Ensembles of dynamically downscaled simulations provide valuable information on regional climate change projections, but their interpretation remains challenging.

Uncertainty in climate change projections arises due to imperfections in climate models, the unpredictable nature of internal climate variability, and uncertainty surrounding the future pathway of greenhouse gas (GHG) emissions. As a result, the climate community relies on ensembles of climate model simulations that attempt to explore and span uncertainty in these factors. In addition to widely used multimodel ensembles (MMEs) of global climate models (GCMs; Meehl et al. 2007; Taylor et al. 2012), a dynamical downscaling technique has been used in an ensemble mode where multiple combinations of GCMs and regional climate models (RCMs) are used to capture information on climate change uncertainty at regional scales. Examples of such RCM ensembles include the North American Regional Climate Change Assessment Program (NARCCAP; Mearns et al. 2009, 2012) and the Prediction of Regional Scenarios and Uncertainties for Defining European Climate Change Risks and Effects (PRUDENCE; Christensen et al. 2007) and Ensemble-Based Predictions of Climate Changes and Their Impacts (ENSEMBLES; Christensen et al. 2009; van der Linden and Mitchell 2009) projects for Europe. The GCMs participating in phase 5 of the Coupled Model Intercomparison Project (CMIP5; Taylor et al. 2012) are now dynamically downscaled for many regions of the world as part of the Coordinated Regional Climate Downscaling Experiment (CORDEX; Giorgi et al. 2009; Jones et al. 2011), including simulations for North America.
(NA-CORDEX). Since downscaling provides climate information at high spatiotemporal (~25–50 km and subdaily) scales, the use of these products in impact assessments and to drive impact models has been commonplace. Numerous studies of hydrological and ecological impacts (Mearns et al. 2015 and references therein), as well as descriptions of regional projections in the U.S. National Climate Assessment (Walsh et al. 2014), have used NARCCAP simulations extensively. Given the utility of RCM ensembles, a discussion about the consequences of their experimental design on the exploration of uncertainty in projections will aid interpretation of their results.

The RCM ensembles are computationally burdensome and their design involves making pragmatic choices regarding the GCMs and RCMs to include, and the resolution and length of the RCM simulations. Selecting a small number of GCMs to drive RCMs is a necessary first step. A systematic approach to selecting GCMs for downscaling involves a thorough understanding of the models’ reliability to produce regional climate drivers (e.g., McSweeney et al. 2015), evaluation of the fields that constitute boundary conditions for the RCMs (e.g., Jury et al. 2015), and additionally also evaluating the interdependency between GCMs to select the most independent models that explore a wide variety of future climate projections (Evans et al. 2014). While useful, such studies are often hindered by data availability, computational constraints, and challenges associated with finding GCMs that have satisfactory performance for many fields across the full RCM domain. Consequently, in the absence of a more systematic approach, selecting GCMs with different equilibrium climate sensitivities (ECSS)\(^1\) is a simple choice to represent uncertainty in projections seen in the MME. The NARCCAP project involved the dynamical downscaling of four CMIP3 GCMs that all have climate sensitivities close to the middle of the Intergovernmental Panel on Climate Change’s (IPCC) likely range (Mearns et al. 2009), while the NA-CORDEX involves the downscaling of six GCMs that were chosen to span the range of ECSS in the CMIP5 ensemble (see https://na-cordex.org/briefing-document). This approach assumes that regional warming scales reasonably well with the ECSS in these models, and as a result, their downscaled projections are roughly consistent with the spread in CMIP5 projections. These assumptions and the impact of GCM selection, in both NARCCAP and NA-CORDEX, on representing uncertainty in regional climate change projections over North America have not yet been discussed thoroughly in the literature. Although a number of different sources of uncertainty remain unexplored in the MME of GCMs, the range produced by such an ensemble of opportunity (Tebaldi and Knutti 2007) is assumed to provide a reasonable spread in climate change projections and sampling this spread well is necessary to evaluate risks to diverse future outcomes.

In addition to exploring uncertainty in regional projections, the NARCCAP and NA-CORDEX projects are also designed to understand the behavior of the RCMs and their utility in providing skillful regional information (Giorgi et al. 2009). In theory, a higher-resolution RCM is expected to improve the quality of the climate information at regional scales. The assessment of the RCM performance over the historical period is then routinely used to discriminate among models and their projections. Numerous studies demonstrate the “added value” aspect of the RCMs (Rummukainen et al. 2016; Luca et al. 2012, 2016; Feser et al. 2011) owing to their strength in representing land surface features (e.g., lakes, mountains, coasts) and regional processes poorly resolved in the coarse-resolution GCMs. But these studies also conclude that generalizations about added value across variables and regions are difficult to make. Besides, model biases can be dependent on the climate state (Kerkhoff et al. 2014) and may not be strongly linked to the projections (Giorgi and Coppola 2010; Racherla et al. 2012), which has implications for using model performance to establish the credibility of their projections. The continuous integration of NA-CORDEX RCMs over the historical period (from 1950 to 2005), longer than the 30-year time slice setup used in NARCCAP, allows evaluation of the relationship between the quality of the RCMs in simulating seasonal mean climatology and the observed temperature trends. This helps us test if the more credible RCMs within the ensemble are likely to provide more reliable climate change projections. If not, exploring the spread in regional projections remains critically important.

Here, I focus on the two factors mentioned above that are critical to the interpretation of the results produced by an ensemble of RCM simulations: i) the impact of GCM selection on representing uncertainty in regional projections and ii) the assessment of model performance and how it is related to models’ ability to capture historical trends in temperature in the RCMs and the driving GCMs. These aspects are evaluated by analyzing NARCCAP and NA-CORDEX downscaled

\(^1\) The ECS is a quantitative measure that indicates a change in global mean temperature in response to an instantaneous doubling of CO2 concentrations after the climate system has reached radiative equilibrium.
Simulations in combination with their driving GCMs, and also by including the estimates of uncertainty in regional projections obtained from the full CMIP3 and CMIP5 ensembles. The results are presented for temperature and precipitation—two key surface variables used in impacts studies. The NARCCAP RCM simulations, in particular, have been thoroughly analyzed in previous studies (e.g., Mearns et al. 2012, 2013; Bukovsky 2012), while the NA-CORDEX project is currently under way. Therefore, this study only focuses on properties of the RCM ensembles within the context of exploring uncertainty in regional projections. Additionally, these cross-generational comparisons allow determination of how the newer CMIP5 and NA-CORDEX ensembles differ from the older and widely used CMIP3 and NARCCAP ensembles.

Data and methods. NARCCAP and NA-CORDEX, as well as CMIP3 and CMIP5 models included in this study, are described in the appendix (see Tables A1 and A2). All the model data are first regridded using bilinear interpolation onto a common 0.5° × 0.5° latitude–longitude grid for the calculation of model biases and onto a common 2.5° × 2.5° grid for comparing climate change projections. The climate model biases for the period 1980–98 are calculated using gridded observations: Climatic Research Unit Time Series (CRU TS version 3.23) data for temperature (Harris et al. 2014). The climate change projections are calculated for the 2042–69 period, which is common to all simulations, for Special Report on Emissions Scenarios (SRES) scenario A2 in the case of NARCCAP and CMIP3, and for representative concentration pathway 8.5 (RCP8.5) in the case of the NA-CORDEX and CMIP5 ensembles. The NA-CORDEX project is currently under way with RCM simulations at 50-, 25-, and 12-km resolutions and two RCPs: RCP8.5 and RCP4.5. This study includes only available NA-CORDEX RCM simulations at 50-km resolution to allow direct comparison with NARCCAP, which is also performed at 50-km resolution. The results are presented for six regions of the contiguous United States (CONUS; Fig. 1) used in the Third National Climate Assessment (NCA; Walsh et al. 2014) that are of suitable size for comparing multiple GCM and RCM ensembles.

Spanning the range of uncertainty in regional projections. The CMIP5 models downscaled in NA-CORDEX are chosen such that they largely span a range in ECS. The implicit expectation in this strategy is that the regional warming scales reasonably well with the ECS and, as a consequence, produces a comparable spread in regional responses. I evaluate if this expectation is true for regional warming in the United States. Figure 2a shows the relationship between the ECS and regional annual and seasonal mean warming for the period 2042–69 under RCP8.5. Moderate (r > 0.5) to strong (r > 0.8) linear relationships are seen between the ECS and annual mean warming across the CMIP5 models, suggesting that the high (low) ECS models such as HadGEM2-ES and CanESM2 (see appendixes for expanded model names) tend to project higher (lower) regional warming across the entire CONUS. While similar relationships are seen for seasonal warming as well, the correlations are not strong in the southeastern and mid-Atlantic regions in December–February (DJF). In fact, for DJF and March–May (MAM), the regional warming by the mid-twenty-first century does not scale very well with the ECS for much of the CONUS except for parts of the Southwest. This is further illustrated in Fig. 2b. Projected winter warming in the Southwest by CMIP5 models is strongly correlated to their ECS, but this is not true for the Southeast even though the selected high- and low-ECS models (HadGEM2-ES and GFDL-ESM2M) also have high and low projected warming. Note that there need not be a 1:1 relationship between the ECS and regional warming across the CMIP5 models because, unlike ECS, the regional warming estimates represent a transient climate response. Precipitation changes over North America, on the other hand, are driven by a combination of thermodynamic (through an increase in specific humidity in response to warming) and atmospheric circulation changes (e.g., storm tracks, subtropical highs), as well as changes in local dynamical processes (e.g., low-level jets, moisture flow) (see Seager et al. 2014). Therefore, a simple relationship between regional precipitation over land and a global measure such as an ECS across models is not expected or found (not shown).
There are cases where the full CMIP5 range in regional warming is not fully spanned by the selected models for downscaling. For instance, for both regions in Fig. 2b, there are at least two CMIP5 models that indicate greater warming (triangles along the x axis = 1; ECS estimates for these models are not available) than all of the models chosen for downscaling. As a result, even in cases where regional warming scales reasonably well with the ECS, the GCMs chosen for downscaling may underrepresent uncertainty in projections seen in the CMIP5 ensemble. To illustrate this further, Fig. 3 shows the fraction of the total range in seasonal temperature and precipitation projections spanned by the driving GCMs (DGCM hereafter) used in NARCCAP and NA-CORDEX relative to CMIP3 and CMIP5 ensembles, respectively. Blue colors denote regions where the range of projections spanned by the DGCMs is less than half of that in the corresponding MME of GCMs, whereas the orange/red colors show regions where DGCMs span 60%-100% of the MME projection ranges. Models that are outliers can affect these fractions since the ranges are calculated by taking the difference between the two extreme projections in the ensemble. But all available data are considered here while calculating the ranges since none of these models are systematically weighted or deemed implausible at simulating North American climate.

Figure 3 shows that, except for summer precipitation (top-right panel), the NARCCAP DGCMs do not span the range of CMIP3 projections for summer and winter temperature and precipitation in most of the CONUS. The fraction of the CMIP3 range spanned by the NARCCAP DGCMs is less than 0.7 across the CONUS for temperature projections in both DJF and June–August (JJA). This is to be expected given the fact that four DGCMs have climate sensitivities between 2.7° and 3.4°C, close to the center of the IPCC likely range (Mearns et al. 2009). In comparison, the NA-CORDEX GCMs span the CMIP5 ranges of projections well with a few exceptions: winter temperature projections in the mid-Atlantic and southeastern United States (Fig. 3, bottom left; <60%), winter precipitation projections in much of the midwestern and the eastern United States (<40%), and summer precipitation projections in the Great Plains region (<40%). The contrast between the seasons is clear for the southeastern United States, where the DGCMs span the temperature and precipitation projection in summer but not in winter, illustrating the difficulty
in selecting a small set of GCMs for downscaling. These results indicate that using ECS to select GCMs for downscaling is a useful method because of its simplicity, but has implications for spanning the ranges of uncertainty in regional climate change projections.

**COMPARISON BETWEEN REGIONAL PROJECTIONS IN GLOBAL AND REGIONAL MODELS.** Given the impact of GCM selection discussed in the previous section, it is important to understand how consistent the spread in downscaled projections is with uncertainties explored by the MMEs. Figure 4 shows comparisons for winter and summer mean temperature and precipitation projections between the CMIP3, CMIP5, NARCCAP, and NA-CORDEX ensembles for the NCA regions. First, the GCM MME ranges in both variables for all regions are of similar magnitude or higher in the CMIP5 ensemble (black boxplots and gray shading) than the CMIP3 ensemble (maroon boxplots and lines). While the lower end of the CMIP3 and CMIP5 regional temperature projections are within 0.5°C of each other during both seasons, more than 25% of the CMIP5 models indicate higher winter warming in all regions than all the CMIP3 models (Fig. 4a). For summer temperatures, however, the differences between the two GCM MME projections are small except for the Midwest (Fig. 4b). For precipitation, ranges produced by the full CMIP3 and CMIP5 ensembles are comparable except for winter precipitation in the Midwest and Northeast, where a number of CMIP5 models project much higher increases. In the two western regions [Northwest (NW), Southwest (SW)], the spread in CMIP5 summer projections is wider and indicates a tendency toward less dry or even wetter projections compared to CMIP3. In spite of significant uncertainty, the NA-CORDEX downscaled GCMs fully span the CMIP5 MME range in summer precipitation projections, as shown in Fig. 3, which is entirely due to two climate models: HadGEM2-ES and CanESM2. Interestingly, both models have high climate sensitivities, with projected summer warming of about 4°C in the Southwest, but with very different precipitation projections. The use of two different emissions scenarios in these MMEs (SRES-A2 and RCP8.5) may partly contribute to wider CMIP5 ranges, but the radiative forcing trajectories for both scenarios are very similar in the latter half of the twenty-first century, reaching roughly 8.5 W m⁻² by 2100 (Nakicenovic et al. 2000; Van Vuuren et al. 2011). Additionally, the broader range in the CMIP5 projections could be because the ensemble size is bigger, and the models included are more complex in their representation of the Earth system components and in their treatment of aerosol forcing compared to their predecessors.

The use of multiple RCMs with diverse physical formulations typically results in ranges of regional projections that are wider than those spanned by the downscaled GCMs. But with regard to spanning the full GCM MME range, the downscaled projections (all filled symbols in Fig. 4) do not sample the higher end of temperature projections well in both seasons. In fact, the winter warming produced by the RCMs in the Midwest (MW), Great Plains (GP), Northeast (NE), and Southeast (SE) regions span only about the lower 75% of CMIP5 projections. On the other hand, the RCMs do a good job at spanning the range in MME winter warming in the two western regions (NW, SW). Precipitation projections for the Northwest
in three NA-CORDEX models are about 10% higher than the highest NARCCAP RCM projections, which is due to downscaling of CanESM2, a GCM with the highest projected increase in precipitation for the region in the CMIP5 ensemble. These findings show that the NARCCAP data have not captured the

Fig. 4. Seasonal mean temperature vs precipitation projections for (a) winter (DJF) and (b) summer (JJA) for six NCA regions for the CMIP3, CMIP5, NARCCAP, and NA-CORDEX ensembles. The projections are calculated for 2042–69 relative to the 1980–98 mean. The open symbols indicate projections for the driving GCMs used in NARCCAP (blue) and NA-CORDEX (black) and the filled symbols show corresponding RCM projections. Black boxplots and gray shading indicate ranges spanned by the 36 CMIP5 models; brown boxplots and lines are for the 19 CMIP3 models.

in three NA-CORDEX models are about 10% higher than the highest NARCCAP RCM projections, which is due to downscaling of CanESM2, a GCM with the highest projected increase in precipitation for the region in the CMIP5 ensemble. These findings show that the NARCCAP data have not captured the
wetter projection in the western United States seen in the newer models. More region-specific analyses are, however, required to understand how potentially higher spatial variability in the RCMs may result in spanning the MME ranges locally.

These ensembles highlight a major uncertainty in summer precipitation projections. All four ensembles contain models with both wet and dry projections for summer in all regions (Fig. 4b). Projected changes in summer precipitation in the Northeast for the down-scaled CMIP3 GCMs have two models each with wet and dry projections. But when downscaled, all RCM projections indicate either a smaller increase or a larger decrease in precipitation than the driving GCMs. For instance, CCSM3 indicates an increase in summer precipitation in the future in all regions (except for NW), but when downscaled, all RCM projections are drier than the DGCM. While CGCM3 indicates both dry and wet projections across the CONUS, again the downscaled projections are all drier than the DGCM. A systematic change in NARCCAP models, where most downscaled projections are drier than the driving GCMs, has been discussed by Mearns et al. (2013) and Bukovsky et al. (2017) (see Fig. 2). In the case of NA-CORDEX, most CMIP5 drivers indicate small changes in precipitation, but when downscaled, the RCMs indicate drier projections than their drivers with the exception of Canadian Regional Climate Model version 4 (CanRCM4) driven by CanESM2. Mearns et al. (2013) noted a negative linear relationship between projected changes in summer temperature and precipitation in the central plains in the NARCCAP RCMs. This relationship holds when the CMIP5 driving GCMs and NA-CORDEX RCMs are included in the analysis, as can be seen in Fig. 4b. The Great Plains (includes the central plains), as well as the Southeast and the Midwest, exhibit anticorrelations between summer warming and precipitation changes in the GCMs and RCMs to varying degrees. These relationships are likely related to a future increase in evaporative demand with warming and a decrease in soil moisture availability (Cook et al. 2014; Berg et al. 2015). For winter precipitation, most CMIP3 and CMIP5 models, as well as their downscaled counterparts, indicate wetter conditions in the future except for a few projections in the two western regions.

Precipitation changes can exhibit large spatial variability and, as a result, simply examining a range in regional average precipitation is not sufficient. Maloney et al. (2014) compare CMIP3 and CMIP5 projections and show that both MMEs indicate increases in wintertime precipitation at mid- and high latitudes over North America and decreases in the far south, whereas the changes in summer precipitation are more spatially heterogeneous. Also, given the interesting response in the NARCCAP RCMs, here I focus on the summer mean precipitation projections in all four ensembles (Fig. 5). The spatial patterns of the ensemble mean projections are very similar in the CMIP3 and CMIP5 ensembles, indicating drying in parts of the western United States from the Northwest to the central plains and wetting in the eastern part of the country. The ensemble mean projections across the NARCCAP and NA-CORDEX DGCMs are also very similar to their parent multimodel ensemble means, suggesting that the downscaled models do indeed represent the MME.
As a result, even though they may not capture the time scales of GHG and aerosols over the historical period, the GCMs are forced by observed concentrations. The RCMs that downscale CanESM2 reproduce the wetting signal, which contributes to the ensemble mean seen in Fig. 5. The Southwest is an arid region and this large percentage increase in summer precipitation should not be overinterpreted. The NARCCAP and NA-CORDEX RCM projections are also different for the Midwest, where the former indicates a decrease in precipitation and the latter indicates an increase even though the driving models produce similar results in both cases (also seen in Fig. 4b). Additionally, the ensemble mean drying signal in NA-CORDEX in the central plains region is modest because of general disagreements among models, unlike the NARCCAP RCMs, which indicated a systematic decrease as mentioned earlier. A thorough analysis of the underlying physical mechanisms is required to understand the degree to which these disagreements are affected by GCM-RCM sampling.

MODEL PERFORMANCE AND REGIONAL TEMPERATURE TRENDS. A topic of important consideration is the ability of the driving GCMs and RCMs to simulate regional trends and their relation to model biases. Pierce et al. (2009) found no relationship between the performance of the CMIP3 models and their future projections in the case of winter temperatures in the western United States. Using a single GCM–RCM setup, Racherla et al. (2012) suggested that the RCM skill in simulating seasonal mean climate for regions in the CONUS is modest because of general disagreements among models, unlike the NARCCAP RCMs, which indicated a systematic decrease as mentioned earlier. A thorough analysis of the underlying physical mechanisms is required to understand the degree to which these disagreements are affected by GCM-RCM sampling.

Figure 6 shows observed and simulated temperature trends over the historical period plotted against model temperature biases for 1980–98. For six NCA regions, the results are shown for one season each in which the region experienced a maximum, statistically significant (at 95% level) warming trend between 1955 and 2004 in the CRU dataset. The estimates of historical climate change are based on linear trends in temperature in the observations and models. It is reasonable to analyze regionally averaged temperature trends as they are spatially uniform across each of the NCA regions (Walsh et al. 2014). The CanESM2 model has a large warm bias and slightly underestimates the observed winter warming in the Northeast and the Midwest (Fig. 6). This warm bias is reduced when CanESM2 is downscaled using three different RCMs, but the RCM estimates of observed warming are not altered substantially. On the other hand, the Max Planck Institute Earth System Model, medium resolution (MPI-ESM-MR), underestimates the observed warming in spring (MAM) in the Northwest and Southwest, but its downscaling counterpart using the CRCM5 shows better skill in capturing the magnitude of observed warming in both regions. Overall, most GCMs and RCMs show a systematic underestimation of seasonal mean temperature and observed warming in spring in the Northwest and Southwest. In the Northeast and Midwest regions, EC-EARTH severely underestimates observed winter warming and its two downscaled counterparts show mixed results, but all these models have comparable small warm biases. The Regional Climate Model,
version 4 (RegCM4), RCM is used to downscale three different GCMs (HadGEM2-ES, MPI-ESM-LR, and GFDL-ESM2M), and the downscaled results do not show improvements in most cases. For instance, in the Midwest, all three RegCM4 simulations show negative biases, as well as very similar observed warming regardless of the different warming estimates produced by the DGCMs. Overall, three RCMs—CanRCM4, CRCM5 (Ouranos Consortium), and the Weather Research and Forecasting (WRF) Model—are spectrally nudged to closely follow the large-scale atmospheric conditions of the driving GCM (see https://na-cordex.org/rcm-characteristics) and consequently produce observed changes in temperature very similar to their driving GCMs for all the regions analyzed in Fig. 6. A similar result was found in the case of nudged NARC-CAP simulations (Bukovsky 2012) as well. These results clearly indicate that the models’ credibility in simulating temperature change cannot be easily tied to their ability to capture historical mean climate. Also, the process of downscaling seems to alter temperature trends detected in the DGCMs, suggesting a strong influence of the RCMs’ physical formulation and their internal variability characteristics. Some of these aspects can be understood using the RCM runs driven by reanalysis data [European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim)], but these runs are not sufficiently long enough (starting in or after 1979) to calculate trends.

Figure 6 also shows the regional temperature change between 1955 and 2004 as captured by different realizations of three NA-CORDEX-driving GCMs. Although the biases in seasonal mean temperature for the period 1980–98 are very similar for different initial conditions realizations of the same model, they have different estimates of warming over the historical period. For instance, four different realizations of CanESM2 all show warm biases in winter of about 5°C in the NE and MW regions, but their warming estimates for the period 1955–2004 range from slight cooling (about −0.5°C) to substantial warming (greater than 2.5°C). The other two models with multiple IC realizations (HadGEM2-ES and MPI-ESM-LR) also show disagreements in their estimates of the observed warming during the winter in the Northeast and Midwest. The spread in estimates of warming across different IC realizations is expected to be less at lower latitudes and in summer when the temperature variability is comparatively low as seen in the results for the Southeast in JJA. Overall, these large variations in regional temperature trends between different members of the same model highlight the importance of modeled internal variability characteristics, mainly in the cold season in the northern United States, in
dictating trends for at least up to 50 years and for spatial scales as large as the NCA regions. The assessment of the driving models’ internal variability characteristics is, therefore, crucial for interpreting RCM results, especially when the RCM is spectrally nudged.

CONCLUSIONS AND DISCUSSION. Regional climate models are extremely valuable in resolving and understanding physical processes at local scales and in providing input at relevant spatial scales for region-specific impacts analysis. One of the primary goals of an ensemble of RCMs is to produce a range in regional climate change projections to allow for the assessment of impacts for diverse future projections. The comparisons between the ranges of projections spanned by the full multi-GCM ensembles and by the models chosen for downscaling in two RCM ensembles clearly show that there are regions across the United States where the choices are inadequate to span known ranges of seasonal mean temperature and precipitation projections. The six NA-CORDEX GCMs do not sample the higher end of the winter temperature projections as they span less than 60% of the full CMIP5 range (36 models) over much of the CONUS. For precipitation, it is the eastern half of the United States in winter and the Great Plains in summer where the NA-CORDEX GCMs do a poor job of representing the full range of CMIP5 projections. For the Southeast, the NA-CORDEX GCMs span the uncertainty in summer but not in winter; this fact clearly illustrates one of the many challenges in selecting GCMs for downscaling for a large domain such as North America. In the case of NARCCAP, apart from summer precipitation, the driving models span only 60% of the full CMIP3 projections. The effect of downscaling using multiple RCMs is to widen the range of regional projections in most cases, but the downscaled projections do not capture the high end of the temperature projections mainly in the eastern (NE and SE) regions. The ranges of downscaled regional precipitation projections, on the other hand, are consistent with or even greater than the GCM MME ranges, mainly during JJA when the modeling uncertainty in simulating precipitation is very large. Ensemble mean projections suggest that the available NA-CORDEX RCMs show a different summer precipitation response to NARCCAP in the western and lower Midwest regions. For the southwestern United States, the disagreement between the dry projections of the NARCCAP models and the wet projections of the NA-COREDEX models is likely a property of the models chosen for downscaling. Note that the ranges spanned by the RCM ensembles are based on averages over large NCA regions and may span the MME ranges locally owing to the higher resolution of the RCMs. The additional simulations in NA-CORDEX with higher spatial resolutions of 12 and 25 km will undoubtedly provide more information about the impact of improved resolution (e.g., Lucas-Picher et al. 2017) on capturing ranges in regional projections. Overall, this cross-generational evaluation highlights the persistence of modeling uncertainty in regional projections despite significant scientific improvements in newer models.

The analysis of the models’ ability to capture seasonal mean climate and temperature trends over the historical period highlights two major issues. First, the NA-CORDEX RCMs and the driving GCMs show large variations in their ability to simulate the observed temperature trends. These trends are affected by internal variability characteristics of both the RCMs and the driving GCMs and whether the RCMs are spectrally nudged. Additionally, multiple realizations of the driving GCMs exhibit very different temperature trends over the 50-yr historical period; therefore, downscaling only one realization of a subset of GCMs undersamples the uncertainty associated with internal climate variability. The NARCCAP RCMs have been shown to exhibit some success in simulating spring and winter temperature trends, but several disagreements between the observed and simulated interannual climate variations reflect complexities in the experimental design, the impact of regional forcings, and varying model structures (Bukovsky 2012; de Elía et al. 2013). The internal climate variability is also a key source of uncertainty in determining future precipitation changes at regional scales (Deser et al. 2012). For instance, the ensemble mean seasonal mean precipitation changes in the northeastern United States in NARCCAP RCMs have been shown to be within the estimated range of natural climate variability (Rawlins et al. 2012). These issues are likely to remain important in the case of NA-CORDEX as well.

The second issue is that the simulated temperature trends are not systematically related to model performance over the historical period. In fact, models with the same skill in capturing seasonal mean climatology have been found to exhibit very different temperature trends over the historical period. This suggests limiting their scope to use model performance to weigh or constrain projections. In any case, constraining RCM projections has proved difficult despite model improvements because it is challenging to identify models that consistently perform well across a range of metrics and regions as the process-based analyses of NARCCAP simulations indicate (e.g., Bukovsky et al. 2013; Thibeault and Seth 2015; Fan et al. 2015).
Additionally, using the ENSEMBLES RCMs, Christensen et al. (2010) discuss challenges and uncertainties in developing schemes to weigh model projections and demonstrate that assigning performance-based weights does not always improve the description of the mean climate. Consequently, spanning a range in regional projections seen in the full GCM ensemble discussed in this study remains an important criterion while selecting GCMs for downscaling (Evans et al. 2014; McSweeney et al. 2015) and while using downscaled scenarios in impacts studies.

ACKNOWLEDGMENTS. This research was supported by the U.S. Department of the Interior's Northeast Climate Adaptation Science Center (NE CASC) and by Grant G12AC00001 from the U.S. Geological Survey. Its contents are solely the responsibility of the author and do not necessarily represent the views of the NE CASC or the USGS. This manuscript is submitted for publication with the understanding that the U.S. government is authorized to reproduce and distribute reprints for governmental purposes. I sincerely thank three reviewers (Linda Mearns and two anonymous) for their careful review and insightful comments that have helped improve the manuscript substantially. I also thank Alexander Bryan, Raymond Bradley, and Marcelo Somos-Valenzuela for their input on an earlier version. I acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and I thank the NARCCAP and NA-CORDEX climate modeling groups for producing and making available the model output.

APPENDIX: CLIMATE MODEL DATA. NARCCAP data. The NARCCAP model pairs used in this study are shown by the letter X in Table A1. NCEP and TMSL indicate National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis driving data and 50-km global time-slice experiments, respectively. The NARCCAP data used in this paper were accessed through the Earth System Grid (www.earthsystemgrid.org/project/narccap.html). The information about the NARCCAP driving data (AOGCMs) was obtained from www.narccap.ucar.edu/about/aogcms.html.

NA-CORDEX data. The NA-CORDEX GCM-RCM pairs used in this study are shown by the letter X in Table A2. Only simulations that are carried out at 50-km (or 0.44°) resolution and for RCP8.5 are shown here. The letter N denotes simulations that are planned but where the data were not unavailable at the time of this study. The NA-CORDEX data (Mearns et al. 2017) used in this study were accessed through the Earth System Grid (www.earthsystemgrid.org/search/cordexsearch.html). The data are available at a common 0.44° grid indicated by an extension (NAM-44i) in the file names.

Table A1. NARCCAP GCM–RCM pairs used in this study. CRCM = Canadian Regional Climate Model; ECP2 = Experimental Climate Prediction Center Regional Spectral Model; HRM3 = Hadley Centre Regional Model version 3; MM5I = Penn State/NCAR Mesoscale Model version 5; WRFG = Weather Research and Forecasting Model.

<table>
<thead>
<tr>
<th>GCM–RCM</th>
<th>CRCM</th>
<th>ECP2</th>
<th>HRM3</th>
<th>MM5I</th>
<th>CRCM3</th>
<th>WRFG</th>
<th>TMSL</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCEP</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>HadCM3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCSM3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFDL</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGCM3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Table A2. NA-CORDEX GCM–RCM pairs used in this study. The pairs used in this study are shown by the letter X and the letter N denotes simulations that are planned but the data were not available at the time of this study. RCA4 = Rossby Center Regional Atmospheric Model, version 4; UQAM = Université du Québec à Montréal; RegCM4 = Regional Climate Model, version 4; WRF = Weather Research and Forecasting Model; CanRCM4 = Canadian Regional Climate Model version 4.

<table>
<thead>
<tr>
<th>GCM–RCM</th>
<th>ECS (°C)</th>
<th>CRCM5 (UQAM)</th>
<th>RCA4</th>
<th>WRF</th>
<th>CanRCM4</th>
<th>HIRHAM5</th>
</tr>
</thead>
<tbody>
<tr>
<td>HadGEM2-ES</td>
<td>4.6</td>
<td>X</td>
<td></td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CanESM2</td>
<td>3.7</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPI-ESM-LR</td>
<td>3.6</td>
<td>X</td>
<td>X</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPI-ESM-MR</td>
<td>3.4</td>
<td></td>
<td></td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC-Earth</td>
<td>3.3</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFDL-ESM2M</td>
<td>2.4</td>
<td></td>
<td>X</td>
<td>N</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**CMIP3 models.** The CMIP3 ensemble used in this study includes the following 19 models: Bjerknes Centre for Climate Research Bergen Climate Model, version 2.0 (BCCR-BCM2.0); Canadian Centre for Climate Modelling and Analysis Coupled Global Climate Model, version 3.1 (CCMC CGCM3.1); Centre National de Recherches Météorologiques Coupled Global Climate Model, version 3 (CNRM-CM3); Commonwealth Scientific and Industrial Research Organisation Mark 3.0 (CSIRO-Mk3.0); CSIRO Mark 3.5 (CSIRO-MK3.5); Geophysical Fluid Dynamics Laboratory Climate Model, version 2.1 (GFDL CM2.1); Goddard Institute for Space Studies Model E, coupled with the Russell ocean model (GISS-ER); Istituto Nazionale Di Geofisica E Vulcanologia ECHAM, version 4 (INGV-ECHAM4); third-generation Institute of Numerical Mathematics of the Russian Academy of Sciences Climate Model (INMCM3.0); L’Institut Pierre-Simon Laplace Coupled Model, version 4 (IPSL-CM4); Model for Interdisciplinary Research on Climate, version 3.2 (medium resolution) [MIROC3.2 (medres)]; Meteorological Institute of the University of Bonn, ECHAM4 and the global Hamburg Ocean Primitive Equation (ECHO-G) Model (MIUBECHOG); MPI ECHAM5 (MPI ECHAM5); Meteorological Research Institute Coupled Atmosphere–Ocean General Circulation Model, version 2.3.2a (MRI-CGCM2.3.2a); NCAR Community Climate System Model, version 3 (NCAR CCSM3); NCAR Parallel Climate Model, version 1 (NCAR PCM1); Met Office Hadley Centre Coupled Model, version 3 (UKMO HadCM3); and Hadley Centre Global Environment Model, version 1 (UKMO HadGEM1).

**CMIP5 models.** The CMIP5 ensemble used in this study includes the following 36 models: Australian Community Climate and Earth–System Simulator, version 1.0 (ACCESS1.0); ACCESS, version 1.3 (ACCESS1.3); Beijing Climate Center, Climate System Model, version 1.1 (BCC_CM1.1); BCC_CSM (moderate resolution) [BCC_CSM1.1(m)]; Beijing Normal University–Earth System Model (BNU-ESM); Second Generation Canadian Earth System Model (CanESM2); Community Climate System Model, version 4 (CCSM4); Community Earth System Model, version 1 (bico-geochemistry, or carbon cycle) [CESM1(BGC)]; CESM1 (Community Atmosphere Model, version 5) [CESM1(CAM5)]; Centro Euro-Mediterraneo per I Cambiamenti Climatici Carbon Cycle Earth System Model (CMCC-CESM); CMCC Climate Model (CMCC-CM); CMCC Stratosphere-resolving Climate Model (CMCC-EMS); CNRM-CM, version 5 (CNRM-CM5); CSIRO Mark 3.6.0 (CSIRO-Mk3.6.0); EC-Earth Consortium (EC-EARTH); Flexible Global Ocean–Atmosphere–Land System Model, gridpoint version 2.0 (FGOALS-g2); First Institute of Oceanography Earth System Model (FIO-ESM); GFDL CM, version 3 (GFDL CM3); GFDL with GOLD component (GFDL-ESM2G); GFDL Earth System Model with MOM, version 4 component (GFDL-ESM2M); Goddard Institute for Space Studies Model E2, coupled with the Hybrid Coordinate Ocean Model (HYCOM) (GISS-E2-H); GISS-E2 coupled with the Russell ocean model (GISS-E2-R); HadGEM2 with the atmosphere and land and ocean and sea ice configurations (HadGEM2-AO); HadGEM2 with the carbon cycle configuration (HadGEM2-CC); HadGEM2 with the Earth System configuration (HadGEM2-ES); fourth-generation INM Climate Model (INMCM4); IPSL Coupled Model, version 5A, low resolution (IPSL-CM5A-LR); IPSL-CM5A, medium resolution (IPSL-CM5A-MR); IPSL CM version 5B, low resolution (IPSL-CM5B-LR); Model for Interdisciplinary Research on Climate, Earth System Model (MIROC-ESM); MIROC-ESM Chemistry Coupled (MIROC-ESM-CHEM); Model for Interdisciplinary Research on Climate, version 5 (MIROC5); MPI-ESM-LR; MPI-ESM-MR; Meteorological Research Institute Coupled Atmosphere–Ocean General Circulation Model, version 3 (MRI-CGCM3); and Norwegian Earth System Model, version 1 (intermediate resolution) (NorESM1-M).

**REFERENCES**


McSweeney, C., R. Jones, R. Lee, and D. Rowell, 2015: Selecting CMIP5 GCMs for downscaling over multiple


