Weather and Climate Extremes 10 (2015) 10–28

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Interpreting climate model projections of extreme weather events

Stephen J. Vavrus *, Michael Notaro, David J. Lorenz

Nelson Institute Center for Climatic Research, University of Wisconsin-Madison, 1225 W. Dayton Street, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 31 December 2014
Received in revised form
4 September 2015
Accepted 24 October 2015
Available online 28 October 2015

Keywords:
Climate Model
Uncertainty
CMIP
Downscaled
Extremes

ABSTRACT

The availability of output from climate model ensembles, such as phases 3 and 5 of the Coupled Model Intercomparison Project (CMIP3 and CMIP5), has greatly expanded information about future projections, but there is no accepted blueprint for how this data should be utilized. The multi-model average is the most commonly cited single estimate of future conditions, but higher-order moments representing the variance and skewness of the distribution of projections provide important information about uncertainty. We have analyzed a set of statistically downscaled climate model projections from the CMIP3 archive to assess extreme weather events at a level aimed to be appropriate for decision makers. Our analysis uses the distribution of 13 global climate model projections to derive the inter-model standard deviation, skewness, and percentile ranges for simulated changes in extreme heat, cold, and precipitation by the mid-21st century, based on the A1B emissions scenario. These metrics provide information on overall confidence across the entire range of projections (via the inter-model standard deviation), relative confidence in upper-end versus lower-end changes (via skewness), and quantitative uncertainty bounds (derived from bootstrapping).

Over our analysis domain, which covers the northeastern United States and southeastern Canada, some primary findings include: (1) greater confidence in projections of less extreme cold than more extreme heat and intense precipitation, (2) greater confidence in relatively conservative projections of extreme heat, and (3) higher spatial variability in the confidence of projected increases in heavy precipitation. In addition, we describe how a simplified bootstrapping approach can assist decision makers by estimating the probability of changes in extreme weather events based on user-defined percentile thresholds.

1. Introduction

Utilizing climate model projections can be challenging for both climatologists and decision makers. Projections from a set of models often exhibit considerable scatter and may even differ on the sign of a future climate change, such as whether a location will become wetter or drier. For many scenarios, there may be consensus on the direction of change (e.g., a warming climate due to greenhouse forcing), but the models may differ greatly on the projected magnitude of the change. Improvements in the quality of climate models have not rectified these discrepancies, as evidenced by the similar spread among projections between the newer Coupled Model Intercomparison Project phase 5 (CMIP5) and the older CMIP3 collection (Knutti and Sedláček, 2013). Differences among the projected changes in extreme weather are of particular relevance for decision makers, because of the disproportionate socioeconomic impact these events exert.

The need to usefully interpret inconsistent model simulations has spurred numerous assessment efforts aimed at quantifying uncertainty in projections and increasing the reliability of projections. The simplest and most widely reported metric is the arithmetic average or multi-model mean among a set of model simulations, which provides the most relevant single piece of guidance on expected change. Information on uncertainty has often been expressed in terms of basic metrics, such as the range of projections among all models (Scherrer and Baettig, 2008) or the inter-model standard deviation of a projection (Maloney et al., 2013). Common alternative approaches are to identify where a vast majority of models agree on the sign of a change (Meehl et al., 2007) and somewhat more refined versions that depict a combination of high inter-model agreement within regions of significant changes (Kirtman et al., 2013; Knutti and Sedláček, 2013). At the other end of the complexity spectrum are highly advanced statistical approaches, such as Bayesian methods (Tebaldi et al., 2005), hierarchical statistical models (Cressie and Wikle, 2011), and Reliability Ensemble Averaging (Giorgi and Mearns, 2002). These more sophisticated strategies are useful in their own right,

* Corresponding author.
E-mail addresses: sjvavrus@wisc.edu (S.J. Vavrus), mnotaro@wisc.edu (M. Notaro), dlorenz@wisc.edu (D.J. Lorenz).

http://dx.doi.org/10.1016/j.wace.2015.10.005

but they are unlikely to be adopted by managers seeking practical guidance on how to digest model information for the purpose of decision making. Regardless of the method, uncertainty assessments are useful for indicating how much confidence should be placed in the inter-model average, but directly translating such information for a particular application is not straightforward.

An alternative approach is to winnow a set of projections by giving more credence to models considered to be the most accurate and downgrading the others. This intuitively satisfying strategy has been explored in a number of studies (e.g., Georgi and Mearns, 2002; Murphy et al., 2004; Annan and Hargreaves, 2010) and has been applied widely in an attempt to optimize various projections (Schmittner et al., 2005; Chapman and Walsh, 2007; Wang and Overland, 2012). Unfortunately, these efforts have been hindered by the lack of a theoretical justification for weighting model projections and by the practical difficulties in doing so (Knutti, 2010), in part because there is no clear relationship between a model's skill in simulating past climate and the magnitude of its projected changes (Knutti et al., 2010). Furthermore, deriving uncertainty information from a weighted model-mean poses additional statistical challenges, and basing a weighting scheme on smaller sample sizes constituting extreme events creates further difficulties.

In this study, we aim to strike a practical balance between simple and complex methods of quantifying uncertainty in climate projections for use by decision makers. Our focus is on extreme weather events, because of their severe societal impacts and overall positive trends (Gleason et al., 2008; Karl and Katz, 2012; Walsh et al., 2014). We concentrate on short-term (daily) extremes of heat, cold, and precipitation over the northeastern United States and southeastern Canada, based on statistically downscaled projections. Although the methodology used here can be applied generally, we focus on simulated climate change by the mid-21st century, a time period of increasing relevance for practical decision-making. Our three primary statistical measures to characterize uncertainty in a set of model projections are considered to be basic to intermediate in complexity: standard deviation (converted to the coefficient of variation), skewness, and percentile ranges derived from bootstrapping. The primary goal of this study is to provide an assessment of projected changes in extreme weather that is relevant for decision makers. We focus on *statistical* uncertainty among model projections, each of which is deemed equally plausible, rather than addressing the underlying causes of model sensitivity that might explain why the projections differ from each other.

2. Data and methods

We use a high-resolution (0.1°) data set of daily maximum temperature, minimum temperature, and precipitation that was statistically downscaled from 13 global climate model (GCM) simulations included in CMIP3 (Table 1). We focus on the late-20th century (1961–2000) and mid-21st century (2046–2065), based on the "middle-of-the-road" A1B emissions scenario, along with a supplemental analysis using the lower-emission B1 scenario. Our downscaled projections are an improved and spatially expanded version of a data set originally covering the state of Wisconsin, which has been widely used for a variety of climate change studies and assessments (WICCI, 2011; Notaro et al., 2011, 2012; Veloz et al., 2012; Vavrus and Behnke, 2013). Details of the downscaling procedure are given in Notaro et al. (2014), but an important feature is that a particular large-scale atmospheric pattern does not yield a unique temperature or precipitation value at the surface. Instead, the downscaling is probabilistic by virtue of parameter values translated into a probability density function that varies in time and space according to large-scale atmospheric fields. In addition, the late-20th century model output was debiased, following the cumulative distribution function algorithm of Wood et al. (2004). The same debiasing method was applied to the original Wisconsin-based downscaled data, which was found to produce an excellent match with observations of extreme daily weather events (WICCI, 2011), unlike some downscaling methods that rely on linear regression and analogs (Gutmann et al., 2014). The statistical downscaling was trained on observations from the National Weather Service's Cooperative Observer Program and Environment Canada's Canadian Daily Climate Data.

Table 1
List of CMIP3 GCMs used in this study, along with their original resolution before downscaling was applied. Horizontal resolution is listed in degrees of latitude/longitude, translated into approximate values for models that use spectral truncation. Vertical resolution is expressed by the number of levels (L) used by the model.

#	Model Name	Country	Horizontal resolution (deg)	Vertical resolution
1	CGCM3.1 (T47)	Canada	3.75	L31
2	CGCM3.1 (T63)	Canada	2.8	L31
3	CNRM_CM3	France	2.8	L45
4	CSIRO_mk3.0	Australia	2.8	L18
5	CSIRO_mk3.5	Australia	2.8	L18
6	GFDL_CM2.0	United States	2.5	L24
7	GISS_AOM	United States	3.5	L12
8	IAP_FGOALS	China	2.8	L26
9	MIROC3.2 (medres)	Japan	3.5	L20
10	MIROC3.2 (hires)	Japan	1.2	L56
11	ECHO_G	Germany/Korea	3.75	L19
12	MPI_ECHAM5	Germany	2.8	L31
13	MRI_CGCM2.3.2a	Japan	3.5	L30

In this study, we focus on a domain approximating that of the Northeast Climate Science Center (NECSC) and associated Landscape Conservation Cooperatives (LCCs) that are part of the United States Department of the Interior (Fig. 1). The NECSC is one member of a federal network of eight regional centers created to provide scientific information, tools, and techniques that managers can use to anticipate, monitor, and adapt to climate change. The use of this relatively small area in our study allows more in-depth analysis of spatial variations and makes absolute temperature and precipitation thresholds of extremes more meaningful. This domain encompasses several of the LCCs, the primary stakeholders and partners of the Climate Science Centers, whose mission is to connect scientific information with on-the-ground conservation

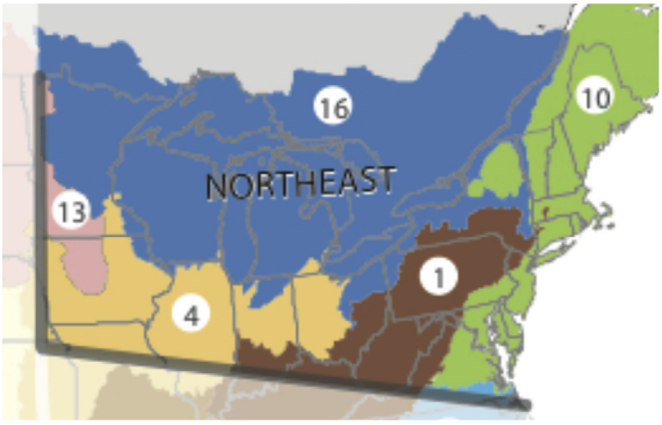

Fig. 1. Domain used in this study, encompassing the Northeast Climate Science Center area (thick gray lines) and several Landscape Conservation Cooperatives (numbered): (1) Appalachian, (4) Eastern Tallgrass Prairie and Big Rivers, (10) North Atlantic, (13) Plains and Prairie Potholes, and (16) Upper Midwest and Great Lakes.

management strategies. We utilize the older CMIP3 GCM projections, rather than the newer CMIP5 data, based on downscaled data availability and the similarity in climate sensitivities and temperature extremes between the two versions (Knutti and Sedláček, 2013; Yao et al., 2013). Likewise, we focus on climate conditions for the A1B scenario during the mid-21st century, rather than late century, because of the similarity in projections among greenhouse gas emissions scenarios at mid-century, as well as the more immediate relevance for decision makers in the earlier period. In addition, we demonstrate that the results are similar when considering the weaker B1 emissions scenario (Appendix D). Furthermore, this study is designed to be more of a demonstration of the methodology than an exhaustive presentation of all possible scenarios and time periods, but our approach is applicable to other regions and model ensembles.

Various criteria have been used to define extreme events, grouped broadly between those that use an absolute threshold (e.g., daily maximum temperature above 25 °C) and a relative threshold (e.g., daily minimum temperature below the 10th percentile at a location). A relative measure may be preferable when making comparisons across very broad regions with large climatic differences from one place to another, but an absolute threshold is often easier to understand and can be more relevant for certain impacts, such as the occurrence of hard freezes. Here we define extremes in absolute terms, based on thresholds adopted for the United States in previous studies. Hot days are defined as a daily maximum temperature of at least 90 F (32.2 °C) (Kucharik et al., 2010; Maloney et al., 2013; Patz et al., 2014), extreme cold is taken to be a daily minimum temperature of 0 F (−17.8 °C) or lower (Vavrus and Van Dorn, 2010; WICCI, 2011), and heavy precipitation is a day with at least 2 in. (50.8 mm) of rainfall or liquid-equivalent snowfall (Karl et al., 1995; Groisman et al., 2004; Vavrus and Behnke, 2013).

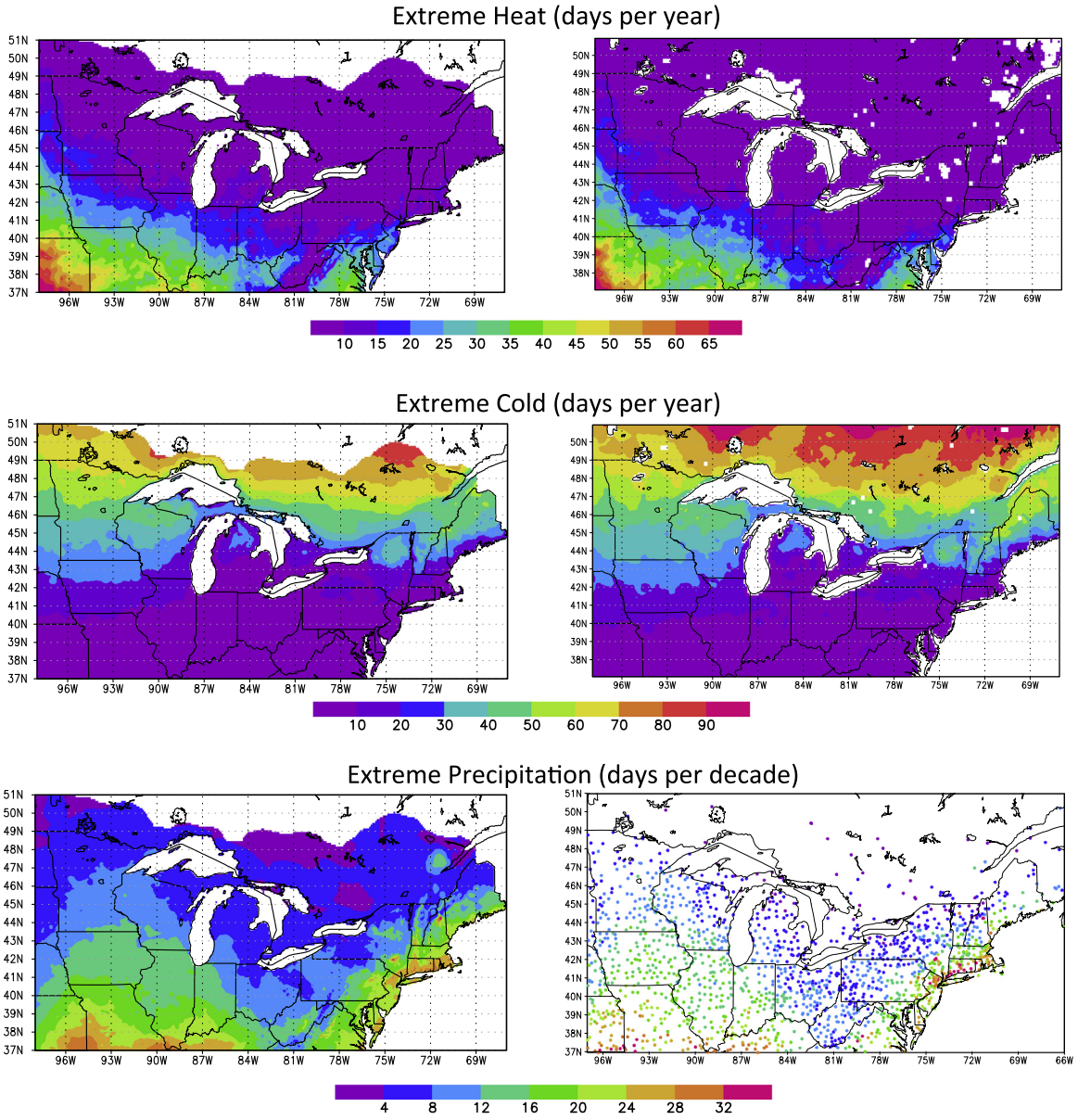

Fig. 2. Climatological frequency of extreme heat (days per year), cold (days per year), and precipitation (days per decade) during the 1961–2000 reference period, as simulated (left column) by the statistically downscaled CMIP3 models and observed (right column). Observations of extreme heat and cold are obtained from the Maurer et al. (2002) gridded product, which interpolates point observations to a 12-km grid. Station observations of precipitation are taken from the National Weather Service's Cooperative Observer Program and Environment Canada's Canadian Daily Climate Data.

3. Results

3.1. Climatology of extreme events in late 20th-century climate (1961–2000)

The frequency of extreme heat, cold, and precipitation varies considerably across the domain but generally exhibits coherent patterns dictated by latitude, continentality, distance to inland water bodies, and elevation (Fig. 2). Very hot days are rare (< 10 days per year) in northern and eastern regions, as well as along the Appalachian Mountains. They are much more common in the far southwestern sector, which experiences a continental climate that is more conducive to extreme heat than comparably southern locations along the Atlantic coast. By contrast, occurrences of extreme cold are more zonally oriented but also influenced by the Great Lakes, with maximum frequencies along the northernmost fringe and a broad minimum across the south. The spatial pattern of heavy precipitation is regulated by both latitude and proximity to moisture sources. One maximum is in the far south and west that extends northward, broadly along the Mississippi River Valley with access to Gulf of Mexico moisture, while the other peak follows the Atlantic seaboard. All of these simulated patterns in the downscaled models strongly resemble observationally derived data (Fig. 2), indicating that the downscaled multi-model means are highly accurate and/or that the debiasing procedure effectively reproduces the tails of the temperature and precipitation distributions. By contrast, the representation of extremes in the raw GCMs is much less realistic, severely undersimulating the occurrence of heavy rainfall and failing to account for the moderating influence of elevation in the frequency of extreme heat along the Appalachian Mountains (Figs. A1–A3).

3.2. Changes in future extremes (2046–2065)

By the mid-21st century, the models project generally large changes in the frequency of extreme heat, cold, and precipitation (Fig. 3, Figs. B1–B3, Table 2). Not surprisingly, occurrences of extreme heat are expected to increase, while bitter cold becomes less common, and the sign of these changes is uniform across the entire domain. The changes in all extremes are statistically significant at the 90% confidence level virtually everywhere, based on a Monte Carlo simulation described in Section 3.3. The largest simulated increase in hot days occurs in the far southeast coastal region, rather than where extreme heat has been most frequent in the recent past (the far southwest). Otherwise, the changes resemble the pattern of historical climatology, in that the increases are smallest where hot days are least common: to the north and in regions of high elevation. Likewise, projected reductions in extreme cold are generally largest in northern regions, where bitter cold is more common, although the pattern of change differs somewhat from historical climatology by virtue of larger declines in extreme cold over the far northeast (southeastern Canada) than over the far northwest. A consistent response across the domain is also seen in extreme precipitation, which is expected to increase in agreement with other studies of recent and future large-scale change (Trenberth, 1999; Groisman et al., 2004; Wehner, 2005; Zhang et al., 2007). However, the pattern of these changes is less coherent than for extreme temperature, except for a clear maximum along the Eastern seaboard, particularly along the New England coast. That regional peak is consistent with a local maximum in the historical climatology, but the other favorable region for heavy rainfalls in the far south does not exhibit an enhanced increase in the future. The enhanced spatial variability provided by downscaling is evident by comparing Fig. 3 with the corresponding projected changes from the raw GCM output (Figs. C1–C3), which generally exhibit much smaller increases in very heavy precipitation, weaker topographic influence, and less pronounced moderating effects from the Great Lakes.

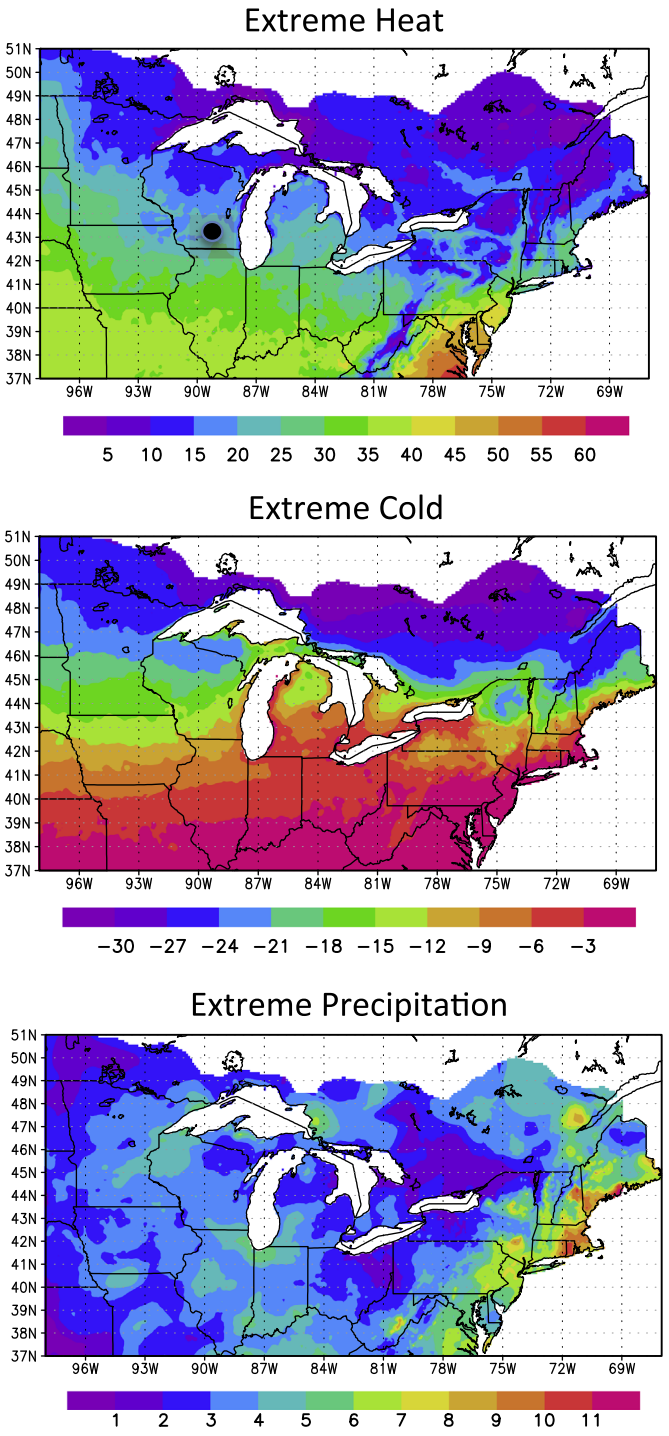

Fig. 3. Multi-model mean change in the simulated frequency of (top) extreme heat, (middle) extreme cold, and (bottom) extreme precipitation between 2046–2065 versus 1960–2000. The changes in temperature extremes are expressed as days per year, while changes in precipitation extremes are in days per decade (see text for definitions of these extremes). Madison, Wisconsin, is denoted with a black dot in the top panel.

These multi-model mean (MMM) patterns of change arguably provide the most useful single piece of information about future conditions, particularly because the MMM typically outperforms any single model in an ensemble (Doblas-Reyes et al., 2003; Cantelaube and Terres, 2005; Gleckler et al., 2008). However, the MMM can mask features that are important for interpretation and

Table 2
Areally averaged statistics of simulated extremes across the study domain, expressed as the multi-model mean, coefficient of variation (COV), and skewness of change. The frequencies and changes for extreme heat and cold are listed in days per year, while extreme precipitation is in days per decade.

Type of extreme	Mean frequency (1961–2000)	Mean change (2046–2065)	COV of change	Skewness of change
Heat	13.58	21.89	0.52	1.05
Cold	26.12	−13.41	0.20	−0.47
Precipitation	11.03	3.48	0.63	0.10

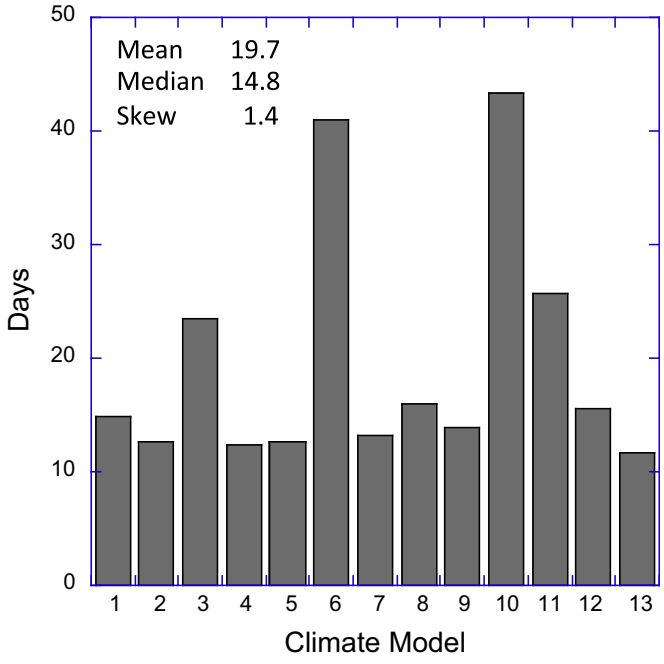

Fig. 4. Projected change in the number of extremely hot days per year in Madison, Wisconsin, by mid-century (2046–2065) compared with the late 20th century (1961–2000) among the 13 downscaled climate model projections.

possibly for applications. As an example, we show the distribution of projected increases in hot days from all 13 models for Madison, Wisconsin, the state capital located in south-central Wisconsin (Fig. 4). The graph clearly shows a large spread about the MMM (19.7 days) that can be detected visually and quantified by the intermodel standard deviation (10.8 days), but the latter metric also fails to capture the most telling feature. In this example, the spread is highly asymmetric (skewness = 1.4), featuring two outlier models that simulate about six more weeks of hot weather per year in the future, more than double the MMM. In contrast, nine of the 13 models are generally consistent with each other in simulating much smaller increases of around two weeks annually. In this case, the inter-model average is not a good representation of the majority, which is captured more closely by the inter-model *median* change of 14.8 days, and it also overlooks information on two extreme projections that are important to consider.

We extend this concept by analyzing the local intermodel spread and skewness over the entire domain. To enable a comparison of the model spread in projections among the three types of extremes, we use the coefficient of variation (COV) to normalize the intermodel standard deviation by the intermodel mean change (MMM) at each grid point. Essentially an inverted signal-to-noise ratio, the COV thus quantifies the degree of model agreement among regions and variables in a comparable manner, thereby providing a measure of confidence in the mean projections. We find considerable differences in the patterns of COV from one

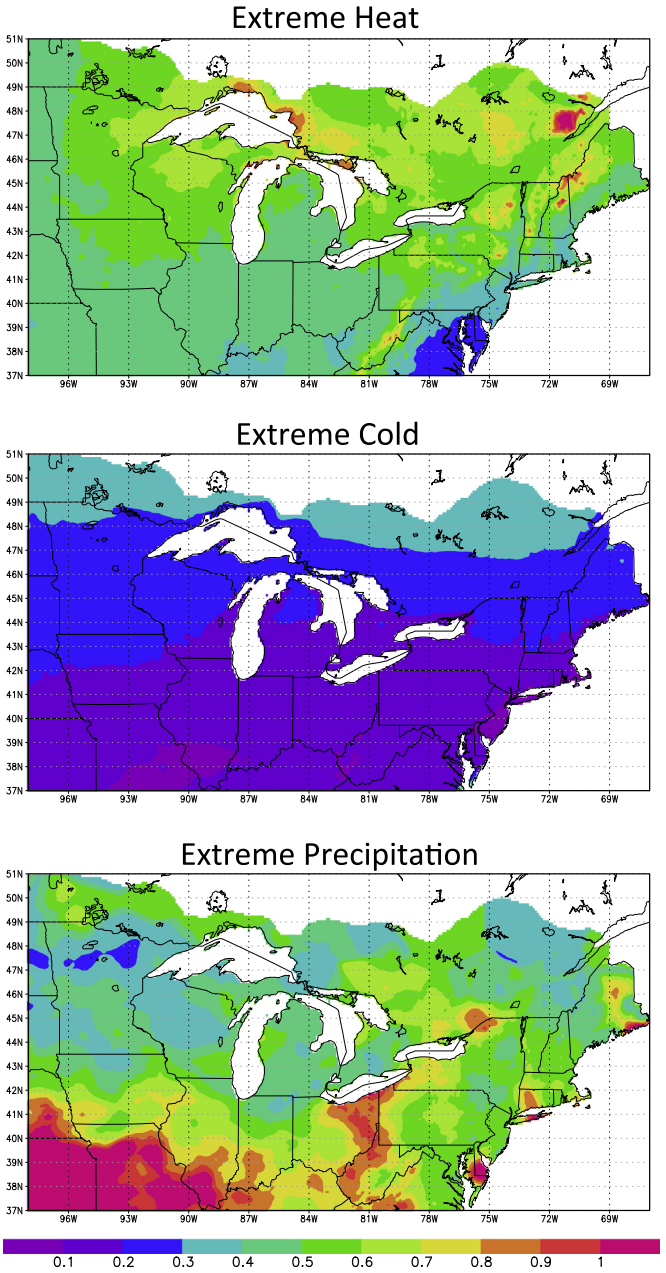

Fig. 5. Like Fig. 3 but for the coefficient of variation (COV) among the 13 model projections, expressed as the inter-model standard deviation divided by the multi-model mean.

location to another and among the three variables (Fig. 5, Table 2). Projected changes in extreme cold exhibit the lowest values, not only in places where bitter cold is infrequent but also in the far north (COV < 0.4), suggesting that a relatively high amount of confidence can be placed in the projected decreases in severe cold weather shown in Fig. 3. COV values are generally higher for the projected increases in extreme heat, especially in northern and elevated regions, where the mean changes are expected to be small. The intermodel spread in extreme precipitation projections is the largest among the variables, shows much less spatial coherency, and does not resemble the pattern of the MMM. The weakest intermodel agreement for extreme precipitation occurs in the far southwest of the domain.

The degree of asymmetry in the spread of projections is reflected in the skewness parameter, whose characteristics vary greatly among the three types of extremes (Fig. 6, Table 2).

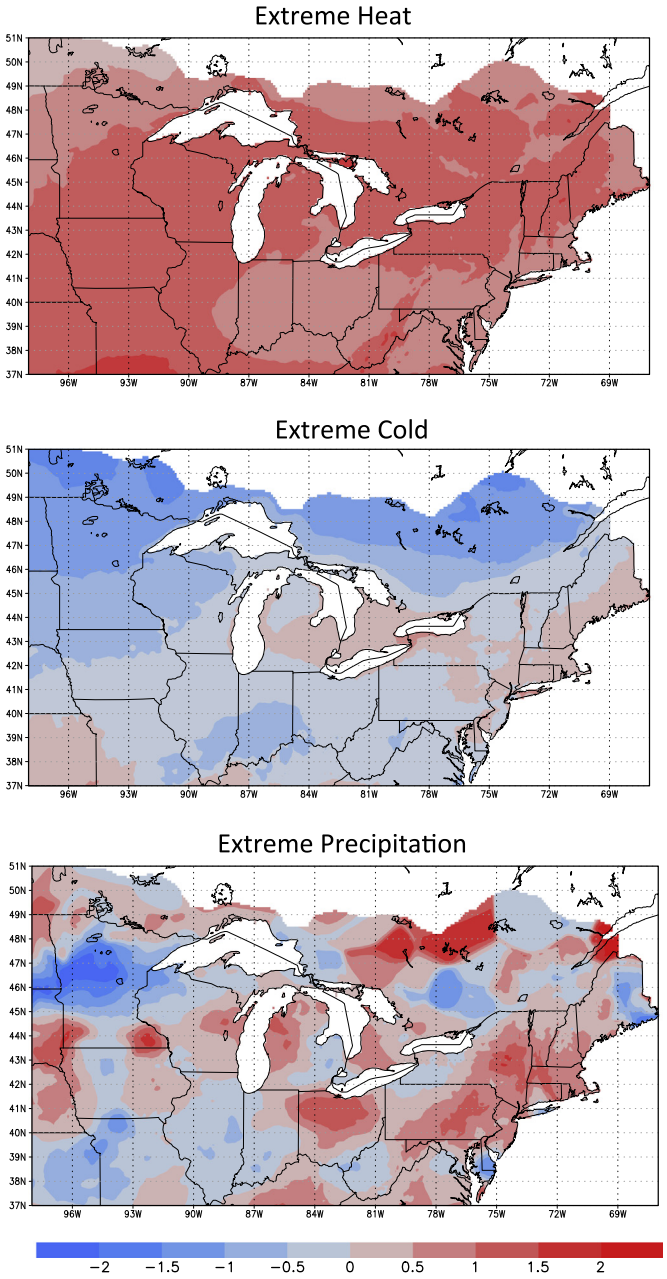

Fig. 6. Like Fig. 5 but for the skewness of the 13 model projections.

Consistent with the results shown above for Wisconsin, positive skewness in the distribution of projected hot days occurs over the entire domain and is relatively large. The skewness exceeds 0.5 almost everywhere, and its domain average is slightly above 1, which is more than double the corresponding magnitude for extreme cold. This finding indicates the presence of a small minority of models that simulate exceptionally large increases in hot days, compared with the more conservative projections by most models. In fact, there are two such models, the GFDL_CM2_0 and the MIROC3_2_HIRES, which are responsible. This pair simulates a much larger increase in hot days averaged over the entire domain compared with the multi-model mean (68% and 104% more, respectively). Interestingly, these are also the two GCMs with the highest horizontal resolution among the 13 models we analyzed.

Simulated reductions in extreme cold exhibit more symmetry across the models, as indicated by the widespread regions with small skewness of either sign (within ± 0.5). There are generally larger, negative skewness values in northern areas, where the greatest declines in bitter cold are projected. Analogous to the projections of extreme heat, that pattern reveals a tendency for a few models to simulate much larger reductions in extreme cold regionally than the intermodel average. The degree of symmetry in the projections of heavy precipitation varies widely and shows no coherent pattern, with skewnesses ranging from high positive to high negative across the domain. This noisiness probably reflects the spotty nature of heavy rainfall, consistent with the relatively high amount of spatial variability in the COV and MMM change in extreme precipitation.

3.3. Probabilistic projections of extremes

The analysis above provides a more in-depth assessment of future change than can be gained from multi-model mean projections alone, but it may still be too general for many applications. For example, how should a decision maker account for the existence of a few outlier models that simulate exceptionally large increases in hot weather? One way to obtain more quantitative estimates is to translate a set of model simulations into probabilistic projections. We do so here through a standard statistical bootstrapping procedure and demonstrate that, to a very good approximation, the transformed distribution is Gaussian and thus any set of desired percentile ranges can be easily obtained from knowledge of the MMM and the standard deviation of the bootstrapped distribution.

To demonstrate, we return to the modeled projections of extreme heat for Madison, Wisconsin shown in Fig. 4. To obtain a large number of possible MMM changes, we resample the 13-member distribution of projected changes in hot days by applying bootstrapping (Efron, 1979). This exercise generates 1000 possible realizations of the MMM by creating new sets of 13 projections selected randomly from the original data. The resulting distribution (Fig. 7) displays much more symmetry than the original (Fig. 4), as measured by the large decrease in skewness from 1.4 to

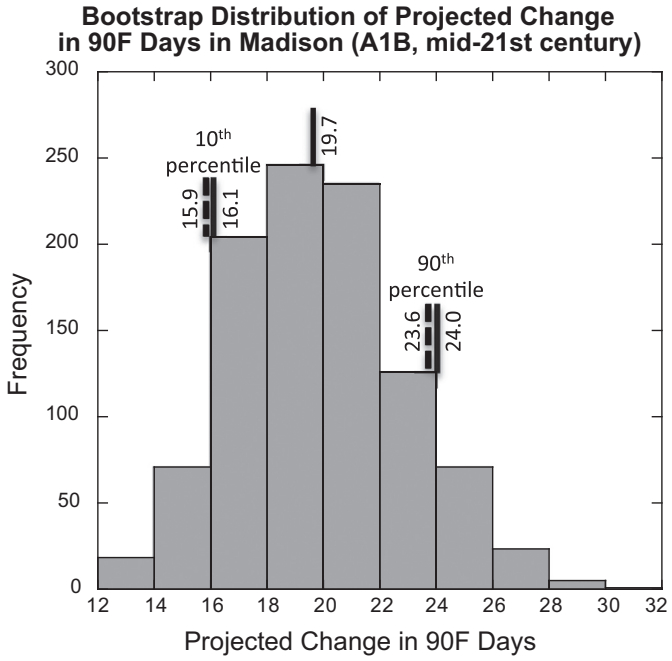

Fig. 7. The bootstrapped distribution of projected changes in extremely hot days by the mid-21st century at Madison, Wisconsin, based on the 13 individual model projections shown in Fig. 4. The histogram is derived from a 1000-member Monte Carlo simulation. The multi-model mean of 19.7 days is denoted, as are the 10th and 90th percentiles based on the actual bootstrapped distribution (solid) and an approximated Gaussian distribution (dashed).

0.45. More importantly, this large set of resampled projections can be used to quantify uncertainty in the MMM (a 19.7 day increase) via percentile bounds. Supposing, for example, that a decision maker is interested in knowing the 10th and 90th percentile range of the MMM projection in hot days, this histogram reveals that these limits are 16.1 days and 24.0 days, respectively. Because of the fairly symmetric shape of this distribution, these limits show similar deviations from the original multi-model mean projection (-3.6 days and 4.3 days, respectively). In fact, similar percentile bounds occur if one assumes a perfectly normal distribution and calculates these limits based on the standard deviation of the histogram: 15.9 days for the 10th percentile and 23.6 days for the 90th percentile, an error of only 0.2 days and 0.4 days, respectively. This approximation has practical value, because it allows a user of the model output to estimate any percentile value of interest simply by obtaining the MMM and the standard deviation of the bootstrap simulation. We find that the bootstrapped distributions are also fairly close to Gaussian elsewhere, as evidenced by skewness values within ± 0.5 over 95% of the domain for all three types of extremes. As a consequence, we find that the approximation of a normal distribution causes only very minor errors ($\leq 1\%$) in the estimated frequency of extreme heat, cold, and precipitation in a test of five widely dispersed cities in our study domain (Madison, WI; Springfield, MO; Richmond, VA; Indianapolis, IN; and Burlington, VT).

By assuming a normal distribution everywhere, uncertainty ranges can easily be constructed to put the multi-model mean projections into context and create probabilistic estimates of future change. In Fig. 8, we show the percentile bounds of changes in extreme heat, cold, and precipitation expressed in terms of the difference between the 90th percentile and the inter-model mean (the deviation for the 10th percentile is of equal magnitude and opposite sign). By adding and subtracting these deviations from the MMM, one obtains the 10th and 90th percentile bounds of the projected MMM changes in extremes.

These patterns bear some resemblance to those for COV (Fig. 5) but also display noteworthy differences that reflect the raw intermodel standard deviation without the normalizing MMM term used to calculate COV. In particular, there is more coherency in the relatively high values for extreme precipitation across the southern and eastern portions of the domain, where heavy precipitation is most common (Fig. 2). These results provide not only quantitative, probabilistic estimates of changes in extremes, but also offer interesting interpretations of the projections. For example, the models simulate a maximum increase in hot days over the far southeast of the domain (Fig. 3), yet the uncertainty is smaller there than over the far southwest. Consequently, although Richmond, VA, is expected to have a considerably greater increase in hot weather occurrences than Jefferson City, MO—53.4 days vs. 34.4 days—the 10th/90th percentile uncertainty bounds are smaller for Richmond (4.5 days) than Jefferson City (5.7 days). This result suggests greater confidence in the inter-model mean projection for Richmond and provides a practical way of utilizing the region's low COV expressed in Fig. 5.

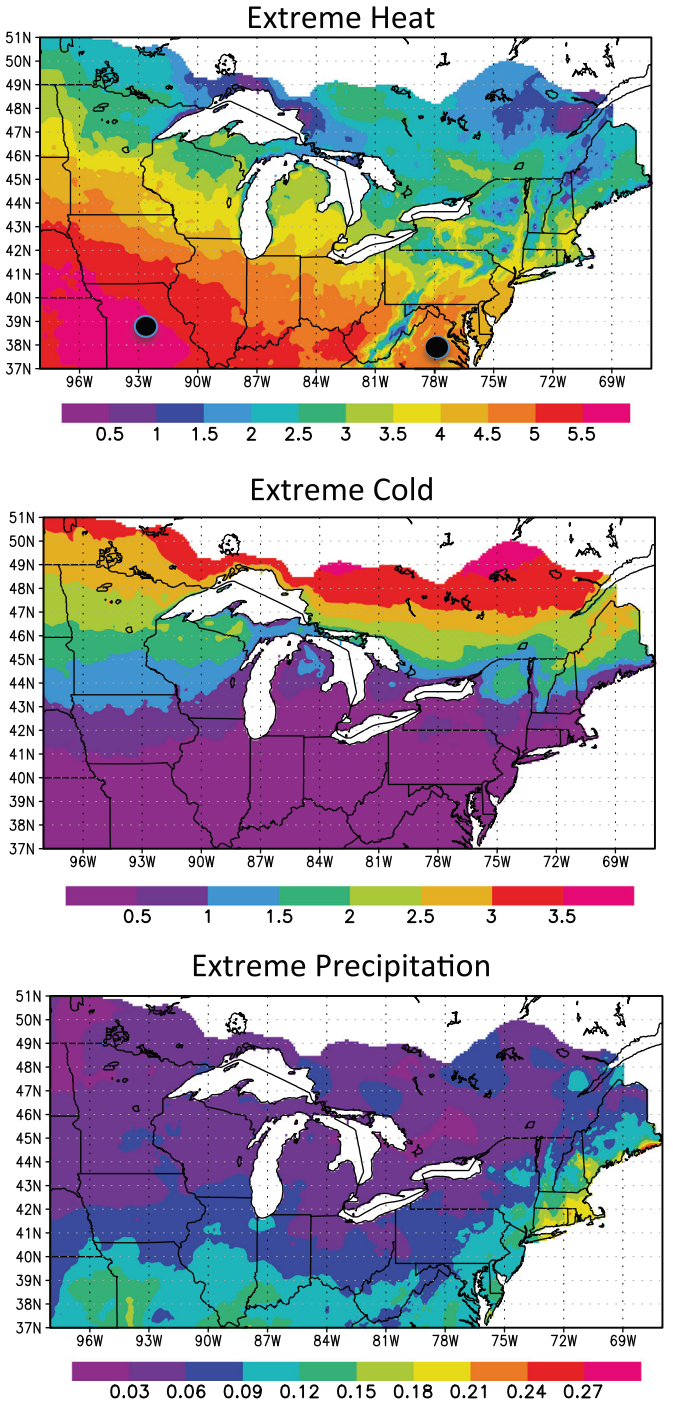

Fig. 8. Like Fig. 3 but for the difference between the 90th percentile and the multi-model mean of the 13 model projections. The difference between the 10th percentile and the MMM is equivalent in magnitude and of opposite sign to the patterns shown here. The locations of Jefferson City, MO, and Richmond, VA, are denoted with black dots in the top panel.

4. Discussion and conclusions

We have conducted an uncertainty assessment using a collection of climate model projections for a “middle-of-the-road” climate change scenario for the mid-21st century. As expected, we find similar, but somewhat muted, responses from the lower-emissions SRES B1 scenario (Figs. D1–D3). The results provide guidance for interpreting projected future changes and demonstrate a methodology that can be applied more generally to other time periods and scenarios. Our aim has been to strike a balance between comprehensive and simple analysis tools that is appropriate for decision makers who desire guidance but are unsure how to utilize the abundance of available model projections. Based on our analysis of first-, second-, and third moments of the distribution of statistically downscaled CMIP3 model projections, some highlights of this study include the following:

- By the middle of this century, *all* locations within our study domain are expected to experience additional extreme heat, less extreme cold, and more extreme precipitation (relative to

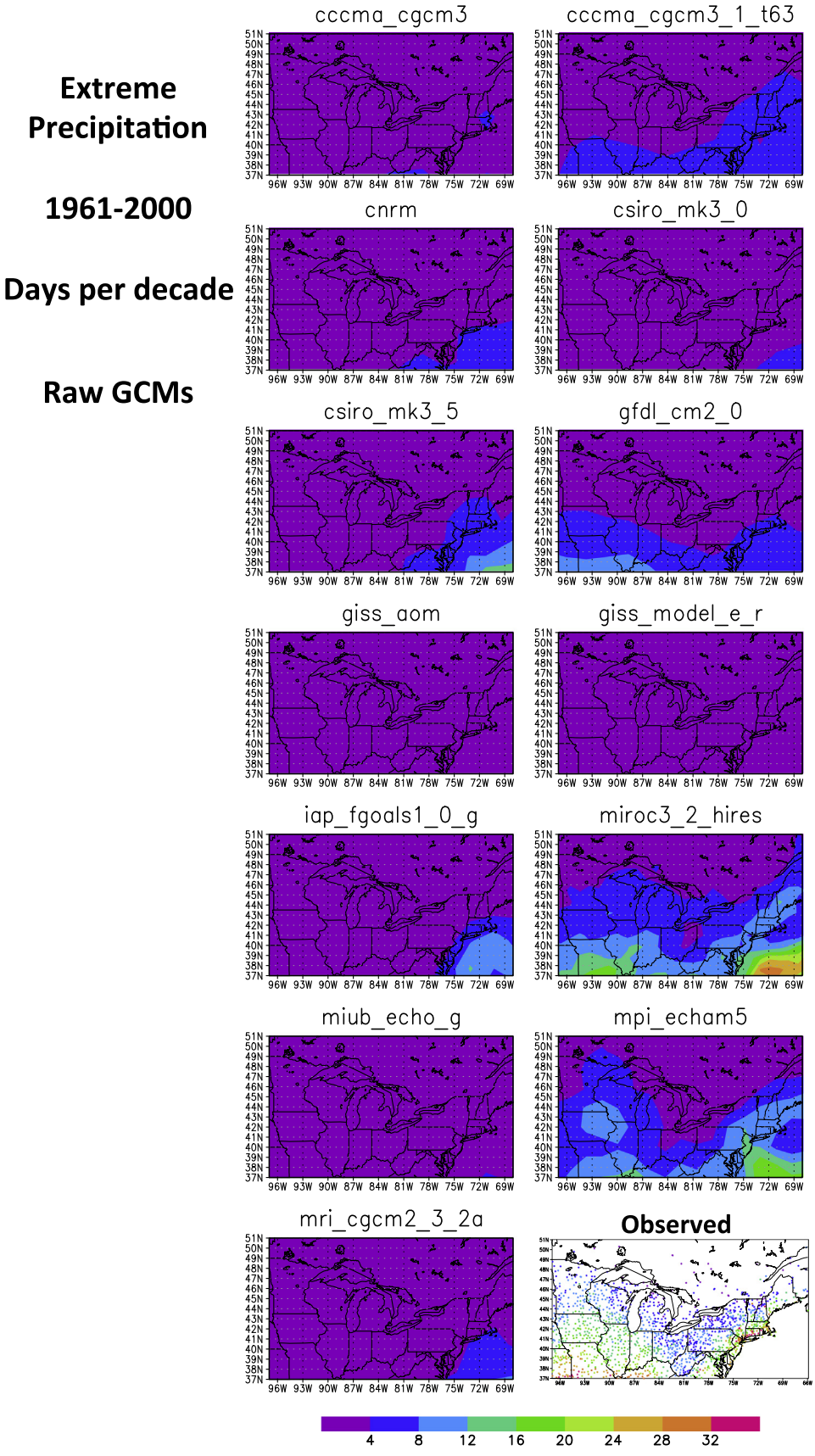

Fig. A1. Annual frequency of extreme precipitation during the late 20th century in the individual GCMs used in downscaling and (bottom right) observed from the National Weather Service's Cooperative Observer Program and Environment Canada's Canadian Daily Climate Data. Extreme precipitation is defined here as daily amounts of at least 2 in. or 50.8 mm.

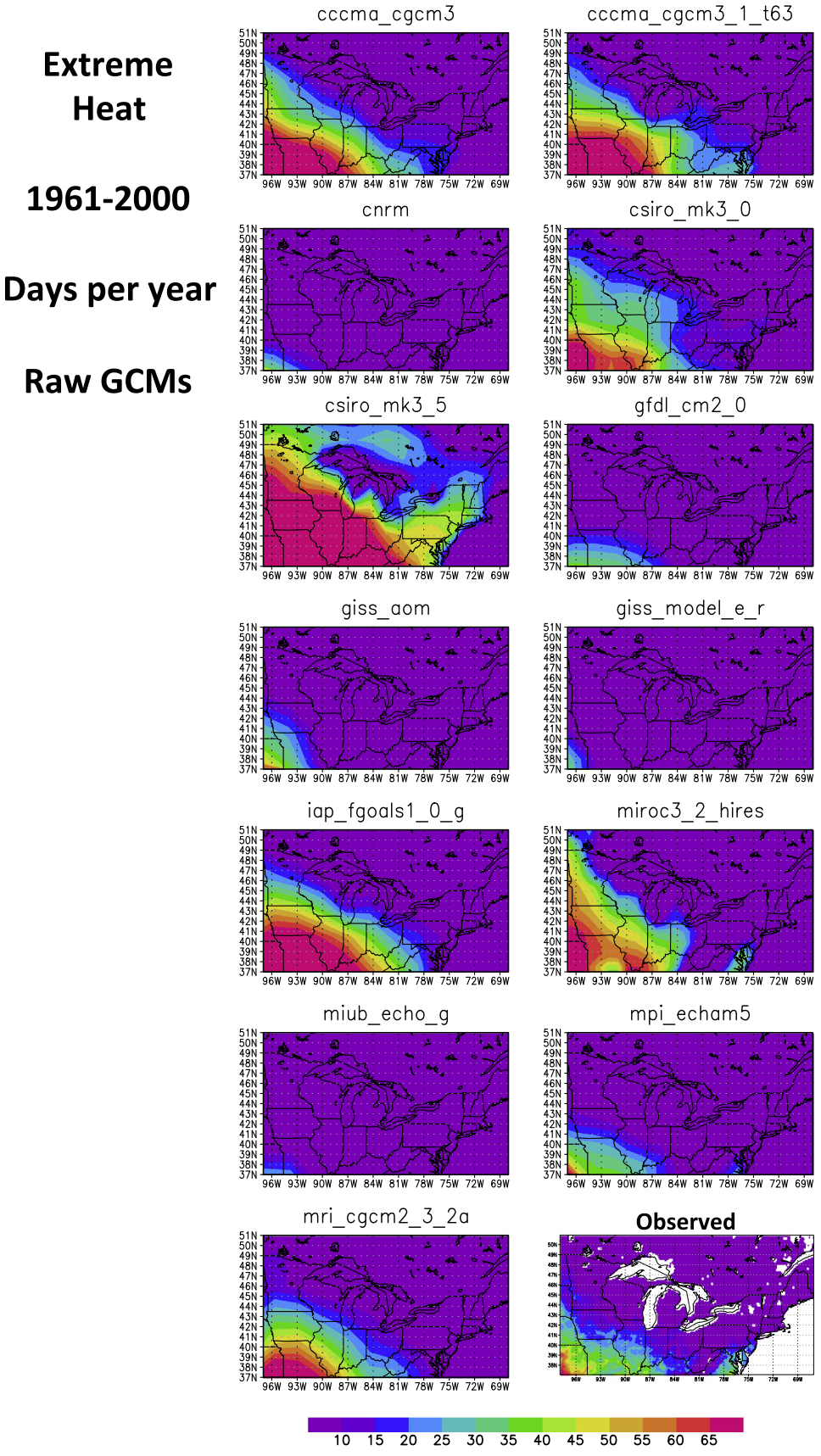

Fig. A2. Annual frequency of extreme heat during the late 20th century in the individual GCMs used in downscaling and (bottom right) observed from the Maurer et al. (2002) gridded product. Extreme heat is defined here as a daily maximum temperature of at least 90 F or 32.2 °C.

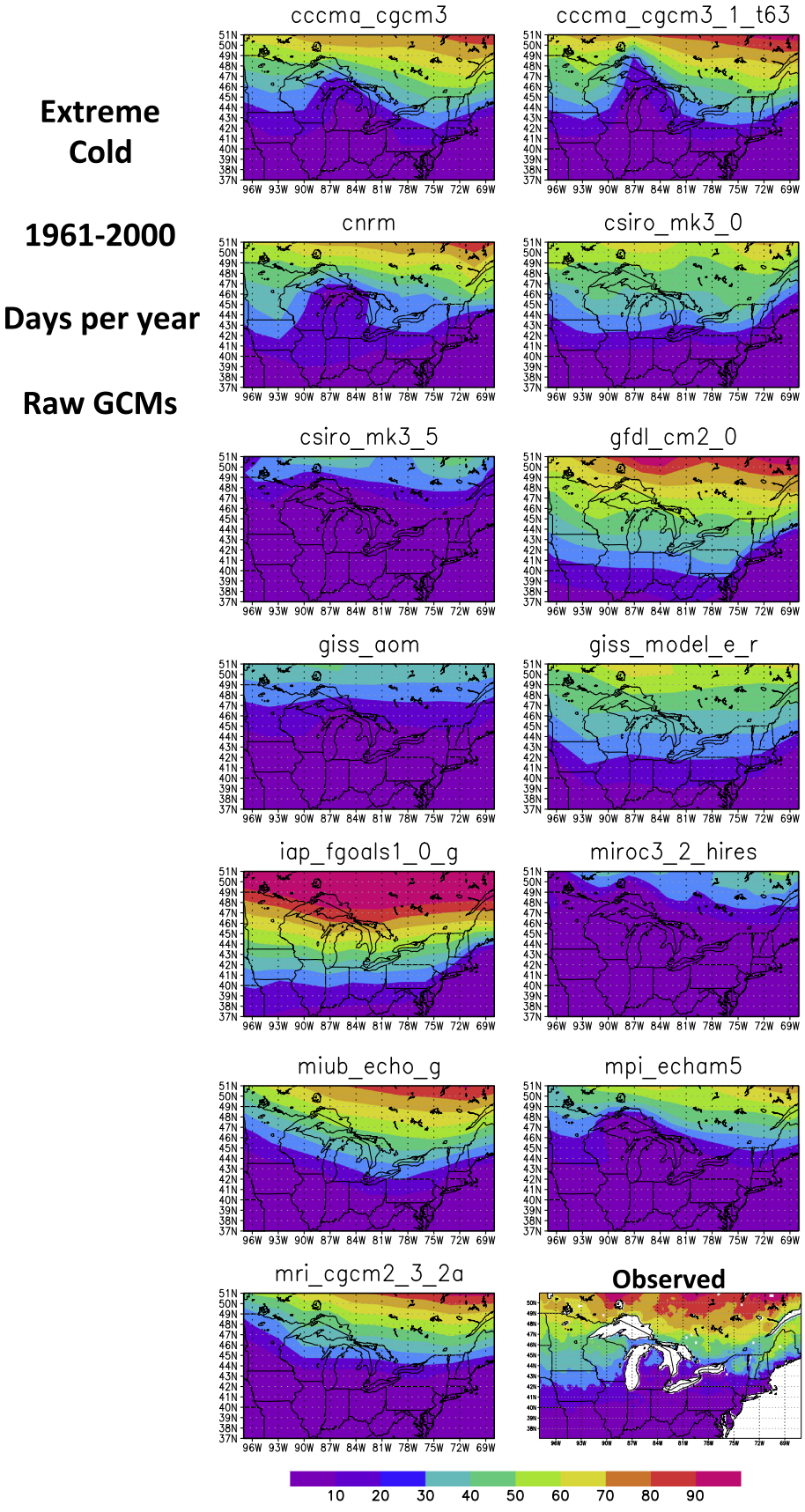

Fig. A3. Annual frequency of extreme cold during the late 20th century in the individual GCMs used in downscaling and (bottom right) observed from the Maurer et al. (2002) gridded product. Extreme cold is defined here as a daily minimum temperature of 0 F (− 17.8 °C) or lower.

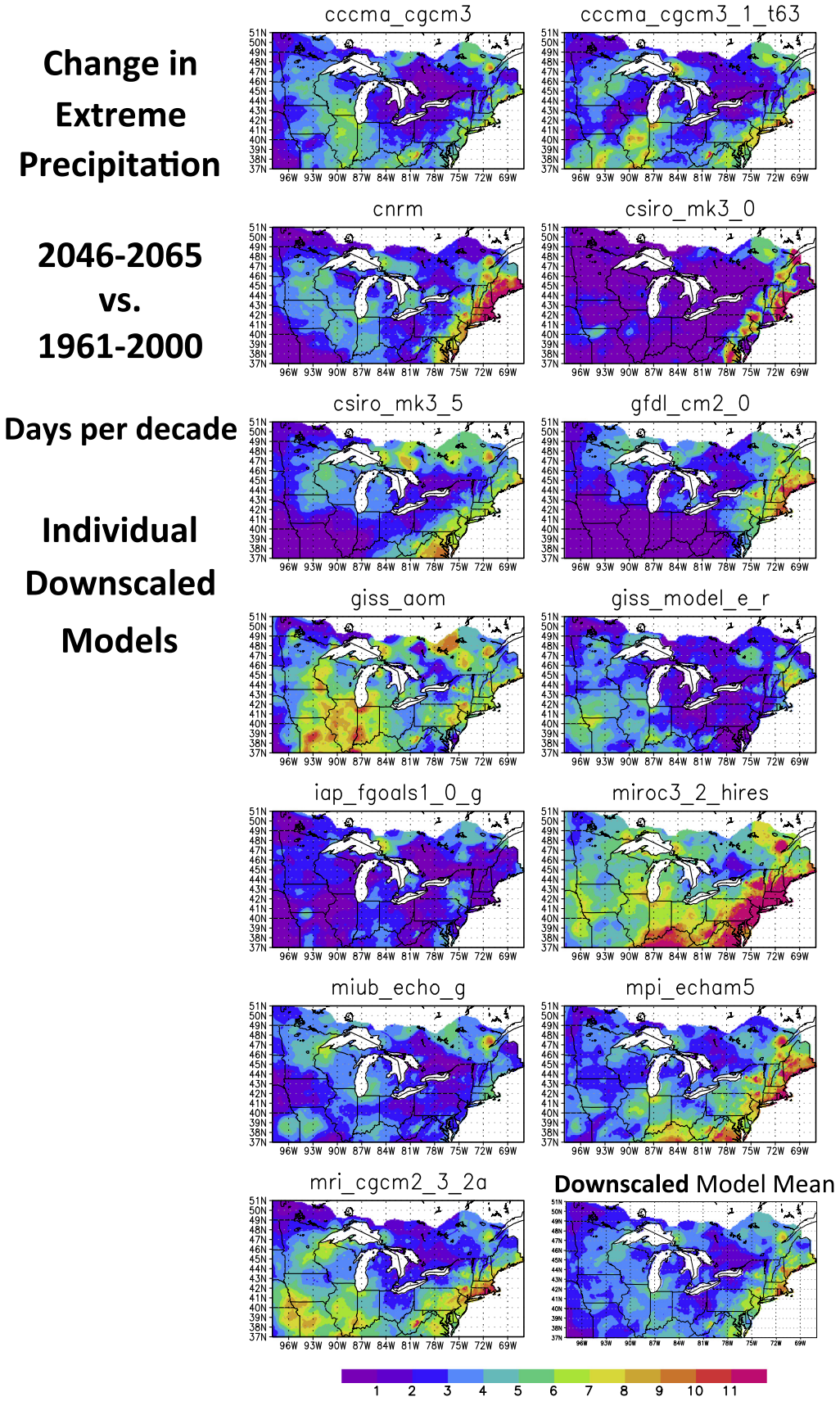

Fig. B1. Change in the simulated frequency of extreme precipitation (days per decade) between 2046 and 2065 versus 1961–2000 in each downscaled GCM and (bottom right) the multi-model mean. Extreme precipitation is defined here as daily amounts of at least 2 in. or 50.8 mm.

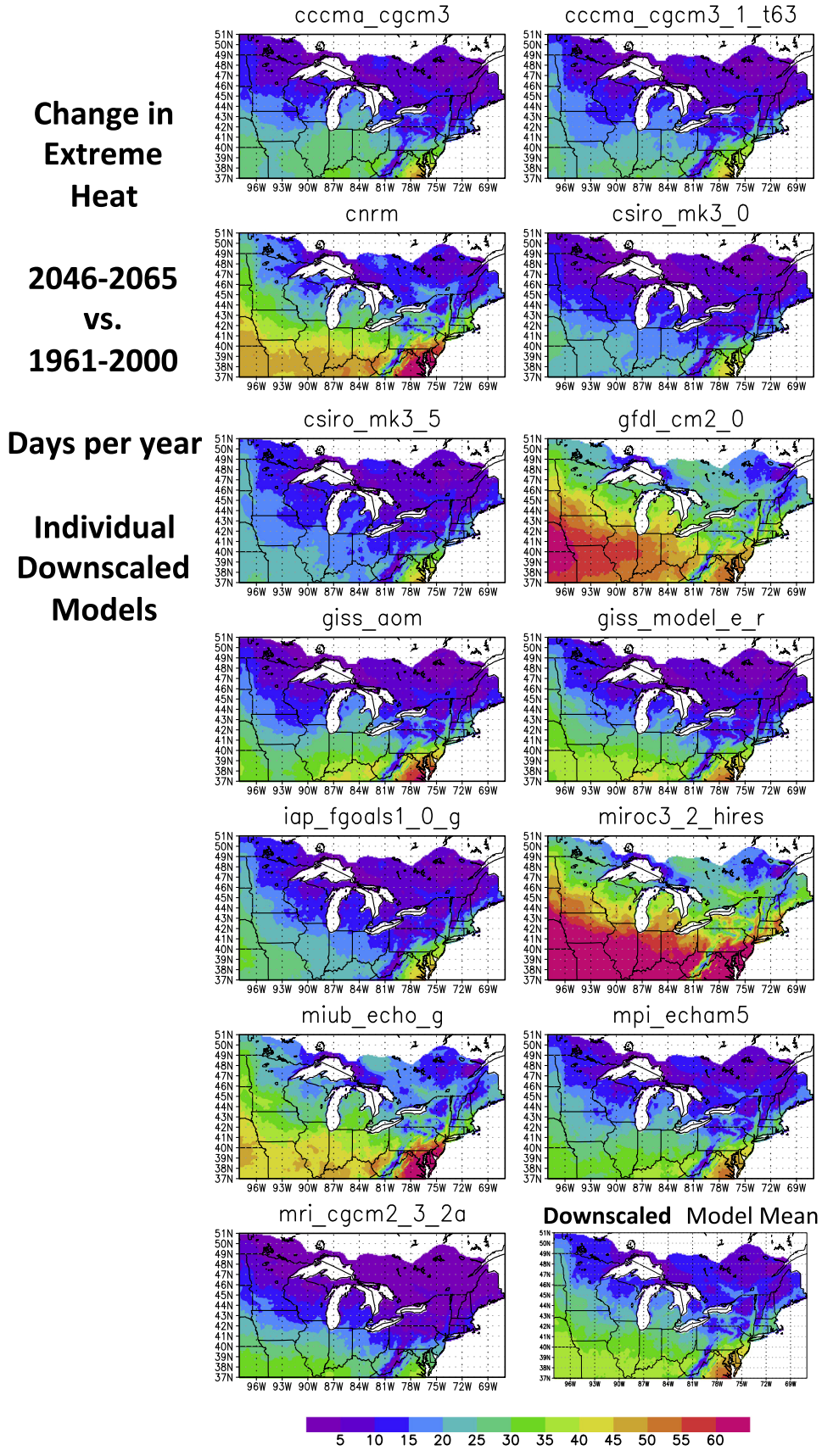

Fig. B2. As in Fig. B1 but for extreme heat (days per year). Extreme heat is defined here as a daily maximum temperature of at least 90 F or 32.2 °C.

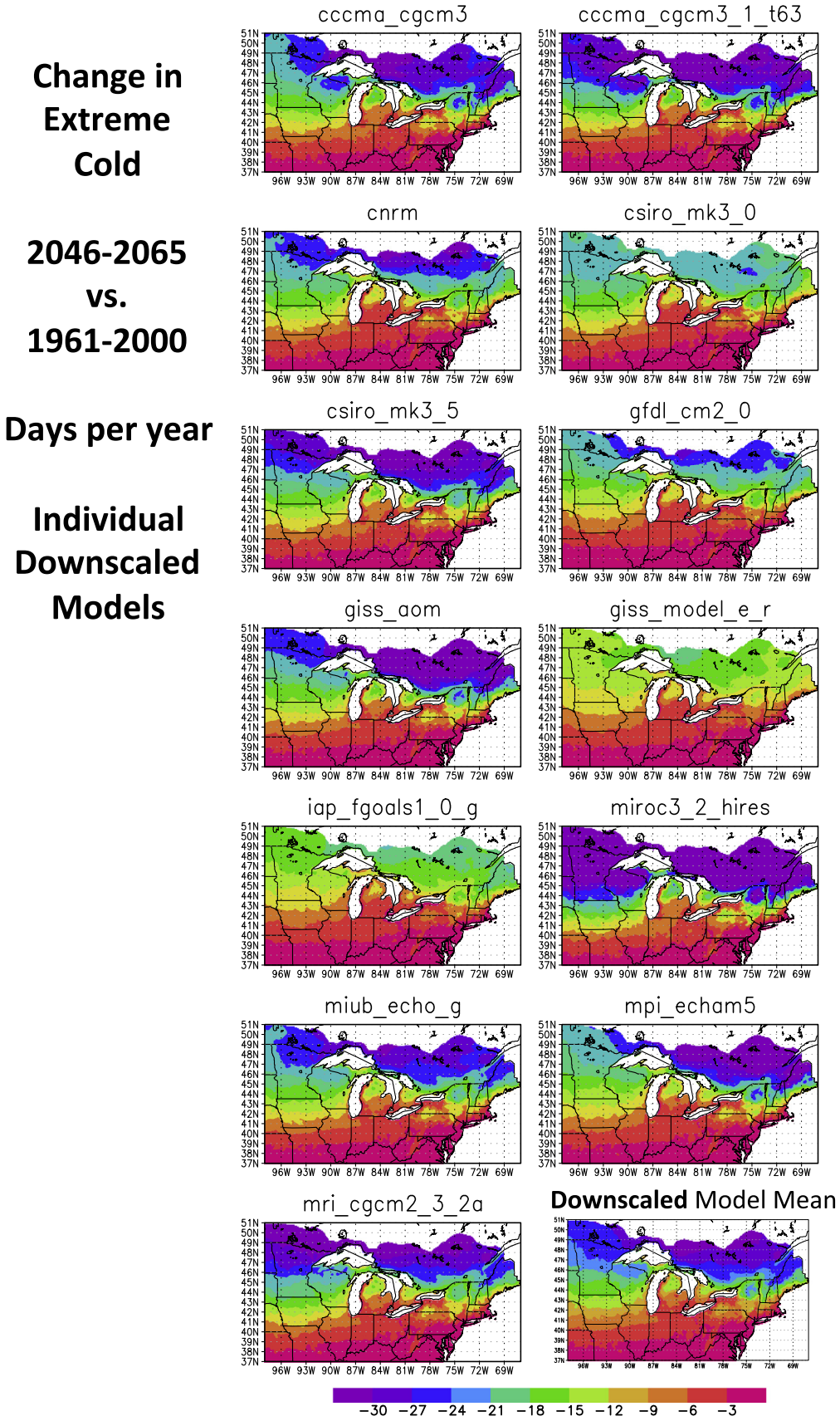

Fig. B3. As in Fig. B1 but for extreme cold (days per year). Extreme cold is defined here as a daily minimum temperature of 0 F (−17.8 °C) or lower.

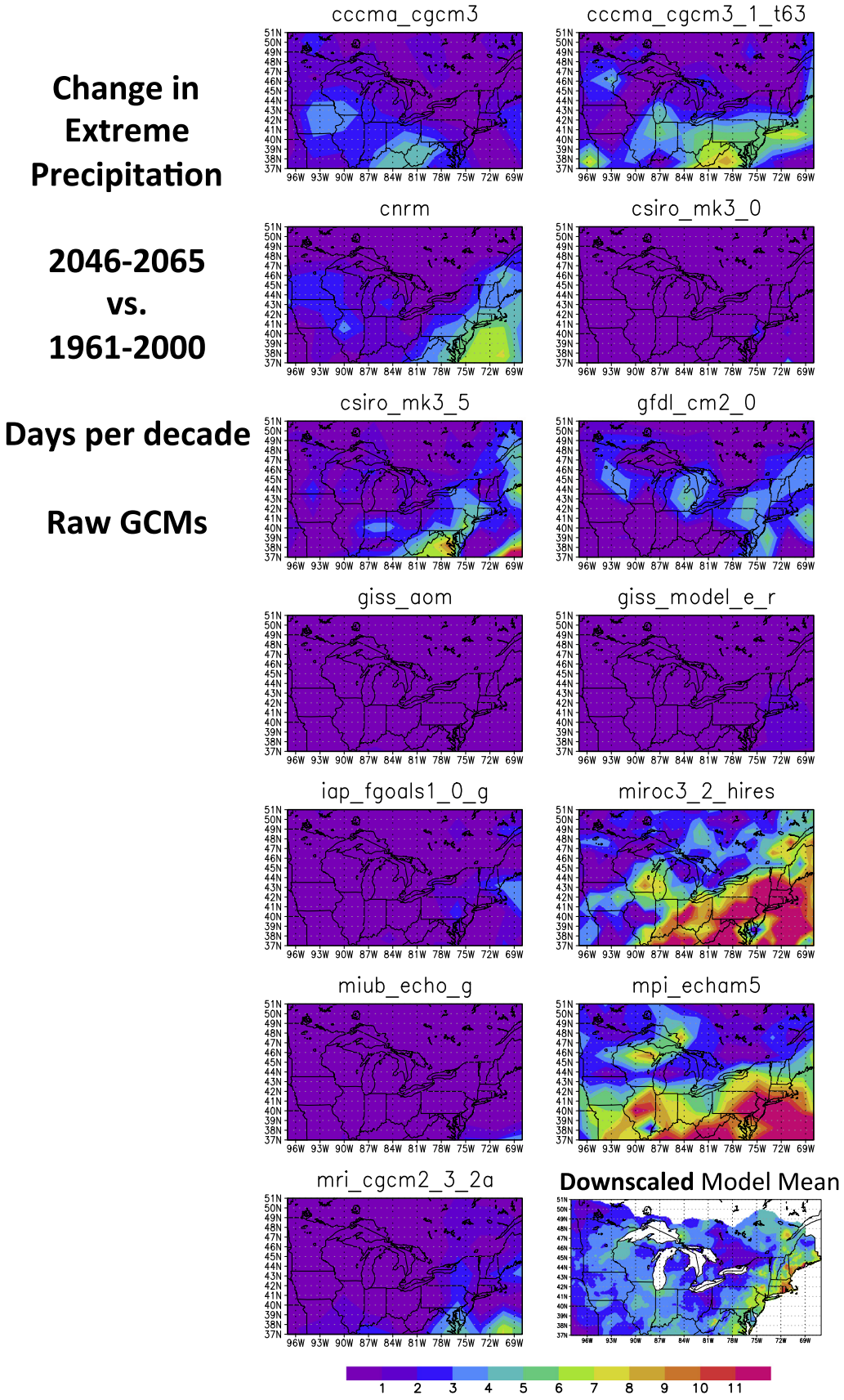

Fig. C1. Change in the simulated frequency of extreme precipitation (days per decade) between 2046–2065 versus 1961–2000 in each raw GCM and (bottom right) the downscaled multi-model mean. Extreme precipitation is defined here as daily amounts of at least 2 in. or 50.8 mm.

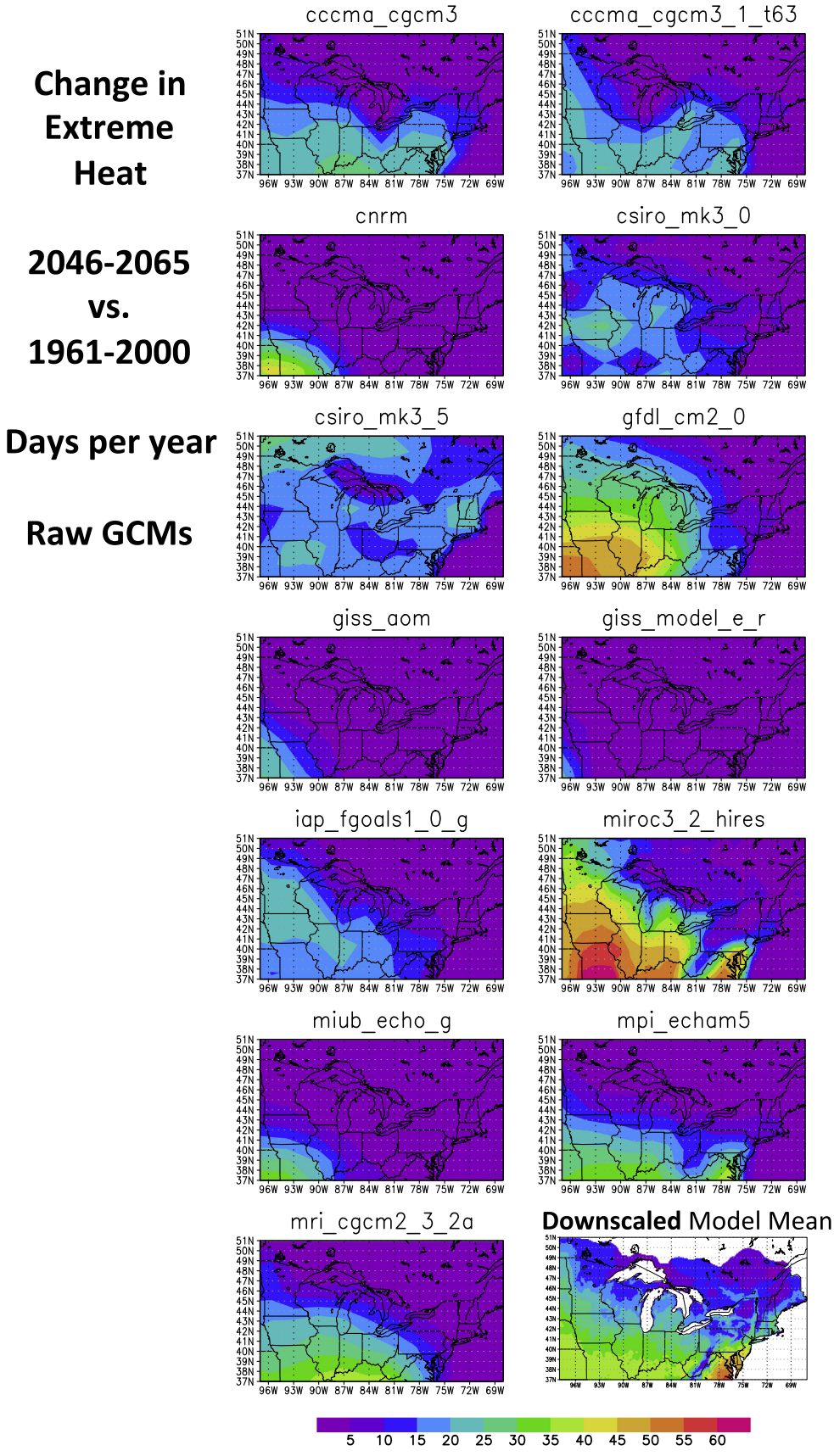

Fig. C2. As in Fig. C1 but for extreme heat (days per year). Extreme heat is defined here as a daily maximum temperature of at least 90 F or 32.2 °C.

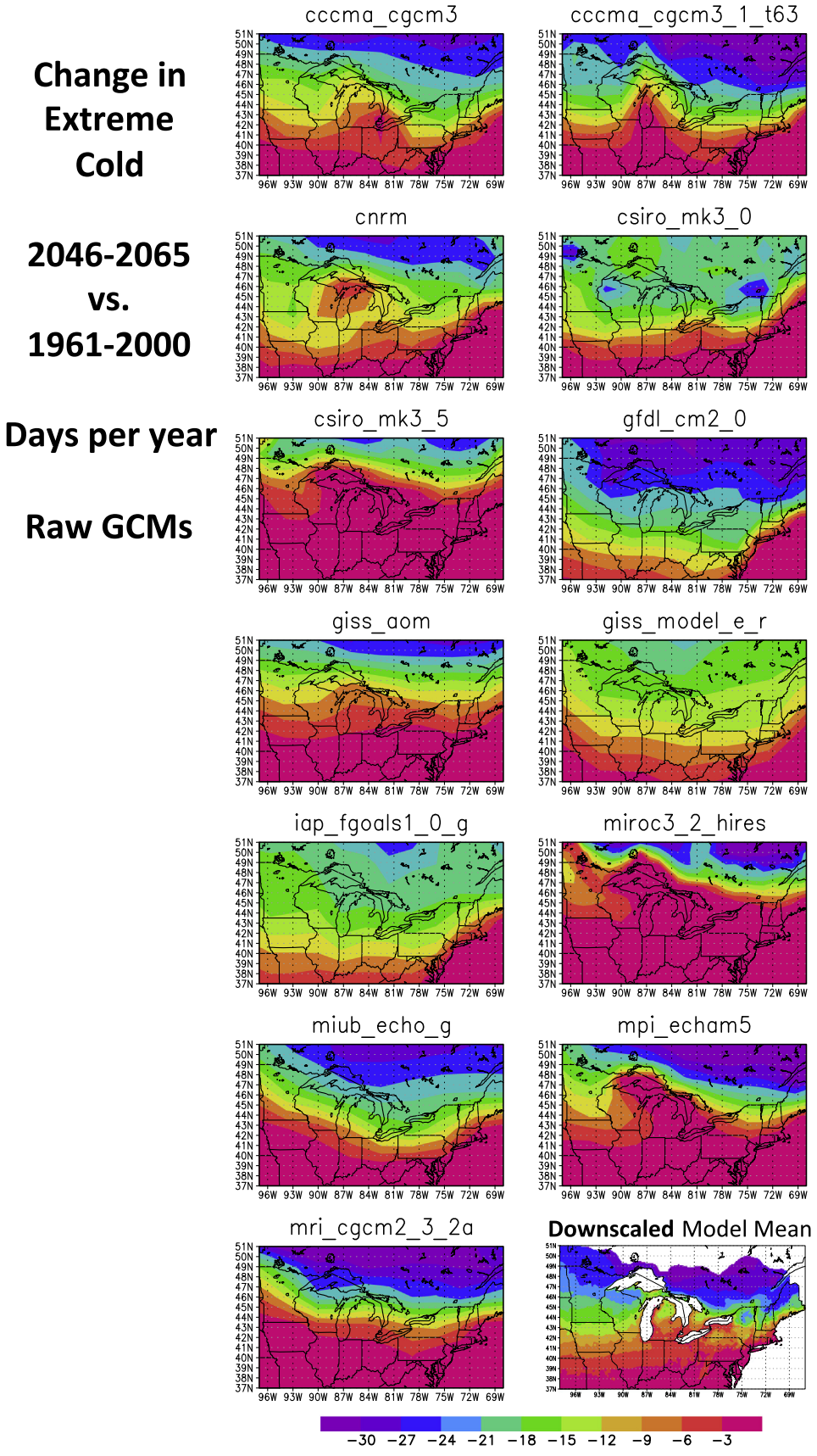

Fig. C3. As in Fig. C1 but for extreme cold (days per year). Extreme cold is defined here as a daily minimum temperature of 0 F (−17.8 °C) or lower.

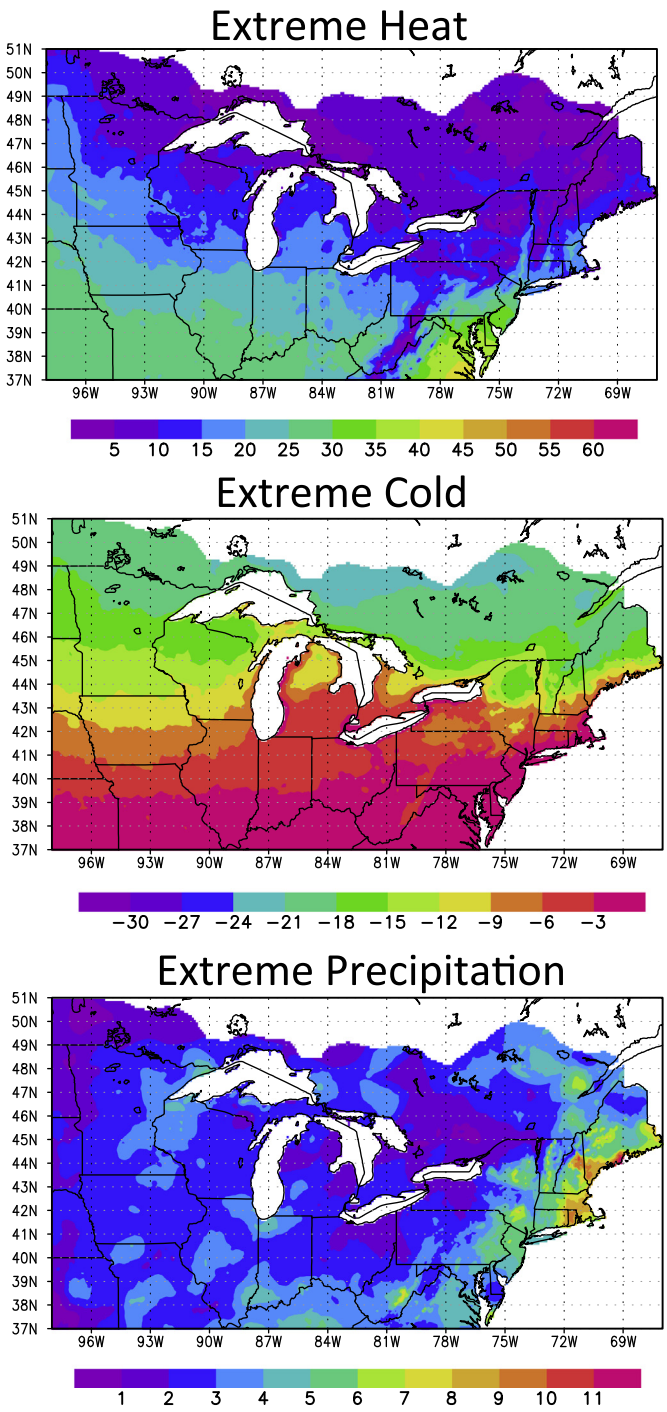

Fig. D1. Like Fig. 3 but for the SRES B1 emissions scenario.

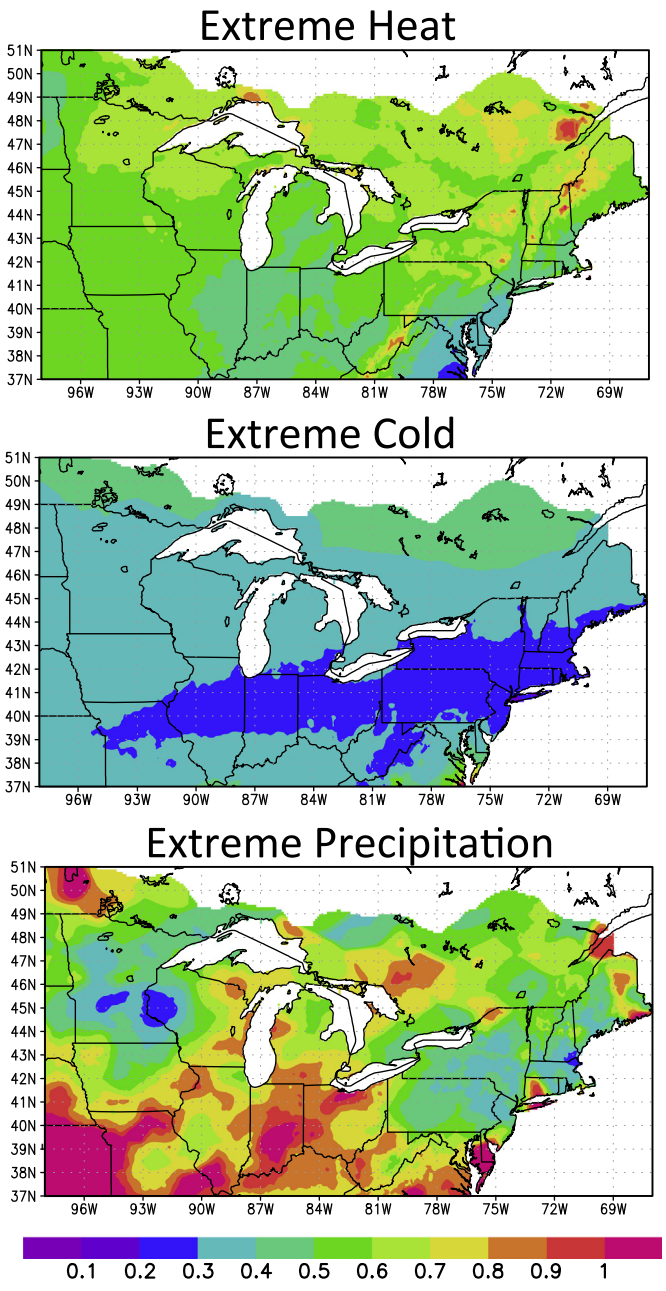

Fig. D2. Like Fig. 5 but for the SRES B1 emissions scenario.

the late 20th century). These projected changes are statistically significant at virtually all locations, based on the nearly unanimous agreement in the sign of projected changes among the 13 climate models.

- Assuming that consistency across models provides an overall measure of reliability, the results indicate that a higher level of confidence should be placed in projections of reduced extreme cold than enhanced extreme heat and precipitation. This finding holds both in northern regions, where bitter cold is more common, and across the south, where very cold conditions rarely occur. The precise reasons for the stronger inter-model agreement in projections of extreme cold are beyond the scope of this study, but the development of extremely cold air masses may depend less on complex feedback processes involving surface coupling that are known to influence extreme heat and precipitation (Schär et al., 1999; Seneviratne et al., 2006).
- Greater confidence appears to be warranted in the more conservative projections of increases in extreme heat. The existence of a couple of outlier models that simulate exceptionally more frequent hot days is widespread throughout the domain. By contrast, our results suggest that closer-to-equal confidence can be placed in the high-end and low-end projections across the 13 models of decreased extreme cold and increased extreme precipitation, although this finding is less useful for heavy precipitation due to the high spatial variability exhibited by that variable.
- Bootstrapping offers a way to quantify extreme weather projections probabilistically and may be a relatively easy and useful tool for decision makers seeking specific uncertainty bounds. A Gaussian approximation of the bootstrapped distribution

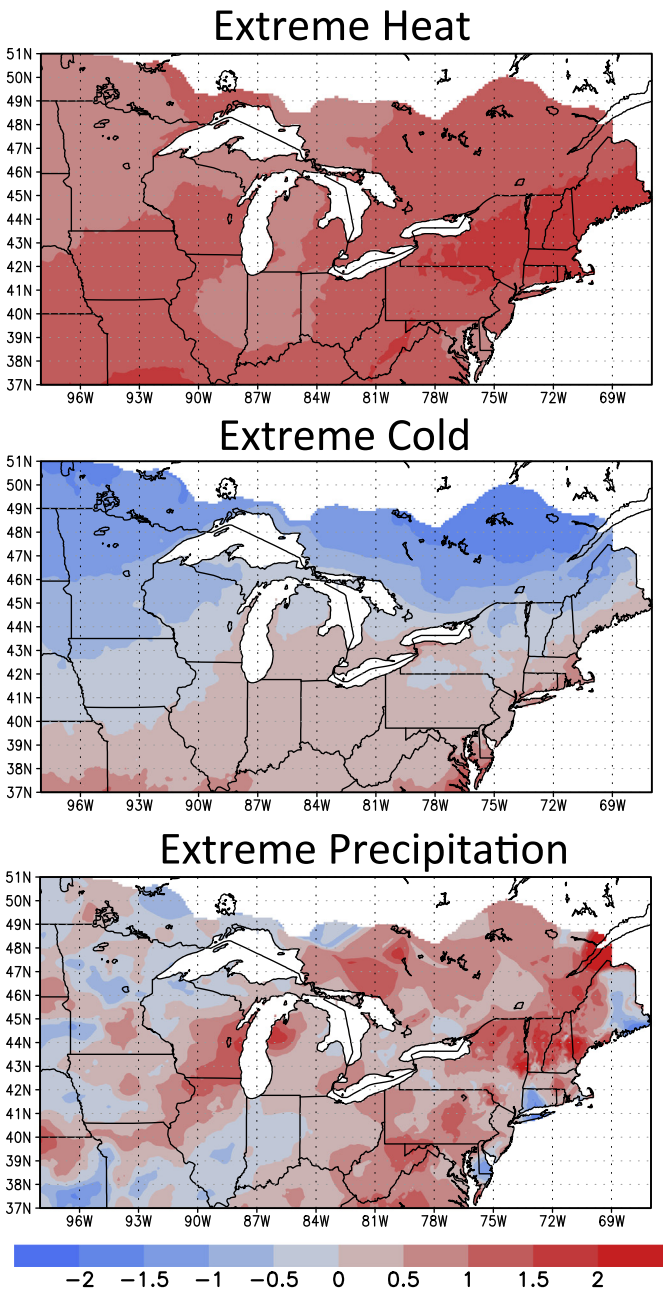

Fig. D3. Like Fig. 6 but for the SRES B1 emissions scenario.

facilitates calculations of the MMM percentile ranges and is probably sufficient for most purposes.

Some important caveats are in order, so that readers can appreciate the context of these findings. First, this study is exclusively a statistical uncertainty assessment, rather than incorporating a physically based component to diagnose the cause of noteworthy features, such as outlier projections. Physically based assessments should also be conducted to identify potentially biased processes or feedbacks that might predispose a model to simulate unrealistic changes in extreme events. Second, we have implicitly invoked "model democracy" in our confidence assessments by assuming that each projection has an equal likelihood. Given the difficulties of justifying model weighting (Stainforth et al., 2007; Knutti, 2010; Weigel et al., 2010), this simpler approach is defensible. Third, however, our methodology is not at odds with model weighting and could be applied in combination with it. To do so would simply require that the formulas for the inter-model mean, standard deviation (COV), and skewness be modified to accommodate the chosen weighting coefficients. Fourth, our study is focused narrowly on temperature and precipitation extremes at *daily* timescales, even though prolonged occurrences of extreme weather can trigger greater societal impacts (Pielke and Downton, 2000; Anderson and Bell, 2011). Fifth, the collection of 13 climate models used in this analysis is far from being exhaustive or mutually independent and thus can be considered an "ensemble of opportunity" (Knutti and Sedláček, 2013). Therefore, in addition to our constrained focus on a single time period and emissions scenario, we have likely underestimated the actual magnitude of uncertainty, including the quantitative probability bounds of the MMM derived from bootstrapping. The implications of using a limited number of model projections that are not independent samples constitute a topic that has been addressed by many others (e.g., Stainforth et al., 2007; Tebaldi and Knutti, 2007; Masson and Knutti, 2011).

Future work could follow up on this research by conducting a similar analysis using the newer CMIP5 collection of models and expanding the spatial domain to cover the entire United States or beyond. Additional climatic variables and alternative definitions of extremes, such as relative thresholds and multi-day events, could also be included to provide broader insights into both the likeliest future outcomes and the robustness of the methodology. As stated above, this work would also be enhanced by a complementary, physically based analysis to provide a more complete uncertainty assessment of greatest use for decision makers.

Acknowledgments

This work was primarily supported by award F12AP01074 from the Upper Midwest Great Lakes Landscape Conservation Cooperative through the U.S. Fish and Wildlife Service. We also acknowledge the support provided to conduct the statistical downscaling: award MSN118446 from the Energy Center of Wisconsin and awards 751B0200072 and 751P1301081 from the Michigan Department of Natural Resources.

Appendix A

See Figs. A1–A3.

Appendix B

See Figs. B1–B3.

Appendix C

See Figs. C1–C3.

Appendix D

See Figs. D1–D3.

References

Anderson, G.B., Bell, M.L., 2011. Heat waves in the United States: mortality risk during heat waves and effect modification by heat wave characteristics in 43 U. S. communities. Environ. Health Perspect. 119, 210–218.

Annan, J.D., Hargreaves, J.C., 2010. Reliability of the CMIP3 ensemble. Geophys. Res. Lett. 37, L02703. http://dx.doi.org/10.1029/2009GL041994.
Cantelaube, P., Terres, J.M., 2005. Seasonal weather forecasts for crop yield modeling in Europe. Tellus A 57, 476–487.
Chapman, W.L., Walsh, J.E., 2007. Simulations of Arctic temperature and pressure by global coupled models. J. Clim. 20, 609–632.
Cressie, N., Wikle, C.K., 2011. Statistics for Spatio-Temporal Data. John Wiley & Sons, Hoboken, NJ.
Doblas-Reyes, F.J., Pavan, V., Stephenson, D.B., 2003. The skill of multi-model seasonal forecasts of the wintertime North Atlantic Oscillation. Clim. Dyn. 21, 501–514.
Efron, B., 1979. Bootstrap methods: Another look at the jacknife. Ann. Statistics 7, 1–26.
Giorgi, F., Mearns, L.O., 2002. Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "Reliability Ensemble Averaging" (REA) method. J. Clim. 15, 1141–1158.
Gleason, K.L., Lawrimore, J.H., Levinson, D.H., Karl, T.R., Karoly, D.J., 2008. A revised U.S. Climate Extremes Index. J. Clim. 21, 2124–2137.
Gleckler, P.J., Taylor, K.E., Doutriaux, C., 2008. Performance metrics for climate models. J. Geophys. Res. 113, D06104. http://dx.doi.org/10.1029/2007JD008972.
Groisman, P.A., Knight, R.W., Karl, T.R., Easterling, D.R., Sun, B., Lawrimore, J.H., 2004. Contemporary changes of the hydrological cycle over the contiguous United States: trends derived from in situ observations. J. Hydrometeorol. 5, 64–85.
Gutmann, E., Pruitt, T., Clark, M.P., Brekke, L., Arnold, J.R., Raff, D.A., Rasmussen, R. M., 2014. An intercomparison of statistical downscaling methods used for water resource assessments in the United States. Water Resour. Res. 50, 7167–7186.
Karl, T.R., Knight, R.W., Plummer, N., 1995. Trends in high-frequency climate variability in the twentieth century. Nature 377, 217–220.
Karl, T.R., Katz, R.W., 2012. A new face for climate dice. Proc. Nat. Acad. Sci. USA 109 (37), 14720–14721.
Kirtman, B., Power, S., Adedoyin, J.A., Boer, G., Bojariu, R., Camilloni, I., Doblas-Reyes, F.J., Fiore, A.M., Kimoto, M., Meehl, G.A., Prather, M., Sarr, A., Schär, C., Sutton, R., van Oldenborgh, G.J., Vecchi, G., Wang, H.J., 2013. Near-term climate projections and predictability. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. Cambridge and New York.
Knutti, R., 2010. The end of model democracy? Clim. Change 102, 395–404.
Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., Meehl, G.A., 2010. Challenges in combining projections from multiple climate models. J. Clim. 23, 2739–2758.
Knutti, R., Sedláček, J., 2013. Robustness and uncertainties in the new CMIP5 climate model projections. Nat. Clim. Change 3, 369–373.
Kucharik, C.J., Serbin, S.P., Hopkins, E.J., Vavrus, S., Motew, M.M., 2010. Patterns of climate change across Wisconsin from 1950 to 2006. Phys. Geogr. 31, 1–28.
Maloney, E.D., et al., 2013. North American climate in CMIP5 experiments: part III: assessment of twenty-first-century projections. J. Clim. 27, 2230–2270.
Masson, D., Knutti, R., 2011. Climate model geneology. Geophys. Res. Lett. 38, L08703.
Maurer, E.P., Wood, A.W., Adam, J.C., Lettenmaier, D.P., Nijssen, B., 2002. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. J. Clim. 15, 3237–3251.
Meehl, G.A., Stocker, T.F., Collins, W.D., Friedlingstein, P., Gaye, A.T., Gregory, J.M., Kitoh, A., Knutti, R., Murphy, J. M., Noda, A., Raper, S.C.B., Watterson, I.G., Weaver, A.J., Zhao, Z.-C. 2007. Global climate projections. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press. Cambridge, and New York. pp. 747–845.
Murphy, J.M., Sexton, D.M.H., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M., Stainforth, D.A., 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 430, 768–772.
Notaro, M., Lorenz, D.J., Vimont, D., Vavrus, S., Kucharik, C., Franz, K., 2011. 21st century Wisconsin snow projections based on an operational snow model driven by statistically downscaled climate data. Int. J. Clim. 31, 1615–1633.
Notaro, M., Williams, J.W., Lorenz, D., 2012. 21st century vegetation and land carbon projections for Wisconsin. Clim. Res. 54, 149–165.
Notaro, M., Lorenz, D., Hoving, C., Schummer, M., 2014. 21st century projections of snowfall and winter severity across central-eastern North America. J. Clim. 27, 6526–6550.
Pielke, R.A., Downton, M.W., 2000. Precipitation and damaging floods: trends in the United States, 1932–97. J. Clim. 13, 3625–3637.
Patz, J.A., Frumkin, H., Holloway, T., Vimont, D., Haines, A., 2014. Climate change: Challenges and opportunities for global health. J. Am. Med. Assoc. 312, 1565–1580.
Schär, C., Lüthi, D., Beyerle, U., Heise, E., 1999. The soil-precipitation feedback: A process study with a regional climate model. J. Clim. 12, 722–741.
Scherrer, S.C., Baettig, M.B., 2008. Changes and inter-model spread in 21st century scenarios for temperature and precipitation extremes as seen with the climate change index (CCI). Environ. Res. Lett. 3. http://dx.doi.org/10.1088/1748-9326/3/3/034005.
Schmittner, A., Latif, M., Schneider, B., 2005. Model projections of the North Atlantic thermohaline circulation for the 21st century assessed by observations. Geophys. Res. Lett. 32, L23710. http://dx.doi.org/10.1029/2005GL024368.
Seneviratne, S.I., Lüthi, D., Litschi, Schär, 2006. Land-atmosphere coupling and climate change in Europe. Nature 443, 205–209.
Stainforth, D.A., Allen, M.R., Tredger, E.R., Smith, L.A., 2007. Confidence, uncertainty, and decision-support relevance in climate predictions. Philos. Trans. R. Soc. A 365, 2145–2161.
Tebaldi, C., Smith, R.L., Nychka, D., Mearns, L.O., 2005. Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multimodel ensembles. J. Clim. 18, 1524–1540.
Tebaldi, C., Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate projections. Philos. Trans. R. Soc. A 365, 2053–2075.
Trenberth, K.E., 1999. Conceptual framework for changes of extremes of the hydrological cycle with climate change. Clim. Change 42, 327–339.
Vavrus, S., Van Dorn, J., 2010. Projected future temperature and precipitation extremes in Chicago. J. Gt. Lakes Res. 36, 22–32.
Vavrus, S.J., Behnke, R., 2013. A comparison of projected future precipitation in Wisconsin using global and downscaled climate model simulations: implications for public health. Int. J. Clim. 34, 3106–3124.
Veloz, S., Williams, J., Vimont, D., Vavrus, S., Lorenz, D., Notaro, M., 2012. Identifying climatic analogs for Wisconsin under 21st-century climate-change scenarios. Clim. Change 112, 1037–1058.
Walsh, J., Wuebbles, D., Hayhoe, K., Kossin, J., Kunkel, K., Stephens, G., Thorne, P., Vose, R., Wehner, M., Willis, J., Anderson, D., Doney, S., Feely, R., Hennon, P., Kharin, V., Knutson, T., Landerer, F., Lenton, T., Kennedy, J., Somerville, R., 2014. Chapter 2: our changing climate. In: Melillo, J.M., Richmond, Terese (T.C.), Yohe, G.W. (Eds.), Climate Change Impacts in the United States: The Third National Climate Assessment. U.S. Global Change Research Program, Washington, DC, pp. 19–67.
Wang, M., Overland, J.E., 2012. A sea ice free summer Arctic within 30 years: an update from CMIP5 models. Geophys. Res. Lett. 39, L18501. http://dx.doi.org/10.1029/2012GL052868.
Wehner, M., 2005. Changes in Daily Precipitation and Surface Air Temperature Extremes in the IPCC AR4 Models. US CLIVAR Variations, vol. 3. pp. 5–9.
Weigel, A.P., Knutti, R., Liniger, M.A., Appenzeller, C., 2010. Risks of model weighting in multimodel climate projections. J. Clim. 23, 4175–4191.
WICCI, 2011. Wisconsin's Changing Climate: Impacts and Adaptation. Wisconsin Initiative on Climate Change Impacts. Nelson Institute for Environmental Studies, University of Wisconsin–Madison and the Wisconsin Department of Natural Resources. Madison, Wisconsin.
Wood, A.W., Leung, L.R., Sridhar, V., Lettenmaier, D.P., 2004. Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. Clim. Change 62, 233–256.
Yao, Y., Luo, Yong, Huang, J., Zhao, Z., 2013. Comparison of monthly temperature extremes simulated by CMIP3 and CMIP5 models. J. Clim. 26, 7692–7707.
Zhang, X., Zwiers, F.W., Hegerl, G.C., Lambert, F.H., Gillett, N.P., Solomon, S., Stott, P. A., Nozawa, T., 2007. Detection of human influence on twentieth-century precipitation trends. Nature 448, 461–465.