How Well Do CMIP6 Historical Runs Match Observed Northeast US Precipitation and Extreme Precipitation-related Circulation?

By Laurie Agel[1], Mathew Barlow[1,2]

[1]Department of Environmental, Earth, and Atmospheric Sciences, University of Massachusetts Lowell, Lowell, MA

[2]Climate Change Initiative, University of Massachusetts Lowell, Lowell, MA

Corresponding Author: Laurie Agel, Department of Environmental, Earth, and Atmospheric Sciences, University of Massachusetts Lowell, One University Avenue, Lowell, MA 01854, Email: Laurie_Agel@uml.edu

Early Online Release: This preliminary version has been accepted for publication in *Journal of Climate*, may be fully cited, and has been assigned DOI 10.1175/JCLI-D-19-1025.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

Abstract

Sixteen Coupled Model Intercomparison Project Phase 6 (CMIP6) historical simulations (1950–2014) are compared to Northeast US observed precipitation and extreme precipitation-related synoptic circulation. A set of metrics based on the regional climate is used to assess how realistically the models simulate the observed distribution and seasonality of extreme precipitation, as well as the synoptic patterns associated with extreme precipitation. These patterns are determined by *k*-means typing of 500-hPa geopotential heights on extreme precipitation days (top 1% of days with precipitation). The metrics are formulated to evaluate the models' extreme precipitation spatial variations, seasonal frequency, and intensity; and for circulation, the fit to observed patterns, pattern seasonality, and pattern location of extreme precipitation.

Based on the metrics, the models vary considerably in their ability to simulate different aspects of regional precipitation, and a realistic simulation of the seasonality and distribution of precipitation does not necessarily correspond to a realistic simulation of the circulation patterns (reflecting the underlying dynamics of the precipitation), and vice versa. This highlights the importance of assessing both precipitation and its associated circulation. While the models vary in their ability to reproduce observed results, in general the higher resolution models score higher in terms of the metrics. Most models produce more frequent precipitation than that for observations, but capture the seasonality of precipitation intensity well, and capture at least several of the key characteristics of extreme precipitation-related circulation. These results do not appear to reflect a substantial improvement over a similar analysis of selected CMIP5 models.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

1. Introduction

The Northeast US is a region that experiences heavy rainfall throughout the year, due to tropical systems and convective events in the summer, and strong extratropical storms throughout the year (Hoskins and Hodges 2002, Hawcroft 2012, Agel et al. 2015, Barlow 2011, Howarth 2019). The region is susceptible to storms that track from the Great Lakes and the Central US, as well as coastal storms, that travel up the East Coast and impact the area with subtropical moisture feeds and strong surface low pressure (Collow et al. 2016, Collins et al. 2014). In addition, recent studies have shown that precipitation is increasing in this region in recent decades, and is expected to continue to do so in accordance with climate change (IPCC 2014; Easterling et al. 2017). Because of these vulnerabilities, it is important to accurately interpret climate model projections for this region. We ask two key questions: which climate models best simulate the various traits of Northeast US precipitation and extreme precipitation, and do they do so for the "right" reasons (that is, under similar synoptic regimes)?

Release of the Coupled Model Intercomparison Project Phase6 (CMIP6; Eyring et al. 2016) data sets has recently begun. This effort aims to build on the previous CMIP Phase 5 (CMIP5; Taylor et al. 2012) experiments, which are part of a long-term effort by the World Climate Research Programme (WCRP)'s Working Group of Coupled Modelling (WGCM) to advance our understanding of the complete Earth system. The goal of CMIP is to provide a framework of common experiment protocols and forcings, and prescribed output to the climate science community, which will lead to increased process understanding in many areas including clouds, aerosols, and internal variability. Improvements from the preceding experiment (CMIP5) are expected particularly for

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

decadal predictions, based on improvements in the models, as well as the methods of initialization and ensemble generation. As such, the CMIP6 model suite provides a rich data set through which to examine our key questions, and to compare to solutions generated by the CMIP5 models.

Previously, Colle et al. (2013) investigated CMIP5 models for their ability to reproduce eastern North American and western North Atlantic cyclone genesis, tracks, rate of development, and intensity, and found that resolution played a large role in the model performance. Fereday et al. (2018) also recognized circulation variability between CMIP5 models to be a key player in precipitation variations for the North Atlantic and European regions. For the Northeast, Karmalkar et al. (2019) evaluated CMIP5 monthly precipitation and temperature (1950–2005) against a set of process-based metrics. Although no single model performed well for every metric described, they identified a subset of 16 models that generated “credible” and “diverse” simulations of precipitation and associated circulation.

Previously, we assessed Northeast US precipitation and extreme precipitation for the CMIP5 model suite. In that study, we identified four patterns of 500-hPa geopotential heights associated with extreme precipitation for each of 14 models. Northeast extreme precipitation and extreme precipitation-related circulation has been previously examined using pattern analysis, by Ning and Bradley (2014), Roller et al. (2016), Collow et al. (2016), and Agel et al. (2018, 2019). Pattern-based analysis techniques associated with extreme precipitation are additionally reviewed in Barlow et al. (2019). Here, we use the same technique with a newly-available sampling of CMIP6 models, and explore how well the models meet certain metrics based on observed precipitation and extreme

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

precipitation circulation patterns. The identical metrics are used here as in the previous study, in order to address a third key question: does the CMIP6 model suite provide an improvement over the CMIP5 model suite in terms of simulating representative aspects of Northeast US precipitation?

Our method for exploring these questions involves 1) establishing key characteristics of observed Northeast US precipitation, including seasonal frequency and intensity, as well as regional characteristics, 2) identifying observed extreme precipitation days, and 3) creating a set of observed circulation patterns that occur in conjunction with extreme precipitation, and identifying key aspects of this circulation. These key characteristics are combined into a set of metrics by which we evaluate CMIP6 "historical run" model output. This study is organized as follows: data and methods are presented in Section 2, results are presented in Section 3, and a summary and conclusion are presented in Section 4.

2. Data and Methods

a. Observed Data

The National Oceanic and Atmospheric Administration (NOAA) Climatological Prediction Center's Unified daily gridded precipitation product (CPCU; Chen et al. 2008), based on daily station data and subjected to a number of quality control checks, and available on a 0.25° x 0.25° grid from 1950–present, is used to calculate Northeast US daily precipitation intensity and extreme precipitation (99^{th} percentile for days with precipitation over 0.2 mm, 1980–2017) at each grid point within the Northeast US (Maine, New Hampshire, Vermont, New York, Massachusetts, Connecticut, Rhode Island, New Jersey, Pennsylvania, Delaware, Maryland, and West Virginia). This results

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

in 3009 days where extreme precipitation occurs concurrently at one or more grid locations. In addition to the top 1% thresholds, we also compute monthly cycles of precipitation and extreme precipitation frequency and intensity at each grid point. Although gridded precipitation often overestimates precipitation frequency and underestimates intensity compared to point sources (Chen and Knutsen 2008), we find that this gridded dataset is effective at qualitatively capturing the precipitation characteristics we examine here.

National Aeronautics and Space Administration (NASA) Modern Era Retrospective Reanalysis for Research and Application (MERRA-2; Gelaro et al. 2017) 500-hPa geopotential heights and mean sea-level pressure (MSLP) are used to represent observed circulation on extreme precipitation days. The daily means (1980–2017) for each field are used, and converted to anomalies by removing the long-term daily mean (i.e. the mean of 01-Jan, 02-Jan, etc.) at each grid point. The long-term-daily mean is smoothed with a 14-day running mean.

Although we use a single precipitation dataset (CPCU) and reanalysis dataset (MERRA-2) for this study, we have used these datasets in tandem for multiple Northeast studies (Roller et al. 2016, Agel et al. 2018, Agel et al. 2019a, Agel et al. 2019b), and find the products to provide realistic analysis, which is both consistent with and complementary to other studies done by other researchers, including Collow et al. (2016), Ning and Bradley (2014), and Howarth et al. (2019).

b. CMIP6 data

Model precipitation and circulation for 16 CMIP6 “r1i1p1f1” historical daily simulations are used, including the 500-hPa geopotential height fields, MSLP, and

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

precipitation flux fields, for the years 1950–2014. The models are listed in Table 1, in order of decreasing resolution. For the purposes of this study, we consider climate models with resolution below 1.0° as "high-resolution" (3 models), those between 1.0–2.0° as "medium-resolution" (9 models) and those over 2.0° as "low-resolution" (4 models). The models range from the high-resolution CNRM-CM6-1-HR and EC-Earth3 to the low-resolution BCC-ESM1 and CanESM5. The https://es-doc.org webpage contains expanded information for each data set, including the atmospheric, ocean, land, and ice components, as well as the physics and moist process parameterizations. The datasets are processed identically to that for the observations, where extreme precipitation is determined at each model grid point by the 99th percentile of days with precipitation over 0.2 mm. The number of model grid points in the domain, the mean 99th-percentile threshold, and the unique number of extreme days for all grid points are shown in Table 1. As for observations, monthly cycles of precipitation and extreme precipitation frequency and intensity are also calculated.

c. Typing

K-means typing (Diday and Simon 1976, Michelangeli et al. 1995) is performed on MERRA-2 500-hPa geopotential heights for the 3009 extreme precipitation days (identified in Section 2a), as well as on the CMIP6 models' 500-hPa geopotential heights for the models' extreme precipitation days, within the area bounded by 30–50°N and 90–60°W, using MATLAB's built-in "kmeans" function. Before processing, the long-term daily mean is removed at each grid point, and the field is reduced through empirical orthogonal function (EOF) analysis to 90% of its variance.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

K-means typing is a technique to separate input data into non-overlapping clusters, where individual input data is assigned to a cluster based on nearest Euclidean distance to the cluster centroid (the mean of the inputs assigned to the cluster). The centroid is then recalculated, and the process is reiterated until further iterations no longer reduce the sum of the intra-cluster variances.

To determine a reasonable number of clusters, k-means is applied for k=1..8, and the most reproduceable clustering is found using the method of Michelangeli et al. (1995). In this method, a "Classifiability Index" (CI) is determined for each k, based on the mean anomaly cross-coefficient between a particular cluster in a single partitioning to each cluster in every other partitioning, over a large number of partitionings. The resulting CI is compared to that produced using random red noise based on the input field, so that any CI greater than the 90th percentile of the red-noise results represents a k that is consistently reproduceable across a large number of iterations. For this study, the CI test for CPCU/MERRA-2 suggests k=4 and k=6 to be the best choices. Further examination shows that the 6-pattern solution breaks two of the k=4 solution patterns into two subsets each. These subsets do not substantively change the results of this study, therefore we use the k=4 solution to simplify and streamline the analysis. *K*-means is subsequently applied to each of the CMIP6 models using k=4, and the results are compared to those for CPCU/MERRA-2.

d. Additional Data Notes

We note that resolution is much higher for the observed precipitation and circulation fields than for each of the CMIP6 models. This can make direct comparison of precipitation characteristics problematic (Gehne et al. 2016). For most studies,

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

observations must first be regridded to the resolution of a climate model before comparison. However, the specific characteristics we examine here (mean top 1% threshold and seasonal cycles of precipitation intensity and frequency) are insensitive to regridding (that is, the mean results are nearly identical whether or not we regrid observations to model resolution). Furthermore, CPCU has coverage for only US land. Regridding near coastlines, the Great Lakes, and Canada result in data loss along the region's borders, which affects the variability of the underlying observed data, if not the mean. For this reason, we compare the observations to model output without regridding.

We also note that the time period used for the CMIP6 historical runs (1950–2014) differs from that for CPCU/MERRA-2 observations (1980–2017). While there are likely underlying trends in the data, we find that the mean top 1% thresholds, and cycles of precipitation frequency and intensity are nearly identical between 1950–2014 and 1980–2017 for CPCU, as well as for the CMIP6 models between 1950–2014 and 1980–2014. In addition, there are only minor differences in the 10^{th}–90^{th}-precentile values for precipitation intensity and frequency. Because underlying trends do not have a substantial impact on our results, we use different time periods for observations and models to maximize our sample sizes.

3. Results

3.1 Observations

Characteristics of observed precipitation, based on CPCU gridded precipitation, 1950–2017, are shown in Figure 1. The grid density and extreme precipitation threshold are shown in Figure1a, and 1b, respectively. The extreme precipitation threshold increases from approximately 30 mm day^{-1} in the northwest to approximately 60 mm day^{-}

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

[1] to the southeast. This gradient is an important factor in determining Northeast US precipitation climatology (Agel et al. 2015), allowing for a separate coastal and inland climatology.

The monthly precipitation frequency, daily intensity aggregated by month, and total monthly precipitation is shown for all precipitation in Figure 1c and extreme precipitation in Figure 1d. Precipitation occurrence peaks in summer and Dec–Jan, with a peak in intensity during the warm months. Although the frequency of extreme precipitation peaks during late summer, the intensity of extreme precipitation tends to be consistently around 50 mm day^{-1} regardless of month. We note that Figure 1 panels c–d show the mean of all grid locations – a more nuanced monthly climatology separated by subregion can be found in Agel et al. (2015). For the purposes of this study, we will compare the CMIP6 model results to observations using the mean of all grid locations, and account for the coastal/inland differences using the gradient of extreme threshold (Figure 1b).

K-means typing of MERRA-2 500-hPa geopotential heights, 1980–2017, on observed extreme precipitation days reveals 4 patterns (Figure 2a). The first (top left, labeled O1, 43.4% of extreme days) exhibits nearly zonal circulation, with a slight troughing to the east of the domain. The second (top right, labeled O2, 22.4%) exhibits slight ridging with anomalously high heights to the east of the domain. The third pattern (bottom left, labeled O3, 21.8%) features a trough/ridge couplet, with the trough draped from the Great Lakes south to Louisiana, and a ridge over the ocean to the east of Massachusetts. The fourth pattern (bottom right, labeled O4, 12.4%) features a deep trough across the Ohio Valley, with surface low pressure centered over New England.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

The favored locations for extreme precipitation (dots) within each pattern are shown in Figure 2b, along with anomalous precipitation (shaded). O1 features the least intense extreme precipitation, which occurs in two locations - along the spine of the Appalachians in Pennsylvania and West Virginia, and in the extreme north regions of the domain along the Canadian border. For O2, the majority of extremes occur in the southwestern portions of the domain. For O3, which features the most widespread and heaviest precipitation, most extremes occur in the center of the domain, and for O4, the extremes occur predominately in Maine and along the far eastern coast of northern New England. Grey lines in Figure 2b separate the domain into 4 regions, which we use to evaluate how well the models capture the extreme locations per pattern type.

The seasonal frequency of each pattern is shown in Figure 2c, where red (blue) bars indicate frequencies higher (lower) than expected based on random sampling. Pattern O1 occurs more frequently than expected during JJA, and less frequently than expected for other seasons, while O2, O3 and O4 exhibit the opposite behavior – occurring less frequently than expected during JJA, and more frequently than expected during the other seasons.

To explore how well the observed patterns reflect circulation on the days assigned to the patterns, Figure 2d shows histograms of the spatial correlations of 500-hPa height anomalies on individual days to the assigned anomaly pattern. The highest correlations occur for pattern O3 (non-summertime trough/ridge couplet), while the lowest correlations occur for pattern O1 (summertime slight trough). Histograms of root-mean-squared-error (RMSE) are shown in Figure 2e. Since the *k*-means algorithm used here assigns days to patterns based on minimum RSME, it follows that cluster centroids with

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

smaller RMSEs are more representative of the underlying days. Here, we find O1 (summertime slight trough) to have slightly better matching to the underlying days than the other patterns.

3.2 CMIP6 models

For each CMIP6 model, a similar analysis is done as for observations. Precipitation flux is analyzed to create a set of extreme precipitation days, that is, days where precipitation is higher than the 99th percentile of all days with precipitation greater than 0.2 mm for one or more grid points. The number of grid points per model within the Northeast domain is listed in Table 1. The regional thresholds for extremes and the monthly frequency and intensity are examined in terms of how well these match observations. Next, the model 500-hPa heights for these days are separated into four patterns using *k*-means, as for observations, and these are compared to those related to observed extremes/patterns. We ask 1) how well does the model simulate Northeast US precipitation, and 2) how well does the model capture the four main circulation patterns associated with Northeast US extreme precipitation? We create a set of 6 precipitation-related metrics and 12 circulation-related metrics (3 metrics per each of 4 patterns) to objectively examine how well the models capture key characteristics of precipitation and related circulation that are representative of Northeast observations. The metrics are identical to those used to examine the CMIP5 model suite, and are listed in Table 2.

The results of comparing the 16 models' output to observations based on the Table 2 metrics are summarized in Figure 3. Metrics that are reasonably met by the model are shown with a green dot. The average "score" (number of green dots) for the precipitation metrics is 3.1 out of 6 (results range from 0 to 5); while the average score

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

for the circulation metrics is 8.2 out of 12 (ranging from 5 to 12). The mean total score is 11.3 out of 18. Clearly, no individual model meets all metrics, and skill at reproducing precipitation characteristics does not necessarily predict skill at reproducing circulation characteristics, and vice versa.

The CNRM-CM6-1-HR model compares the best to observational metrics, with a total score of 16 out of 18; while CNRM-CM6-1 and MPI-ESM1-2-HR both have scores of 15. Other models that simulate observations well based on these metrics include ACCESS-CM2, EC-Earth3, and HadGEM3-CG21-LL, with total scores of 13. However, EC-Earth3, despite scoring well for circulation metrics, scores low for the precipitation metrics (2 out of 6), while ACCESS-CM2 scores better for the precipitation metrics (5 out of 6) than for the circulation metrics (8 out of 12). The poorest performing models for these metrics include NorESM2-LM and BCC-ESM1, with total scores of 8 or less.

Resolution appears to play a role in how well the models capture the combined precipitation and circulation characteristics, with the three high-resolution models in the top third and the four low-resolution models in the bottom third of the total metric scores. The relationship to resolution is weaker when looking at precipitation or circulation metrics alone. For precipitation metrics, the medium-resolution MIROC6 and BCC-CSM2-MR model score better than high-resolution MPI-ESM1-2-HR and EC-Earth3 models. For the circulation metrics, BCC-CSM2-MR (medium-resolution) performs worse than all four low-resolution models, while NorESM2-LM (low-resolution) scores as well as or better than many of the medium-resolution models. The ACCESS-CM2 and MPI-ESM1-2-HR models are discussed in detail below, as examples of models that simulate observed extreme precipitation well (but not necessarily the related circulation),

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

and those that simulate observed circulation on extreme days well (but not necessarily the extreme precipitation itself), respectively.

Precipitation and related circulation characteristics for ACCESS-CM2 are shown in Figures 4 and 5. Despite having lower resolution than observations (Figure 4a), the areal-mean top 1% threshold is reasonable, and the northwest-southeast gradient in precipitation is similar to observations (Figure 4b). However, precipitation near the Great Lakes appears to be too intense. While the model produces too many days of precipitation in all months but December and January, the daily intensity matches observations well (Figure 4c). The model also matches observations well for extreme precipitation seasonal frequency and intensity (Figure 4d). Visually, the circulation patterns associated with extreme precipitation (labeled P1–P4, Figure 5a) have key differences with observational patterns. Specifically, there appears to be a shortwave in the flow across the southeastern states for P2, the ridging over the Northeast is much stronger than in observations for P3, and the deep trough in P4 is located too far west. The location of anomalous precipitation is similar to observations, but the location of extremes in P3 is concentrated farther south (Figure 5b). For P2, there is no significant decrease in the frequency of JJA dates, as for observations, and there are less DJF and SON dates by percentage than for observations (Figure 5c). While not explored here, this may be related to the shortwave in the 500-hPa flow, which is relevant to the generation of precipitation extremes (Agel et al. 2019a). The presence of the shortwave in otherwise zonal flow may cause more of these fields to be grouped into O2-like patterns as opposed to O1-like patterns by the clustering algorithm. Finally, Figure 5d explores how well P1–P4 match O1–O4 in terms of RMSE and spatial correlation. Results that are significantly

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

lower for RMSE or higher for correlation than expected by chance (.05 level of significance), as determined by random sampling, are indicated by asterisks. RMSE between P1/O1 and P2/O2 are lower than between P1 and O2/O3/O4, and P2 and O1/O3/O4, as we would expect. However, RMSE between P3/O3 is not much lower than that between P3/O2, and RMSE between P4/O3 is lower than P4/O4. Similarly, correlations between P1/O1, P2/O2 are highest, but correlation between P4/O4 is less than that between P4/O3, and while P3/O3 correlation is the highest, it is not significantly higher than that due to chance, and is very close in value to P3/O2 (which is significantly higher than expected by chance). In summary, although ACCESS-CM2 precipitation characteristics are similar to observations, the circulation associated with extreme precipitation has some key differences from observations. It is beyond the purposes of this study to ascertain why this occurs, but possibilities include model feedback mechanisms which enhance troughs and ridges during extreme precipitation, or model physics and parameterizations that only produce extreme precipitation under the conditions of enhanced synoptic flow.

Characteristics of precipitation/circulation for MPI-ESM1-2-HR are shown in Figures 6 and 7. Despite the high resolution of this model, the model does not fully capture the northwest-to-southeast gradient of precipitation (Figure 6b). While the inland values for the top 1% threshold are reasonable, the coastal values are much lower than for observations. The monthly frequency of precipitation is too high, but the daily intensities of precipitation (Figure 6c) and extreme precipitation (Figure 6d) match observations well. The four model patterns associated with extreme precipitation are shown in Figure 7a. The patterns are visually similar to observations, except for P2, which has more

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

enhanced ridging over the Northeast, and P4, which features a deeper trough. Anomalous precipitation over land is slightly higher than observations, but is qualitatively similar, in terms of where the heaviest precipitation occurs (Figure 7b). Spatially, the location of extremes is similar to observations. Seasonally, the extreme pattern frequencies match observations, in that P1 occurs more frequently than expected due to chance during JJA, while the other patterns occur less frequently than expected during JJA (Figure 7c). The patterns match those from observations well, based on the RMSE values and spatial correlation values between the model patterns and the observational patterns (Figure 7d). The lowest RSME values and highest positive correlation values occur between P1/O1, P2/O2, P3/O3, and P4/O4, as we would expect. The correlation value for P1/O1 is not significantly higher than expected by chance, but that is not surprising for the predominantly zonal pattern, where small variations in anomalous flow can cause large correlation differences. In this case, RMSE may be a better overall measure of fit. In summary, MPI-ESM1-2-HR appears to produce less heavy precipitation than observations, particularly along the coast; however, the heavy precipitation appears to be generated within similar circulation constraints to observations.

Similar figures for all 16 models are available in Supplemental Information. Overall, BCC-ESM1, EC-Earth3, and NorESM2-LM all produce noticeably less heavy precipitation than observations, as can be seem in the top 1% threshold values and daily intensity values; while too much heavy precipitation is produced by CanESM5 and MIROC6 inland, HadGEM3-CG31-LL throughout New Jersey and Delaware, ACCESS-CM2 along the coast, and IPSL-CM6A-LM throughout the domain. CNRM-CM6-1-HR (the highest-resolution model examined here) shows the closest match to observations for

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

the top 1% values and regional gradient. All models produce too many days of precipitation, but several show reasonable seasonal cycles, including ACCESS-CM2, CESM2, CESM2-WACCM, CNRM-CM6-1-HR, MIROC6, and NorESM2. In contrast, CanESM5 produces too much summer precipitation, while EC-Earth3, MPI-ESM1-2-HR, and MRI-EMS2-0 produce too much spring precipitation. Daily intensity is simulated well by ACCESS-CM2, CNRM-CM6-1-HR, MIROC6, and MRI-ESM2-0; while other models struggle to match observations. BCC-ESM1 and BCC-CSM2-MR both are biased too low for each month, while CESM2 and CESM2-WACCM produce too little summer daily intensity, and IPSL-CM6A-LR produces too much May–June daily intensity.

For the circulation characteristics, CNRM-CM6-1, CNRM-CM6-1-HR, and EC-Earth3 reasonably reproduce observed patterns in terms of spatial correlation, pattern seasonality, and location of extreme precipitation within the patterns. While there is good visual matching between P1/O1 for all models, 9 out of 16 models do not match the metric for fit (correlation and RMSE) between P1/O1. This is likely due to poor correlation rather than low RMSE, which may be related to the zonal pattern itself, where anomalous flow can cause large deviations in correlation. All models meet the metric for fit between P2/O2, however this too is somewhat misleading: CESM2, HadGEM3-CG31-LL, IPSL-CM6A-LR, MIROC6, MPI-ESM1-2-HR, and NorESM2-LM all feature much more pronounced ridges over the Northeast than that observed in O2. The models show varied success in visual matching (and metric matching) for the ridge/trough in P3/O3 and the deep trough in P4/O4, which is likely related to the intensity and relative location of the ridge/trough in P3. In these cases, days with deeper and eastward-shifted troughs

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

may get split between P3 and P4 during the *k*-means separation, rather than all assigned to P3. While all models but BCC-ESM1 capture the observed location of extremes in P4/O4, only MPI-ESM1-2-HR and MRI-ESM2-0 capture the observed locations for P3/O3. Again, this is likely related to the relative location of the trough/ridge axis in P3, and how the *k*-means algorithm splits these days. While all models capture the relative seasonality of the P1/O1 and P3/O3 patterns, a number of models struggle with the seasonality for P4/O4. HadGEM3-CG31-LR, IPSL-CM6A-LR, and MIROC6 each have higher frequency in JJA than expected (whereas observations show lower frequency than expected), which is likely related to a shallower P4 trough than that for O4. A shallow trough across the Ohio Valley is a common summer pattern associated with extreme precipitation for the Northeast (Agel et al. 2017). These three models may generate extreme precipitation for shallower troughs in general, since they also overproduce heavy precipitation, as seen in the overdone top 1% thresholds.

3.3 Comparison to CMIP5 results

One of the main motivations for this study is to determine if the CMIP6 models improve the simulation of Northeast precipitation and associated circulation over the CMIP5 models, per the set of metrics devised here. Six of the CMIP6 model families examined here were also included in the CMIP5 study. Table 3 shows a summary of the results for CMIP6 compared to CMIP5. ACCESS-CM2, HadGEM3-CG31-LL, and NorESM2-LM perform about the same as their CMIP5 counterparts. Noticeably, model resolution does not improve between CMIP5 and CMIP6 for these models. For CMIP6 models with increased resolution compared to their CMIP5 counterparts, including CNRM-CM6-1-HR and MPI-ESM1-2-HR, scores increase 2–3 points overall, split

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

between the precipitation and circulation metrics. However, IPSL-CM6A-LR, (here with a higher resolution than its IPSL-CMIP5A-LR counterpart), only improves by one point for the precipitation metrics. In addition, CNRM-CM6-1, with no increase in resolution over CNRM-CM5, improves by 2 points, which is likely related to improvements in the physical parameterizations in the atmospheric and land model components (Voldoire et al. 2019).

Until additional datasets become available, it is not possible to compare all of the previously examined CMIP5 model families to their CMIP6 counterparts; however, we can make some general statements. The mean score for precipitation metrics does not change (~3 out of 8) between the CMIP5 and CMIP6 results, while the score for associated extreme precipitation circulation increases slightly from 10.9 to 11.3 out of 12. The mean resolution (latitude x longitude) for the models increases from a mean 1.72° x 2.26° for the CMIP5 models examined to a mean 1.33° x 1.58° for the CMIP6 models examined. Despite several of the higher resolution models meeting the study's metrics better, we cannot yet state with certainty that the overall higher resolution of the CMIP6 models appreciably increase the scores for these metrics above those for CMIP5.

4. Summary and Conclusions

In this study, we examine how well CMIP6 climate models simulate Northeast US precipitation and extreme precipitation, as well as extreme precipitation-related circulation, based on a set of four observationally-determined 500-hPa geopotential height patterns for observed extreme precipitation days. We establish a set of metrics that best capture key aspects of Northeast precipitation observations and circulation, and

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

evaluate each model within the framework of those metrics. In addition, we compare these results to those for a previous study that considered CMIP5 models.

Specifically, we examine 16 models with historical 'r1i1pf1p' geopotential heights and precipitation, 1950–2014. The results are varied in how well the models meet the different metrics. Some models simulate the seasonality and spatial distribution of precipitation reasonably well, but do not successfully simulate all aspects of the associated circulation and spatial/temporal characteristics of the established patterns for extreme precipitation. That is, the extreme precipitation is not produced via the same dynamical mechanisms as the corresponding observed extreme precipitation. This highlights the importance of assessing circulation in association with precipitation. Other models do not capture the key aspects of precipitation well, but do generate extreme precipitation within the context of the four observed circulation patterns. We do note that for all models, the *k*-means typing results are at least very broadly visually similar to the basic four observed patterns, whether or not each specific precipitation or circulation metric is met. The range of model limitations in reproducing both aspects of the precipitation and the associated circulations suggests that CMIP6 precipitation projections for the region should be considered very cautiously.

In general, higher resolution models simulate precipitation closer to observed precipitation. However, resolution is not an absolute predictor of success regarding the metrics used here – for example, the relatively high-resolution EC-Earth3 does not score well on the precipitation metrics despite scoring very well on the circulation metrics. Nevertheless, models with resolution finer than 1.0° scored overall better in both precipitation and circulation metrics.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

One of the important goals of this research is to evaluate the CMIP6 models relative to their CMIP5 counterparts. As a preliminary assessment, although the resolution on average increases in the suite of CMIP6 considered here, the performance is not substantially better in terms of the regional precipitation and circulation metrics. However, we have at this time evaluated only a subset of the CMIP6 data expected to be available. As more datasets become available, we expect to add to these results. Additionally, as a starting point, this analysis has focused on four basic extreme-precipitation circulation patterns spanning the whole year. More detailed, season-specific analysis would be useful follow-on work.

Data Availability

CPCU data is downloaded from ftp://ftp.cdc.noaa.gov/Projects/Datasets/cpc_us_precip, as of November 2018. MERRA-2 data is downloaded from https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/data_access as of November 2018. CMIP6 model data is downloaded from https://esgf-node.llnl.gov/projects/cmip6, as of November 2019.

Acknowledgements

The work in this study is funded by National Science Foundation Project AGS 1623912.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

References

Agel, L., M. Barlow, F. Colby, H. Binder, J. L. Catto, A. Hoell, and J. Cohen, 2019a: Dynamical analysis of extreme precipitation in the US northeast based on large-scale meteorological patterns. *Climate Dyn*., **52**: 1739. https://doi.org/10.1007/s00382-018-4223-2.

Agel, L., M. Barlow, M. Collins, E. Douglas, and P. Kirshen, 2019b: Hydrometeorological Conditions Preceding Extreme Streamflow for the Charles and Mystic River Basins of Eastern Massachusetts. *J. Hydrometeor*., **20**, 1795–1812. https://doi.org/10.1175/JHM-D-19-0017.1.

Agel, L., M. Barlow, S. B. Feldstein, and W. J. Gutowski, 2018: Identification of large-scale meteorological patterns associated with extreme precipitation in the US northeast. *Climate Dyn.***,** **50**: 1819. https://doi.org/10.1007/s00382-017-3724-8.

Agel, L., M. Barlow, J.-H. Qian, F. Colby, E. Douglas, and T. Eichler, 2015: Climatology of Daily Precipitation and Extreme Precipitation Events in the Northeast United States. *J. Hydrometeor*., **16,** 2537-2557. https://doi.org/10.1175/JHM-D-14-0147.1.

Barlow, M., 2011: Influence of hurricane-related activity on North American extreme precipitation. *Geophysical Research Letters*, **38,** L04705, doi:10.1029/2010GL046258.

Barlow, M., and Coauthors, 2019: North American extreme precipitation events and related large-scale meteorological patterns: a review of statistical methods, dyanmics, modeling, and trends. *Clim. Dyn.,* in press.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

Chen, C.-T., and T. Knutson, 2008: On the Verification and Comparison of Extreme Rainfall Indices from Climate Models. *J. Climate*, **21,** 1605-1621.

Chen, M., P. Xie, and Co-authors, 2008: CPC Unified Gauge-based Analysis of Global Daily Precipitation. *Western Pacific Geophysics Meeting, Cairns, Australia, 29 July - 1 August, 2008*.

Colle, B. A., Z. Zhang, K. A. Lombardo, E. Chang, P. Liu, and M. Zhang, 2013: Historical Evaluation and Future Prediction of Eastern North American and Western Atlantic Extratropical Cyclones in the CMIP5 Models during the Cool Season. *J. Climate*, **26,** 6882-6903.

Collins, M. J., and Coauthors, 2014: Annual floods in New England (USA) and Atlantic Canada: synoptic climatology and generating mechanisms. Physical Geography, 35, 195-219.

Collow, A. B. M., M. G. Bosilovich, and R. D. Koster, 2016: Large-Scale Influences on Summertime Extreme Precipitation in the Northeastern United States. Journal of Hydrometeorology, 17, 3045-3061.

Diday, E., and J. C. Simon, 1976: Clustering Analysis. *Digital Pattern Recognition*, K. S. Fu, Ed., Springer Berlin Heidelberg, 47-94.

Easterling, D.R., and Coauthors, 2017: Precipitation change in the United States. In: *Climate Science Special Report*: Fourth National Climate Assessment, Volume I [Wuebbles, D.J., D.W. Fahey, K.A. Hibbard, D.J. Dokken, B.C. Stewart, and T.K. Maycock (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, pp. 207-230, doi: 10.7930/J0H993CC.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9,** 1937-1958.

Fereday, D., R. Chadwick, J. Knight, and A. A. Scaife, 2018: Atmospheric Dynamics is the Largest Source of Uncertainty in Future Winter European Rainfall. *J. Climate*, **31,** 963-977.

Gehne, M., T.M. Hamill, G.N. Kiladis, and K.E. Trenberth, 2016: Comparison of Global Precipitation Estimates across a Range of Temporal and Spatial Scales. *J. Climate*, **29**, 7773–7795, https://doi.org/10.1175/JCLI-D-15-0618.1

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Climate*, **30,** 5419-5454.

Hawcroft, M. K., L. C. Shaffrey, K. I. Hodges, and H. F. Dacre, 2012: How much Northern Hemisphere precipitation is associated with extratropical cyclones? *Geophysical Research Letters*, **39,** n/a-n/a.

Hoskins, B. J., and K. I. Hodges, 2002: New Perspectives on the Northern Hemisphere Winter Storm Tracks. Journal of the Atmospheric Sciences, 59, 1041-1061.

Howarth, M. E., C. D. Thorncroft, and L. F. Bosart, 2019: Changes in Extreme Precipitation in the Northeast United States: 1979–2014. Journal of Hydrometeorology, 20, 673-689.

IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

Karmalkar, A. V., J. M. Thibeault, A. M. Bryan, and A. Seth, 2019: Identifying credible and diverse GCMs for regional climate change studies—case study: Northeastern United States. *Climatic Change*, **153**(3), 367-386.

Michelangeli, P.-A., R. Vautard, and B. Legras, 1995: Weather Regimes: Recurrence and Quasi Stationarity. *J. Atmos. Sci.*, **52,** 1237-1256.

Ning, L., and R. S. Bradley, 2014: Winter Climate Extremes over the Northeastern United States and Southeastern Canada and Teleconnections with Large-Scale Modes of Climate Variability. *J. Climate*, **28,** 2475-2493.

Roller, C. D., J.-H. Qian, L. Agel, M. Barlow, and V. Moron, 2016: Winter Weather Regimes in the Northeast United States. *J. Climate*, **29,** 2963-2980.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An Overview of CMIP5 and the Experiment Design. *Bull. Amer. Meteor. Soc.*, **93,** 485-498.

Voldoire, A., D. Saint-Martin, S. Sénési, B. Decharme, A. Alias, M. Chevallier, et al. (2019). Evaluation of CMIP6 DECK experiments with CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, 11, 2177– 2213. https://doi.org/10.1029/2019MS001683

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

Table 1. CMIP6 models and observations (MERRA-2/CPCU) in order of decreasing resolution. The grid resolution is shown both in terms of latitude/longitude (degrees), but also in terms of the number of grid points that overlap the Northeast region. Also given are the top 1% precipitation threshold values (mm day^{-1}), and the number of unique extreme days 1980–2017. Asterisks indicate model families also considered in an earlier CMIP5 analysis.

Model/Observations	Lat.	Lon.	Number of grids	Extreme threshold (mm day^{-1})	Number of extremes
CPCU	0.25	0.25	925	40.72	3009
MERRA-2	0.50	0.63	n/a	n/a	n/a
CNRM-CM6-1-HR*	0.50	0.50	232	41.19	2655
EC-Earth3	0.70	0.70	122	36.28	2247
MPI-ESM1-2-HR*	0.94	0.94	62	39.70	1665
CESM2	0.94	1.25	51	39.32	1479
CESM2-WACCM	0.94	1.25	51	39.38	1383
BCC-CSM2-MR	1.12	1.12	48	39.20	1653
GFDL-CM4	1.00	1.25	48	41.26	1657
MRI-ESM2-0	1.12	1.12	48	38.79	1518
CNRM-CM6-1*	1.4	1.41	29	39.96	1405
MIROC6	1.40	1.40	29	45.43	1403
ACCESS-CM2*	1.25	1.88	24	44.82	1669
HadGEM3-CG31-LL*	1.25	1.88	24	47.26	1586
IPSL-CM6A-LR*	1.27	2.50	20	53.19	1553
NorESM2-LM*	1.89	2.50	11	31.97	675
BCC-ESM1	2.79	2.81	6	32.02	575
CanESM5	2.79	2.81	6	45.79	600

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

Table 2. Metrics used to determine how well CMIP6 model precipitation simulates observed precipitation (metrics 1–6), and how well *k*-means clustering of CMIP6 500-hPa geopotential heights on extreme precipitation days matches observed patterns of circulation on observed extreme precipitation days (metrics 7–18). The assessment criteria describe approximate correspondence to observations.

	Metric	Assessment Criteria
1	Top 1% threshold	Mean threshold within 25^{th}–75^{th} percentiles of obs thresholds
2	Range of top 1% thresholds	10^{th}–90^{th} percentile thresholds greater than obs 25^{th}–75^{th} percentile thresholds
3	Monthly prec frequency	At least 11 out of 12 months within 10^{th}-90^{th} percentile for obs
4	Monthly prec daily intensity	At least 11 out of 12 months within 10^{th}-90^{th} percentile for obs
5	Monthly extreme prec frequency	At least 11 out of 12 months within 10^{th}-90^{th} percentile for obs
6	Monthly extreme prec daily intensity	At least 11 out of 12 months within 10^{th}-90^{th} percentile for obs
7	P1: Spatial Distribution	Greater than 15% decrease from TQR to EQR in SE quadrant, where TQR=grids per quadrant/total grids, and EQR=extreme grids per quadrant/total extreme grids (or if no SE grids, no increases greater than 15% in any other quadrant)
8	P2: Spatial Distribution	Greater than 5% increase from TQR to EQR in NE and SE quadrants, and greater than 5% decrease in NW and SW
9	P3: Spatial Distribution	Less than |15%| difference from TQR to EQR in all quadrants
10	P4: Spatial Distribution	Greater than 15% decrease in SW quadrant from TQR to EQR, and greater than 15% increase in NE quadrant
11	P1: Seasonal Freq.	JJA *higher* than 5-95% confidence interval for all extreme days (not just P1)
12	P2: Seasonal Freq.	JJA *lower* than 5-95% confidence interval for all extreme days (not just P2)
13	P3: Seasonal Freq.	JJA *lower* than 5-95% confidence interval for all extreme days (not just P3)
14	P4: Seasonal Freq.	JJA *lower* than 5-95% confidence interval for all extreme days (not just P4)
15	P1->O1	P1→O1 corr/rmse at least 10% larger/smaller than P1→O2,O3,O4
16	P2->O2	P2→O2 corr/rmse at least 10% larger/smaller than P2→O1,O3,O4
17	P3->O3	P3→O3 corr /rmse at least 10% larger/smaller than P3→O1,O2,O4
18	P4->O4	P4→O4 corr/rmse at least 10% larger/smaller than P4→O1,O2,O3

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

Table 3. Comparison of resolution and metric scores between similar CMIP5 and CMIP6 models, and the overall score for all sampled CMIP5 (14 models) and CMIP6 models (16 models).

MODEL FAMILY	CMIP5					CMIP6				
	Lat	Lon	Prec	Circ	Tot	Lat	Lon	Prec	Circ	Tot
ACCESS1-0 / ACCESS-CM2	1.25	1.88	4	8	12	1.25	1.88	5	8	13
CNRM-CM5/ CNRM-CM6-1	1.40	1.41	4	9	13	1.40	1.41	5	10	15
CNRM-CM5/ CNRM-CM6-1-HR	1.40	1.41	4	9	13	0.50	0.50	5	11	16
HadGEM2-CC/ HadGEM3-CG31-LL	1.25	1.88	4	8	12	1.25	1.88	4	9	13
IPSL-CM5A-LR/ IPSL-CM6A-LR	1.89	3.75	1	7	8	1.27	2.5	2	7	9
MPI-EMS-LR/ MPI-ESM1-2-HR	1.87	1.88	3	10	13	0.94	0.94	3	12	15
NorESM1-M/ NorESM2-LM	1.90	2.50	0	8	8	1.89	2.50	0	8	8
All CMIP5 / All CMIP6	1.72	2.26	3.0	7.9	10.9	1.33	1.58	3.1	8.2	11.3

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

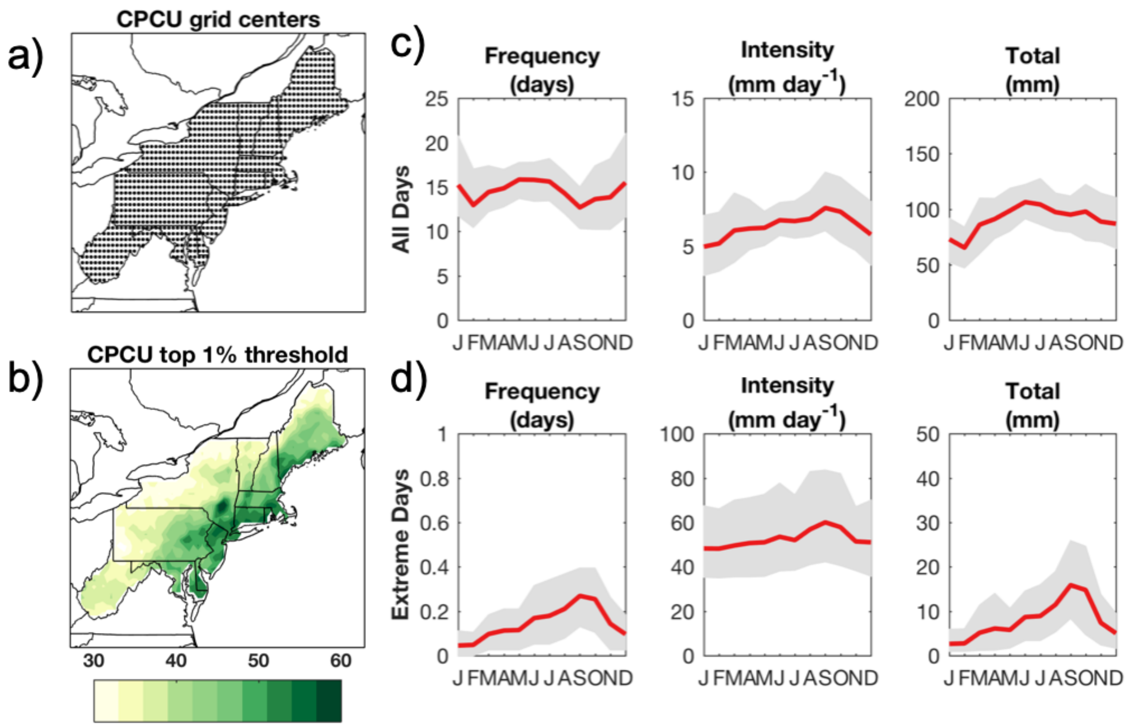


Figure 1. Observed precipitation (CPCU) characteristics, 1980–2017, with a) CPCU grid center locations, b) top 1% wet-day daily intensity threshold (shaded, in mm), c) grid-level mean wet-day monthly precipitation frequency (red line, in days), mean daily intensity (red line, in mm), and mean total daily precipitation (red line, in mm), and d) same as (c), but for extreme precipitation only. The grey shading for (c) and (d) represents the grid-level 10–90th percentile values.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

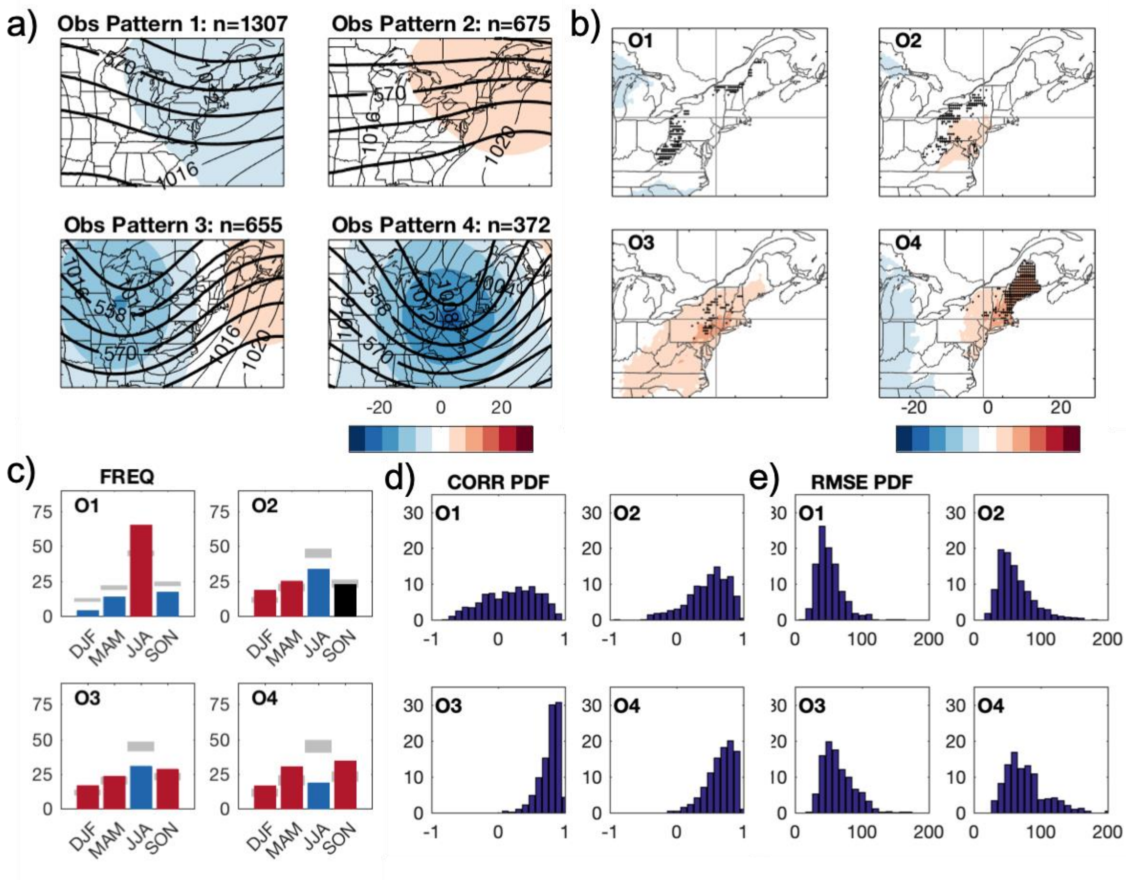


Figure 2. *K*-means separated (O1–O4) extreme precipitation a) patterns of 1980–2017 MERRA-2 500-hPa geopotential height anomalies (shaded) and total fields (thick black contours, in 6-dam increments) and MSLP (thin black contours, in 4-hPa increments), b) CPCU daily precipitation anomalies (shaded, in mm) and location of extreme precipitation (black dots, where each dot represents a grid location where the frequency of extremes exceeds 0.15%), and divided into 4 quadrants separated by grey lines, c) seasonal frequency of patterns, with frequency that is similar to, less than, or more than expected by chance represented by black, blue, and red bars, respectively, d) histograms of 500-hPa geopotential height spatial correlations of individual pattern days to pattern mean, and e) histograms of 500-hPa geopotential height RMSE (blue bars, in m) for individual patterns days to pattern mean.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

	Extr Threshold	Extr Gradient	Frequency	Intensity	Extr Frequency	Extr Intensity	Fit P1	Fit P2	Fit P3	Fit P4	Spatial P1	Spatial P2	Spatial P3	Spatial P4	Seasonal P1	Seasonal P2	Seasonal P3	Seasonal P4	Prec Count	Pattern Count	Total Count
CNRM-CM6-1-HR	●	●	×	●	●	●	●	●	●	●	●	●	×	●	●	●	●	●	5	11	16
CNRM-CM6-1	●	●	×	●	●	●	●	●	●	●	●	×	×	●	●	●	●	●	5	10	15
MPI-ESM1-2-HR	●	×	×	●	×	●	●	●	●	●	●	●	●	●	●	●	●	●	3	12	15
ACCESS-CM2	●	●	×	●	●	●	●	●	×	×	●	●	×	●	●	×	●	●	5	8	13
EC-Earth3	●	×	×	×	×	●	●	●	●	●	●	●	×	●	●	●	●	●	2	11	13
HadGEM3-CG31-LL	×	●	×	●	●	●	×	●	●	●	●	●	×	●	●	●	●	×	4	9	13
MRI-ESM2-0	●	×	×	●	×	●	●	●	×	×	●	×	●	●	●	●	●	●	3	9	12
GFDL-CM4	●	×	×	●	×	●	●	●	×	×	●	●	×	●	●	×	●	●	3	8	11
MIROC6	●	●	×	●	●	●	×	●	×	×	×	●	×	●	●	●	●	×	5	6	11
BCC-CSM2-MR	●	●	×	●	●	●	×	●	×	×	×	×	×	●	●	×	●	●	5	5	10
CESM2-WACCM	●	×	×	×	×	●	×	●	●	●	×	●	×	●	●	●	●	×	2	8	10
CanESM5	●	●	×	×	×	●	×	●	×	×	×	×	×	●	●	●	●	●	3	6	9
CESM2	●	×	×	×	×	●	×	●	×	●	×	●	×	●	●	●	●	×	2	7	9
IPSL-CM6A-LR	×	●	×	×	×	●	×	●	●	×	×	●	×	●	●	●	●	×	2	7	9
NorESM2-LM	×	×	×	×	×	×	×	●	●	●	●	×	×	●	●	●	●	×	0	8	8
BCC-ESM1	×	×	×	×	×	●	×	●	×	×	●	×	×	×	●	●	●	●	1	6	7

Figure 3. CMIP6 model ability to reproduce precipitation and extreme precipitation-related circulation based on metrics established in Table 2, where a green dot (black X) signifies the model met (did not meet) the criteria of the metric. There are 6 precipitation metrics, and 12 circulation metrics, 3 for each of 4 patterns (P1–P4). The two sets of metrics are separated by a thick black line. The three right columns show the total number of metrics that were met for precipitation, circulation, and combined metrics, respectively. Results are arranged in descending order by total number of metrics met.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

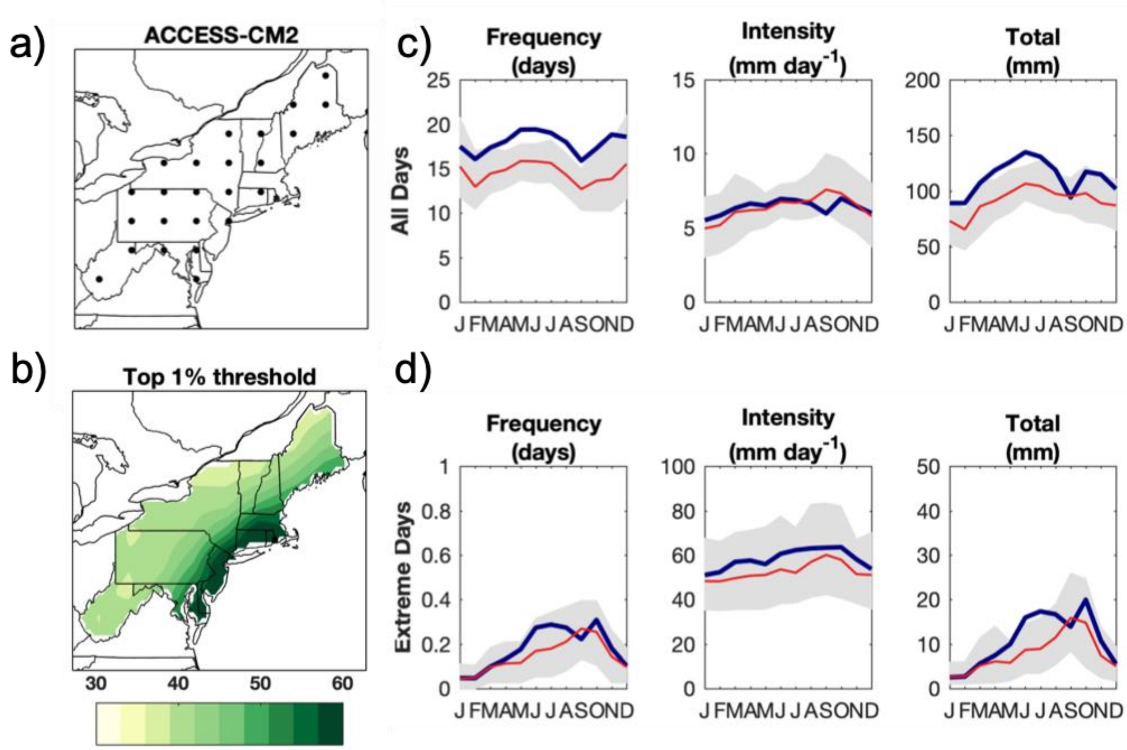


Figure 4. ACCESS-CM2 model precipitation characteristics, with a) grid center locations, b) top 1% wet-day daily intensity threshold (shaded, in mm), c) grid-level mean wet-day monthly precipitation frequency (blue line, in days), mean daily intensity (blue line, in mm), and mean total daily precipitation (blue line, in mm), and d) same as (c), but for extreme precipitation only. The red lines in (c) and (d) represent the observed results from Figure 1, while the grey shading represents the grid-level 10–90th percentile values for the observed results.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

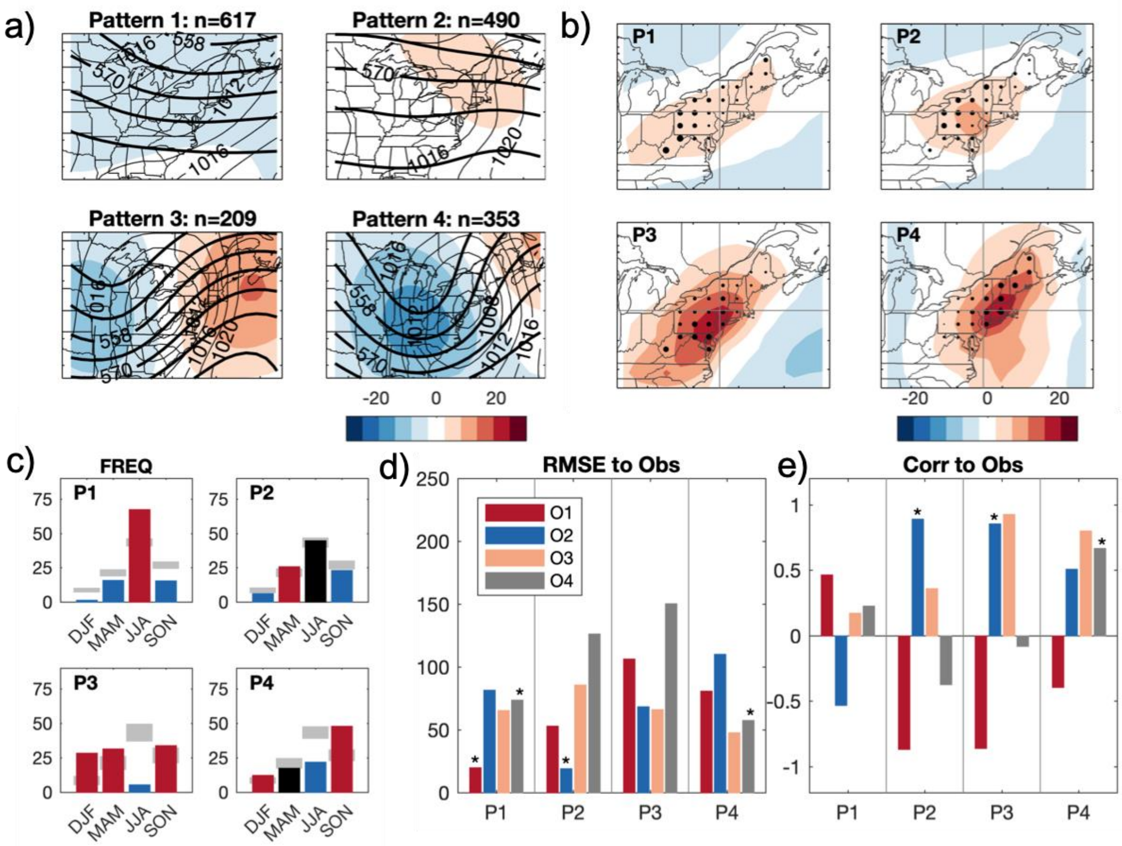


Figure 5. ACCESS-CM2 model *k*-means separated (P1–P4) extreme precipitation day a) patterns of 500-hPa geopotential height (anomalies shaded, and total fields shown as thick black contours, in 6-dam increments) and MSLP (thin black contours, in 4-hPa increments), b) daily precipitation anomalies (shaded, in mm) and location of extreme precipitation (dot size relative to number of days at grid location), c) seasonal frequency of patterns, with frequency that is similar to, less than, or more than expected by chance represented by black, blue, and red bars, respectively, d) bar charts of 500-hPa geopotential height RMSE between model patterns P1–P4 and observed patterns O1–O4, and 3) bar charts of 500-hPa geopotential height correlation between model patterns P1–P4 and observed patterns O1–O4. In (c) and (d), asterisks indicate values that are

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

statistically lower than expected (for RMSE) or higher than expected (for correlation), based on random sampling and a .05 level of significance.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.

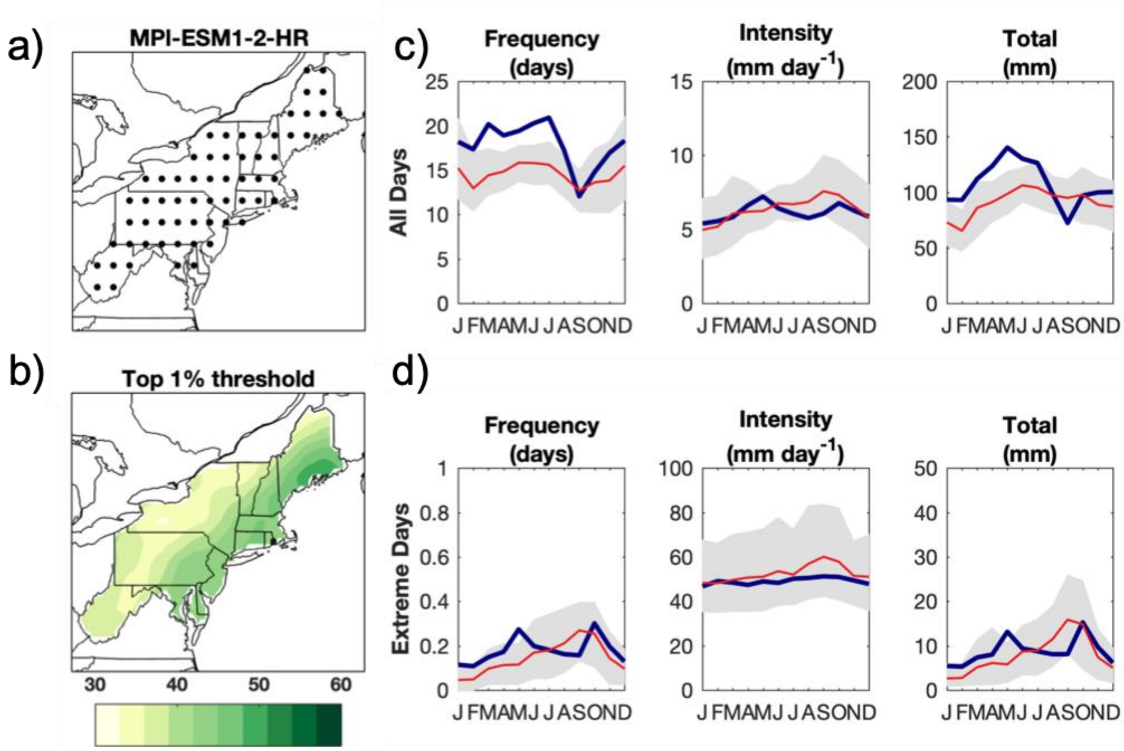
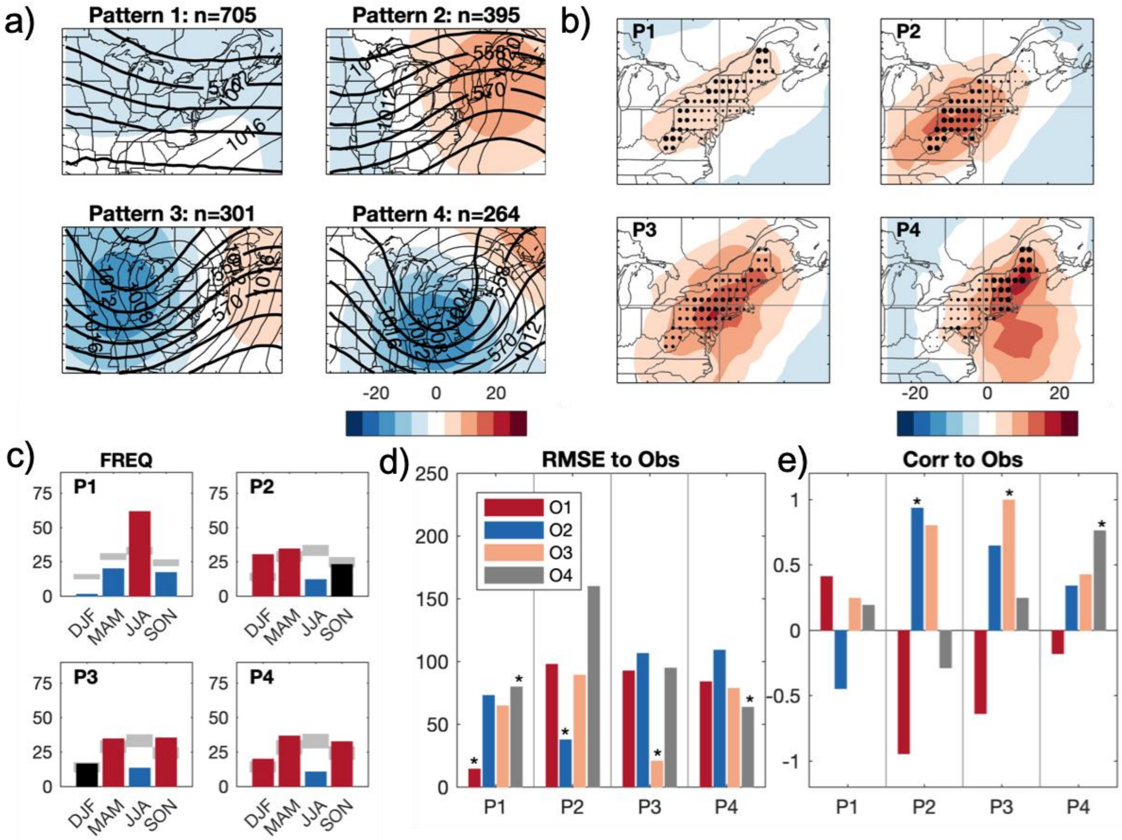


Figure 6. Same as Figure 4, but for MPI-ESM1-2-HR.

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.



637

638 Figure 7. Same as Figure 5, but for MPI-ESM1-2-HR.

639

Accepted for publication in *Journal of Climate*. DOI10.1175/JCLI-D-19-1025.1.