ldots, N$, sort the p_i into ascending order. Note that N is bounded above by $N \leq 175$, the number of GCM grid cells covering CONUS given the resolution used for the GCM spatial grid. Using standard statistical notation, these p values are now denoted with parenthetical subscripts, such that $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(N)}$. Local null hypotheses are now rejected if their respective p values are no larger than p_{FDR}:

$$p_{\mathrm{FDR}} = \max_{i=1,\ldots,N} \left[p_{(i)} : p_{(i)} \leq (i/N)\alpha_{\mathrm{FDR}} \right]. \tag{A1}$$

Thus, it can be seen that this method, known as the Benjamini-Hochberg procedure, requires even smaller p values to reject a local null hypothesis than would be needed when examining individual grid points in isolation.

Acknowledgments
This research was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 as part of their Earth and Environmental Systems Modeling Program and used resources of the National Energy Research Scientific Computing Center (NERSC), also supported by the Office of Science of the U.S. Department of Energy, under Contract No. DE-AC02-05CH11231. Hossein Parishani was supported by the U.S. Department of Energy under Contract No. DE-SC0012548. CPC Hourly U.S. Precipitation data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/. The model can be accessed at DOI: 10.5281/zenodo.3727515. Model data needed to reproduce the figures can be found at https://portal.nersc.gov/project/m1517/cascade/charn2019_spcam_microphys/.

References

Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, *438*(7066), 303.

Benedict, J. J., & Randall, D. A. (2009). Structure of the Madden–Julian oscillation in the superparameterized CAM. *Journal of the Atmospheric Sciences*, *66*(11), 3277–3296.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300.

Bryan, G. H., & Morrison, H. (2012). Sensitivity of a simulated squall line to horizontal resolution and parameterization of microphysics. *Monthly Weather Review*, *140*(1), 202–225.

Collins, W. D., Hackney, J. K., & Edwards, D. P. (2002). An updated parameterization for infrared emission and absorption by water vapor in the National Center for Atmospheric Research Community Atmosphere Model. *Journal of Geophysical Research*, *107*(D22), ACL 17–1-ACL 17-20.

Dai, A. (2006). Precipitation characteristics in eighteen coupled climate models. *Journal of Climate*, *19*(18), 4605–4630.

Elliott, E. J., Yu, S., Kooperman, G. J., Morrison, H., Wang, M., & Pritchard, M. S. (2016). Sensitivity of summer ensembles of fledgling superparameterized US mesoscale convective systems to cloud resolving model microphysics and grid configuration. *Journal of Advances in Modeling Earth Systems*, *8*, 634–649. https://doi.org/10.1002/2015MS000567

Feng, Z., Leung, L. R., Houze Jr, R. A., Hagos, S., Hardin, J., Yang, Q., et al. (2018). Structure and evolution of mesoscale convective systems: Sensitivity to cloud microphysics in convection-permitting simulations over the US. *Journal of Advances in Modeling Earth Systems*, *10*, 1470–1494. https://doi.org/10.1029/2018MS001305

Fildier, B., Parishani, H., & Collins, W. (2018). Prognostic power of extreme rainfall scaling formulas across space and time scales. *Journal of Advances in Modeling Earth Systems*, *10*, 3252–3267. https://doi.org/10.1029/2018MS00146

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Mathematical Proceedings of the Cambridge Philosophical Society* (Vol. 24, pp. 180–190). Cambridge: Cambridge University Press.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S., Collins, W., et al. (2013). Evaluation of climate models. In T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, & V. Bex (Eds.), *P. Midgley*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Grabowski, W. W., Morrison, H., Shima, S.-I., Abade, G. C., Dziekan, P., & Pawlowska, H. (2019). Modeling of cloud microphysics: Can we do better? *Bulletin of the American Meteorological Society*, *100*(4), 655–672.

Higgins, R. W., Janowiak, J. E., & Yao, Y.-P. (1996). *A gridded hourly precipitation data base for the United States (1963-1993) US Department of Commerce*: National Oceanic and Atmospheric Administration, National Weather Service.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., et al. (2013). The Community Earth System Model: A framework for collaborative research. *Bulletin of the American Meteorological Society*, *94*(9), 1339–1360.

IPCC (2014). *Climate change 2014: Impacts, adaptation, and vulnerability. Part A: Global and sectoral aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change Edited by* C. B. Field, V. R. Barros, D. J. Dokken, K. J. Mach, M. D. Mastrandrea, T. E. Bilir, M. Chatterjee, K. L. Ebi, Y. O. Estrada, R. C. Genova, B. Girma, E. S. Kissel, A. N. Levy, S. MacCracken, P. R. Mastrandrea, & L. L. White (pp. 1132). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, *113*, D13103. https://doi.org/10.1029/2008JD009944

Igel, A. L., Igel, M. R., & van den Heever, S. C. (2015). Make it a double? Sobering results from simulations using single-moment microphysics schemes. *Journal of the Atmospheric Sciences*, *72*(2), 910–925.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, *81*(348), 158–171.

Khairoutdinov, M. F., Krueger, S. K., Moeng, C.-H., Bogenschutz, P. A., & Randall, D. A. (2009). Large-eddy simulation of maritime deep tropical convection. *Journal of Advances in Modeling Earth Systems*, *1*, 15. https://doi.org/10.3894/JAMES.2009.1.15

Khairoutdinov, M. F., & Randall, D. A. (2001). A cloud resolving model as a cloud parameterization in the NCAR Community Climate System Model: Preliminary results. *Geophysical Research Letters*, *28*(18), 3617–3620.

Khairoutdinov, M. F., & Randall, D. A. (2003). Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities. *Journal of the Atmospheric Sciences*, *60*(4), 607–625.

Khairoutdinov, M., Randall, D., & DeMott, C. (2005). Simulations of the atmospheric general circulation using a cloud-resolving model as a superparameterization of physical processes. *Journal of the Atmospheric Sciences*, *62*(7), 2136–2154.

Kooperman, G. J., Pritchard, M. S., Burt, M. A., Branson, M. D., & Randall, D. A. (2016). Robust effects of cloud superparameterization on simulated daily rainfall intensity statistics across multiple versions of the Community Earth System Model. *Journal of Advances in Modeling Earth Systems*, *8*, 140–165. https://doi.org/10.1002/2015MS000574

Kooperman, G. J., Pritchard, M. S., O'Brien, T. A., & Timmermans, B. W. (2018). Rainfall from resolved rather than parameterized processes better represents the present-day and climate change response of moderate rates in the Community Atmosphere Model. *Journal of Advances in Modeling Earth Systems*, *10*, 971–988. https://doi.org/10.1002/2017MS001188

Li, F., Rosa, D., Collins, W. D., & Wehner, M. F. (2012). "Super-parameterization": A better way to simulate regional extreme precipitation? *Journal of Advances in Modeling Earth Systems*, *4*, M04002. https://doi.org/10.1029/2011MS000106

Marshall, J. S., & Palmer, W. M. K. (1948). The distribution of raindrops with size. *Journal of Meteorology*, *5*(4), 165–166.

McCrary, R. R., Randall, D. A., & Stan, C. (2014). Simulations of the West African monsoon with a superparameterized climate model. Part II: African easterly waves. *Journal of Climate*, *27*(22), 8323–8341.

Milbrandt, J., & Yau, M. (2005). A multimoment bulk microphysics parameterization. Part I: Analysis of the role of the spectral shape parameter. *Journal of the Atmospheric Sciences*, *62*(9), 3051–3064.

von Mises, R. (1954). La distribution de la plus grande de n valeurs. In (Ed.), *Selected Papers* (Vol. II, pp. 271–294). Providence, RI: American Mathematical Society.

Morrison, H., Curry, J., & Khvorostyanov, V. (2005). A new double-moment microphysics parameterization for application in cloud and climate models. Part I: Description. *Journal of the Atmospheric Sciences*, *62*(6), 1665–1677.

Neale, R., Chen, C., & Gettelman, A. (2010a). Description of the NCAR Community Atmosphere Model (CAM 4.0) NCAR, TN-485+ STR. http://www.cesm.ucar.edu/models/ccsm4.0/cam/docs/description/cam4_desc.pdf

Neale, R. B., Chen, C.-C., Gettelman, A., Lauritzen, P. H., Park, S., Williamson, D. L., et al. (2010b). Description of the NCAR Community Atmosphere Model (CAM 5.0) (*Note NCAR/TN-486+ STR*): NCAR. 1(1), 1-12.

O'Brien, T. A., Collins, W. D., Kashinath, K., Rübel, O., Byna, S., Gu, J., et al. (2016). Resolution dependence of precipitation statistical fidelity in hindcast simulations. *Journal of Advances in Modeling Earth Systems*, *8*, 976–990. https://doi.org/10.1002/2016MS000671

Ooyama, K. V. (2001). A dynamic and thermodynamic foundation for modeling the moist atmosphere with parameterized microphysics. *Journal of the Atmospheric Sciences*, *58*(15), 2073–2102.

Pall, P., Aina, T., Stone, D. A., Stott, P. A., Nozawa, T., Hilberts, A. G., et al. (2011). Anthropogenic greenhouse gas contribution to flood risk in England and Wales in Autumn 2000. *Nature*, *470*(7334), 382.

Pall, P., Patricola, C. M., Wehner, M. F., Stone, D. A., Paciorek, C. J., & Collins, W. D. (2017). Diagnosing conditional anthropogenic contributions to heavy Colorado rainfall in September 2013. *Weather and Climate Extremes*, *17*, 1–6.

Parishani, H., Pritchard, M. S., Bretherton, C. S., Terai, C. R., Wyant, M. C., Khairoutdinov, M., & Singh, B. (2018). Insensitivity of the cloud response to surface warming under radical changes to boundary layer turbulence and cloud microphysics: Results from the ultraparameterized CAM. *Journal of Advances in Modeling Earth Systems*, *10*, 3139–3158. https://doi.org/10.1029/2018MS001409

Parishani, H., Pritchard, M. S., Bretherton, C. S., Wyant, M. C., & Khairoutdinov, M. (2017). Toward low-cloud-permitting cloud superparameterization with explicit boundary layer turbulence. *Journal of Advances in Modeling Earth Systems*, *9*, 1542–1571. https://doi.org/10.1002/2017MS000968

Phillips, V. T., & Donner, L. J. (2006). Cloud microphysics, radiation and vertical velocities in two- and three-dimensional simulations of deep convection. *Quarterly Journal of the Royal Meteorological Society*, *132*(621C), 3011–3033.

Pickands III, J. (1971). The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability*, *8*(4), 745–756.

Pickands III, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, *3*(1), 119–131.

Rauscher, S. A., O'Brien, T. A., Piani, C., Coppola, E., Giorgi, F., Collins, W. D., & Lawston, P. M. (2016). A multimodel intercomparison of resolution effects on precipitation: Simulations and theory. *Climate Dynamics*, *47*(7-8), 2205–2218.

Redelsperger, J.-L., Brown, P., Guichard, F., How, C., Kawasima, M., Lang, S., et al. (2000). A GCSS model intercomparison for a tropical squall line observed during TOGA. -COARE. I: Cloud-resolving models, *Quarterly Journal of the Royal Meteorological Society*. *126*(564), 823–863.

Risser, M. D., Paciorek, C. J., Wehner, M. F., O'Brien, T. A., & Collins, W. D. (2018). A probabilistic gridded product for daily precipitation extremes over the United States. arXiv preprint arXiv:1807.04177.

Rowe, A. K., Rutledge, S. A., & Lang, T. J. (2012). Investigation of microphysical processes occurring in organized convection during NAME. *Monthly Weather Review*, *140*(7), 2168–2187.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., et al. (2010). The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, *91*(8), 1015–1058.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP climate forecast system version 2. *Journal of Climate*, *27*(6), 2185–2208.

Shima, S.-i., Kusano, K., Kawano, A., Sugiyama, T., & Kawahara, S. (2009). The super-droplet method for the numerical simulation of clouds and precipitation: A particle-based and probabilistic microphysics model coupled with a non-hydrostatic model. *Quarterly Journal of the Royal Meteorological Society: A Journal of the Atmospheric Sciences, Applied Meteorology and Physical Oceanography*, *135*(642), 1307–1320.

Singh, M. S., & O'Gorman, P. A. (2014). Influence of microphysics on the scaling of precipitation extremes with temperature. *Geophysical Research Letters*, *41*, 6037–6044. https://doi.org/10.1002/2014GL061222

Telford, J. (1955). A new aspect of coalescence theory. *Journal of Meteorology*, *12*(5), 436–444.

Van Weverberg, K., Vogelmann, A. M., Morrison, H., & Milbrandt, J. A. (2012). Sensitivity of idealized squall-line simulations to the level of complexity used in two-moment bulk microphysics schemes. *Monthly Weather Review*, *140*(6), 1883–1907.

Verlinde, J., & Cotton, W. R. (1993). Fitting microphysical observations of nonsteady convective clouds to a numerical model: An application of the adjoint technique of data assimilation to a kinematic model. *Monthly Weather Review*, *121*(10), 2776–2793.

Wilcox, E. M., & Donner, L. J. (2007). The frequency of extreme rain events in satellite rain-rate estimates and an atmospheric general circulation model. *Journal of Climate*, *20*(1), 53–69.

Wilhelmson, R. (1974). The life cycle of a thunderstorm in three dimensions. *Journal of the Atmospheric Sciences*, *31*(6), 1629–1651.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, *9*(1), 60–62.

Wilks, D. S. (2016). "The stippling shows statistically significant grid points": How research results are routinely overstated and overinterpreted, and what to do about it. *Bulletin of the American Meteorological Society*, *97*(12), 2263–2273.

Zeng, X., Tao, W.-K., Lang, S., Hou, A. Y., Zhang, M., & Simpson, J. (2008). On the sensitivity of atmospheric ensembles to cloud microphysics in long-term cloud-resolving model simulations. *Journal of the Meteorological Society of Japan Series II*, *86*, 45–65.

Zhang, G. J., & McFarlane, N. A. (1995). Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian Climate Centre general circulation model. *Atmosphere Ocean*, *33*(3), 407–446.