A Comparison of Statistical and Dynamical Downscaling of Winter Precipitation over Complex Terrain

ETHAN D. GUTMANN

Research Applications Laboratory, and Advanced Studies Program, National Center for Atmospheric Research, Boulder, Colorado*

ROY M. RASMUSSEN, CHANGHAI LIU, KYOKO IKEDA, DAVID J. GOCHIS, AND MARTYN P. CLARK

Research Applications Laboratory, National Center for Atmospheric Research, Boulder, Colorado*

JIMY DUDHIA

Mesoscale and Microscale Meteorology Division, National Center for Atmospheric Research, Boulder, Colorado*

GREGORY THOMPSON

Research Applications Laboratory, National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 5 October 2010, in final form 16 June 2011)

ABSTRACT

Statistical downscaling is widely used to improve spatial and/or temporal distributions of meteorological variables from regional and global climate models. This downscaling is important because climate models are spatially coarse (50–200 km) and often misrepresent extremes in important meteorological variables, such as temperature and precipitation. However, these downscaling methods rely on current estimates of the spatial distributions of these variables and largely assume that the small-scale spatial distribution will not change significantly in a modified climate. In this study the authors compare data typically used to derive spatial distributions of precipitation [Parameter-Elevation Regressions on Independent Slopes Model (PRISM)] to a high-resolution (2 km) weather model [Weather Research and Forecasting model (WRF)] under the current climate in the mountains of Colorado. It is shown that there are regions of significant difference in November–May precipitation totals (>300 mm) between the two, and possible causes for these differences are discussed. A simple statistical downscaling is then presented that is based on the 2-km WRF data applied to a series of regional climate models [North American Regional Climate Change Assessment Program (NARCCAP)], and the downscaled precipitation data are validated with observations at 65 snow telemetry (SNOTEL) sites throughout Colorado for the winter seasons from 1988 to 2000. The authors also compare statistically downscaled precipitation from a 36-km model under an imposed warming scenario with dynamically downscaled data from a 2-km model using the same forcing data. Although the statistical downscaling improved the domain-average precipitation relative to the original 36-km model, the changes in the spatial pattern of precipitation did not match the changes in the dynamically downscaled 2-km model. This study illustrates some of the uncertainties in applying statistical downscaling to future climate.

1. Introduction

Accurate knowledge of the current and future spatial distribution of snow in complex terrain is critical for water resource management in the western United States. Water for most residential, agriculture, and local ecosystems comes primarily from spring snowmelt. Despite the

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Ethan Gutmann, RAL, NCAR, P.O. Box 3000, Boulder, CO 80307-3000.
E-mail: gutmann@ucar.edu

DOI: 10.1175/2011JCLI4109.1

importance of snowpack, we do not currently have a good way to estimate future changes in snowpack with adequate spatial resolution over complex terrain to physically model changes in streamflow (Stern and Easterling 1999).

Current predictions of future precipitation come from global climate models with grid sizes of 100–200 km, and even current regional climate models (RCMs) are run at resolutions far too low [e.g., North American Regional Climate Change Assessment Program (NARCCAP), 50 km (Mearns et al. 2009); Climate Change and its Impacts Ensembles (ENSEMBLES), 50–25 km (Hewitt and Griggs 2004); Nested Regional Climate Model (NRCM), 36 km (Hurrell et al. 2008)] to accurately capture this distribution. The most common approach to improve the resolution of precipitation for hydrologic applications is statistical downscaling. Snowfall typically increases over higher topographic elevations, and this relationship is exploited by the most current spatial distribution estimates (e.g., Daly et al. 1994; Maurer et al. 2002; Daly et al. 2008).

Statistical downscaling methods have been developed by numerous researchers (e.g., Hayhoe et al. 2004; Wood et al. 2004; Wilby et al. 1998; Hewitson and Crane 1996), but all of these have suffered from a lack of high-resolution data on which to base their downscaling. These studies downscale climate data to match a time series at a point, or a gridded-reanalysis-type product, which may have a spatial resolution of tens of kilometers (e.g., Wood et al. 2004). Indeed, one of the most commonly used high-resolution downscaled products for current and future climate (Maurer et al. 2007) achieves a grid size of $\frac{1}{8}^{\circ}$ ($\sim$12 km). While this resolution is a great improvement over the hundreds of kilometers from GCM output, it is larger than topographic variations that can drive significant precipitation in complex terrain ($\sim$1–5 km, Colle et al. 2000; Medina et al. 2005; Garvert et al. 2007) and is based on a Parameter-Elevation Regressions on Independent Slopes Model (PRISM) (Daly et al. 1994) type interpolation between measurements. As a result, this method only works in regions of the globe with extensive existing measurement networks such as the United States and Europe; important mountainous regions, such as the Andes and the Himalayas, have few measurements available. For hydrologic simulations, the inability of a 10-km forcing dataset to adequately resolve important topographic variations in the Colorado Rockies presents serious limitations for snowpack modeling.

Much of the recent statistical downscaling literature has focused on reproducing realistic time series of weather at a point (e.g., Gangopadhyay et al. 2005; Vrac et al. 2007). These studies have worked toward creating a better reproduction of the probability distribution function of variables, such as precipitation, with a focus on extreme events, which have been difficult to get correct. While this is an important topic, the appropriate spatial distribution of precipitation in complex terrain has largely been driven by refinement to the PRISM method (e.g., Daly et al. 1994, 2008).

The PRISM method is very powerful, but it is particularly limited in regions with few observations and is primarily controlled by elevation. While elevation may be the primary driver in precipitation spatial variability, it is far from the only one. Other important factors such as local frontal systems, topographic channeling of flow, and the interplay of atmospheric dynamics with hydrometeor microphysics play a large role in the distribution of precipitation (Colle et al. 2000; Medina et al. 2005; Garvert et al. 2007). These factors can be included in a dynamic model, but are difficult—if not impossible—to account for with a statistical topographic index; as such, they are not included in statistically downscaled grids of future climate (e.g., Hayhoe et al. 2004; Maurer et al. 2007). Furthermore, these dynamics may change in a future climate (Rasmussen et al. 2011); as a result, statistical downscaling based on current climatic patterns may not be valid in a future climate. It should be noted that some aspects of climate change might be amenable to statistical techniques.

Current study

Here we use output from the high-resolution (2 km) Weather Research and Forecasting (WRF) model, version 3.0 (Skamarock and Klemp 2008), to develop a statistical downscaling for winter precipitation. Properly validated, these data provide a new method of estimating the spatial distribution of precipitation over complex terrain. This is particularly important as it demonstrates a method to derive spatially distributed statistical downscaling products in regions that currently lack extensive measurements. In addition, a 2-km WRF climate change experiment (Rasmussen et al. 2011) provides a new method to assess the viability of a simple statistical downscaling method in a future climate scenario. By comparing dynamical downscaling to a statistical downscaling based on WRF data instead of the more traditional downscaling based on PRISM data, we are able to stress the differences due purely to the statistical versus dynamical approach. If we used PRISM data to develop the statistical downscaling, then it would be difficult to distinguish the differences due to differences between WRF and PRISM from those due to differences between a statistical and a dynamical climate change methodology.

In section 2, we present the datasets used in this study, including a short description of the WRF model, an overview of the NARCCAP simulations, snow telemetry (SNOTEL) observational data, and the PRISM dataset.

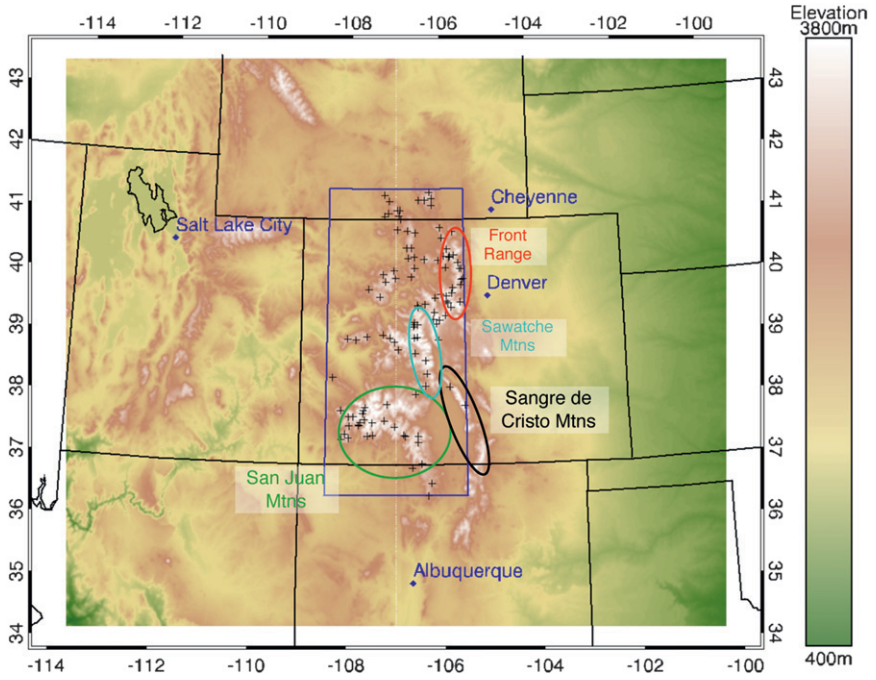

FIG. 1. Two-kilometer gridcell topography within the model domain. The area outside the model domain is in white. The subdomain over which proceeding figures and statistics are based is outlined in blue; SNOTEL stations used in this study are noted with black crosses. The subdomain encompasses the Colorado Rocky Mountains; specific mountain ranges referred to in the text are labeled.

In section 3, we present the simple downscaling method used in this study. In section 4, we present the experiments used to 1) verify the downscaling in the current climate, 2) compare our downscaling to the PRISM dataset, and 3) estimate statistical downscaling skill under an imposed future climate. In section 5, we present the basic results from these experiments followed by a discussion of the differences between various datasets and the implications for statistical downscaling of future climate in section 6.

2. Datasets

a. WRF

1) MODEL PHYSICS

We used the high-resolution WRF simulations described by Rasmussen et al. (2011) for the work presented in this study. These simulations were performed with WRF, version 3.0 (Klemp et al. 2007; Skamarock and Klemp 2008), a compressible, nonhydrostatic atmospheric weather model. The model was configured with the Mellor–Yamada–Janjic (MYJ) planetary boundary layer scheme (Skamarock and Klemp 2008), Community Atmosphere Model (CAM) longwave and shortwave radiation scheme (Collins et al. 2004), Thompson et al. (2004, 2008) cloud microphysics scheme, the Noah land surface model (Chen and Dudhia 2001), and an enhancement to the incident radiation calculation to include the effects of terrain slope and aspect. These schemes were chosen based on comparisons between 14-day simulations and SNOTEL data (Liu et al. 2011). These simulations were performed over a single domain (Fig. 1) 1200 km east to west by 1000 km north to south, with 45 vertical levels, a 2-km horizontal grid spacing, and a 15–20-s time step with model output every 30 min.

In addition to the high-resolution (2 km) WRF simulations described in Rasmussen et al. (2011), we used low-resolution (36 km) WRF simulations using the same parameterizations and forcing data. The primary differences between the 36-km simulations and the 2-km simulation were the horizontal grid spacing and the use of a convective parameterization in the 36-km simulation. The 2-km simulation is able to resolve convection explicitly, while the 36-km parameterization used the Kain–Fritsch (KF) convective parameterization (Kain 2004). Ikeda et al. (2010) found no significant difference between different convective parameterizations for seasonal snowpack evolution. These low-resolution runs are used to allow us to compare two models that are as similar as possible except for their grid spacing.

2) BOUNDARY/INITIAL CONDITIONS

Both high- and low-resolution WRF simulations used the North American Regional Reanalysis (NARR) product (Mesinger et al. 2006) for the initial and boundary conditions. This allows us to perform a retrospective study comparing model output with observations. The NARR dataset has a 3-h time step and a 32-km horizontal grid resolution. A complete (three dimensional) depiction (land surface and atmosphere conditions) from the NARR dataset was used to initialize the model, and the 3-h product was used for lateral boundary conditions for the six-month simulations. This setup was used to drive WRF for four independent 6-month simulations beginning on 1 November in 2001, 2003, 2005, and 2007. These years were selected to be representative of dry, normal, normal, and wet years, respectively. The 36-km model simulations were only performed for the 2007/08 period.

3) PSEUDO GLOBAL WARMING

In addition to these current climate simulations, we perform experiments for an imposed pseudo-global-warming (PGW) climate using the same NARR boundary conditions perturbed with a climate change signal as described in Rasmussen et al. (2011). This climate change experiment is based on the methods of Schar et al. (1996), Hara et al. (2008), and Kawase et al. (2008), with a verification using historical data presented in Hara et al. (2008). In this experiment the original NARR boundary conditions used to force the model are perturbed by adding a temporally constant, spatially variable mean climate shift to each of the primary atmospheric variables: wind speed and direction, geopotential height, air temperature, and water vapor mixing ratio. The initial model conditions are similarly perturbed, including the land surface soil temperatures.

Rasmussen et al. (2011) derived the climate perturbations from the difference between two 10-yr intervals from Community Climate System Model (CCSM) A2 emissions scenarios: 1995–2005 and 2045–55. The CCSM runs were generated by the National Center for Atmospheric Research (NCAR) Climate and Global Dynamics division as part of the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (Collins et al. 2006). The primary perturbation applied to the NARR boundary conditions included a mean temperature increase of 2°C and a 10%–15% increase in the water vapor mixing ratio; the exact perturbations to each varied both horizontally and vertically as derived from the CCSM 1.4° grid. Rasmussen et al. (2011) also applied perturbations to the wind field and pressure levels as specified by the CCSM output. This climate perturbation field was developed separately for each month of the year. Complete details on the climate perturbation methodology are given in Rasmussen et al. and are not repeated here.

Because a mean climate perturbation is applied to the same NARR boundary conditions, the weather in the PGW run is essentially the same as that of the current climate run. For example, a storm system will enter both the PGW and the current climate model domain at the same time with the same magnitude (relative to other storm systems in the same run). This is identified by Rasmussen et al. as one of the greatest weaknesses of the PGW method because it means that shifts in the storm track under a future climate will not be included in the model.

b. PRISM

We use precipitation climatology data from the PRISM dataset (Daly et al. 1994, 2008). This dataset is a spatial interpolation between observational stations. It relies on the inherent correlation between elevation and precipitation or temperature. The PRISM precipitation dataset is created by computing the elevation–precipitation relation for all stations within a subregion, with individual stations weighted differently as described later, and applying that relation to a topographic basemap. Subregions are defined by contiguous regions of consistent topographical aspect (Daly et al. 1994).

This interpolation is further modified by applying different weights to each station depending on the potential presence of an inversion layer, coastal proximity, and by the expected effectiveness of local terrain at orographic lifting (Daly et al. 2008). The inversion layer is only used in the calculation of temperature fields, and the coastal proximity is irrelevant to the current study. The effectiveness of local terrain is defined by a combination of filters applied to an 800-m digital elevation model (DEM). First, the minimum elevation in a 22-km radius is calculated, then the average of those minima over a 11-km radius is calculated, then this smoothed minimum surface is subtracted from the original DEM. Finally, the average over a 15-km radius is computed for this difference map. The exact radii used for these filters may vary, but these are the values presented in Daly et al. This smoothed difference map is finally used to group the terrain into three classes of effectiveness. The least effective class has a value less than 75 m, an intermediate class is defined between 75 and 250 m, and the most effective class has a value greater than 250 m. These classes are then used to place limits on the slope of the local precipitation elevation regression line. The majority of the mountainous terrain in the domain of this study is in the most effective class, although the edges of the mountain are intermediate, and

the flat areas, including the plains east and west of the mountains, and the San Luis Valley are in the least effective class.

We use the 800-m monthly climatology dataset available from the PRISM group Worldwide Web site; this climatology is generated for the period 1971–2000. To account for differences between the resolution of the PRISM climatology (800 m) and the WRF model (2 km), we aggregated the PRISM data to the 2-km WRF grid by averaging all of the PRISM grid cells whose center point was within a coarser 2-km grid cell. While it does not provide a perfect match to the WRF high-resolution simulations, the higher spatial resolution was considered preferable to the 4-km monthly dataset. We also performed the analysis presented in section 4 using the 4-km monthly dataset and found similar results (not presented here).

c. NARCCAP

We use data from the NARCCAP–National Centers for Environmental Prediction (NCEP) experiment. These data include six models—all forced with NCEP reanalysis boundary conditions (Mearns et al. 2009). The six models are the International Center for Theoretical Physics Regional Climate Model, version 3 (RegCM3); the WRF; the Canadian Regional Climate Model (CRCM); the Scripps Experimental Climate Prediction Center Regional Spectral Model (ECPC RSM); the fifth-generation Pennsylvania State University–NCAR Mesoscale Model (MM5); and the Hadley Centre Regional Model, version 3 (HadRM3). The models were run with NCEP boundary conditions from 1979 to 2004. All models have approximately the same domain with a western boundary around −150°, an eastern boundary around −50°, a northern boundary around 70°, and a southern boundary around 20°. While the exact location of the boundaries varies from model to model, in all cases the boundaries are sufficiently far away from the high-resolution WRF model domain so that edge effects are relatively unimportant. These models all have a 50-km grid spacing, and lateral boundary and sea surface temperatures are updated every 6 h throughout the run. These models represent a range of possible regional climate model parameterizations; relevant model characteristics are outlined in Table 1.

Precipitation data from the NARCCAP models were recently evaluated by Wang et al. (2009), who suggest that these models systematically overpredict winter precipitation in the intermountain west, underestimate the seasonal cycle, but reasonably reproduce the semiannual cycle. Their work also suggests that these models do a reasonable job of capturing the first-order variation between major regions (e.g., the Pacific Northwest

TABLE 1. NARCCAP model characteristics [from the NARCCAP Internet site (www.narccap.ucar.edu/)].

	CRCM	ECPC RSM	HadRM3	MM5	RegCM3	WRF
Boundary layer	Local K, gradient Richardson number formulation	Hong–Pan nonlocal K	First-order turbulent mixing	Hong–Pan (MRF)* countergradient, nonlocal K	Nonlocal K, countergradient flux	Explicit entrainment
Explicit moist physics	Removal of supersaturation	Removal of supersaturation	Prognostic cloud liquid and ice; liquid potential temperature	Dudhia simple ice	SUBEX** prog. cloud water	Prognostic cloud liquid and ice, rain, snow
Cumulus parameterization.	Mass flux	Simplified Arakawa–Schubert	Mass flux, including downdraft	KF mass flux	Grell with Fritsch–Chappell closure	KF mass flux
Vertical layers	29	28	19	23	18	35
Vertical coordinate type	Gai–Chen scaled height	Normalized pressure	Hybrid terrain following and pressure	Sigma	Terrain following	Terrain following
Time step (s)	900	100	300	120	150	150
Spectral nudging	No	Yes	No	No	No	Yes

* MRF: medium-range forecast.
** SUBEX: subgrid explicit moisture scheme.

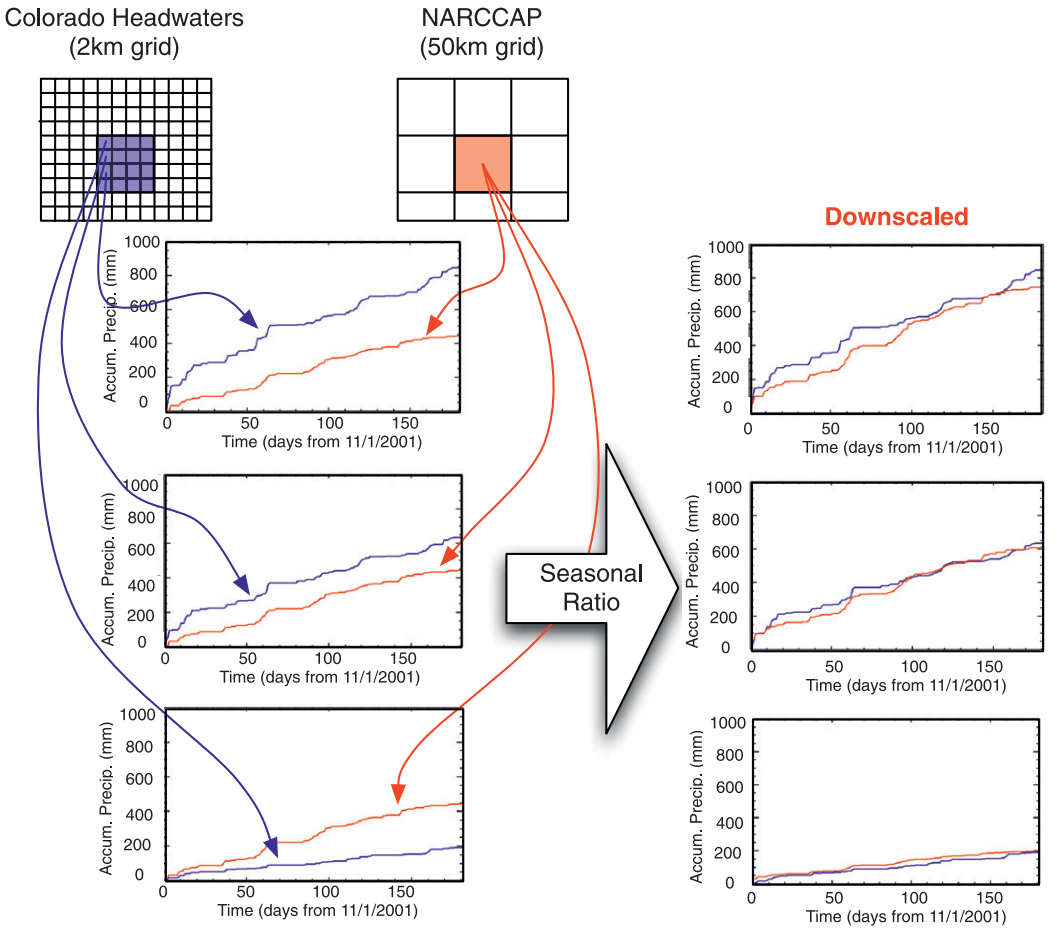

FIG. 2. Statistical downscaling of a time series of accumulated precipitation at three WRF 2-km grid cells (blue), with NARCCAP data (red) coming from the single encompassing grid cell in 2001/02; statistical downscaling was trained with data from 2003 to 2004. (right) The downscaled output matches the dynamically downscaled WRF 2-km model output better than (left) the raw NARCCAP output, but it still does not match perfectly.

Cascades, the Colorado Rockies, and central Arizona), though only CRCM appeared to capture the summer monsoon in central Arizona. They stress that caution must be used in interpreting the NARCCAP precipitation data over the intermountain west.

d. SNOTEL

We use precipitation data from 81 sites in the SNOTEL network (Fig. 1) as verification data. The SNOTEL precipitation measurement is made with an alter-shielded, weighing-type precipitation gauge. These gauges are typically placed in a forest clearing between 2400- and 3600-m elevation. The main problem associated with weighing gauge measurement of snowfall is undercatch that increases with wind speed. Because these gauges are typically located in a forest clearing, relatively low wind speeds (2 m s^{-1}) are common. These wind speeds correspond to an undercatch of 10%–15% (Yang et al. 1998; Serreze et al. 1999). SNOTEL sites are operated by the Natural Resources Conservation Service (NRCS), and all of the data are publicly available [www.wcc.nrcs.usda.gov/snow; see Serreze et al. (1999) for a description.] Of the 81 sites used in this study, only 46 were active prior to 1985, while 65 were active from 1987 forward. To maximize the number of stations with complete records and the number of years of intercomparison, we only compare SNOTEL data with NARCCAP output from after 1987, but we compare all 81 sites with the 2-km WRF simulations.

3. Downscaling

In our downscaling, we calculate the ratio between the low-resolution model precipitation and the high-resolution model precipitation and use that ratio to scale additional low-resolution model output (Fig. 2). We use the ratio

between accumulated precipitation at the end of each snow season (1 November–1 May) rather than individual days or storms. For each grid cell in the high-resolution model, we select the closest (i.e., nearest neighbor) grid cell from the low-resolution model. We select the nearest-neighbor low-resolution model grid cell because any interpolation of the low-resolution model will inherently add a temporal smoothing function to the data. This smoothing occurs because a precipitation event is unlikely to reach two grid cells at the same time.

When applied to two models with identical boundary conditions and time periods, this approach is functionally similar to the bias corrected spatial disaggregation (BCSD) approach of Wood et al. (2004). The BCSD approach takes a substantially more sophisticated approach to derive this ratio because climate models are not directly comparable to observations in any particular year. While this method may not be appropriate for analysis of extreme weather events or frequencies, it is useful for seasonal precipitation totals. Future work will utilize more sophisticated statistical methods.

We then apply this downscaling to the low-resolution model for the entire low-resolution model simulation of interest. Because we only performed the WRF model runs for the mountain snow season (1 November–1 May), the ratio developed here is not applicable to the summer season. As such, we only apply the ratio to low-resolution models during the same winter months.

While the results presented here are for seasonal totals, the downscaling method could be applied to an hourly or other temporal frequency dataset if the low-resolution model is deemed to sufficiently capture the temporal variability. However, this will lead to transient problems because an individual storm may take several hours or even days to traverse a single low-resolution model grid cell, while a simple statistically downscaled product will be forced to assume it reaches the entire grid cell simultaneously.

4. Experiments

First, we compare the dynamically downscaled WRF model output with the statistically interpolated precipitation from PRISM. We compare the two products in two ways: first, in a map view of differences and, second, in three east–west cross-sectional views across different mountain ranges. We then compare both of these products with SNOTEL observations from 81 stations for the WRF simulation periods. While we have no purely independent ground truth to show which is correct, the differences are illustrative of possible improvements gained by basing a statistical downscaling on an initial high-resolution dynamical downscaling.

Second, we test the statistical downscaling under the current climate in two ways. First, we calculate the ratio of NARCCAP 50-km precipitation to WRF 2-km precipitation using accumulated precipitation from water year 2001/02 and apply the ratio to the NARCCAP model output from water year 2003/04. The differences between the downscaled NARCCAP 2003/04 and the WRF 2-km 2003/04 precipitation illustrate the errors between a statistical and a dynamical downscaling, but they are also affected by the real spatial differences between the years, as such it may represent an overestimate of the expected errors. Ideally, a statistical downscaling technique would be trained with a large number of years, thus decreasing the error in the training dataset. However, the use of the same 2-km WRF model parameterization for both development and verification of the downscaling technique will also lead to a compensating underestimate of the expected errors.

We then calculate the downscaling ratio using the accumulated precipitation from water years 2001/02 and 2003/04 and apply this ratio to the NARCCAP data from 1988 through 2000 and compare downscaled precipitation to precipitation measured at 65 SNOTEL sites throughout Colorado, northern New Mexico, and southern Wyoming. Because we were only using winter precipitation to develop the downscaling, we only analyze SNOTEL precipitation over the winter seasons (1 November–1 May). By comparing the relationship between the original NARCCAP data and the SNOTEL data with the relationship between the downscaled NARCCAP data and the SNOTEL data, we illustrate the potential improvements that result from such a downscaling. It should be noted that a downscaling based on PRISM would be likely to produce as good or better results; however, this would not be an independent verification because the PRISM dataset incorporates these same SNOTEL data.

Finally, we test the downscaling approach under a climate change scenario. To do this, we develop the statistical downscaling for the 36-km WRF model of the 2007/08 water year and apply it to the PGW 36-km WRF model run described in section 2a. This downscaled output is then compared with a PGW 2-km WRF model run. Because the current and future climate runs have essentially identical weather patterns, the differences between the two show the changes under a modified climate that would not be captured by most statistical downscaling techniques. Indeed, because we are using the same high-resolution model for validation that we used to develop the statistical downscaling, the results presented here are likely to represent a best case for statistical downscaling, even given the simplicity of the downscaling approach.

5. Results

a. PRISM

Figure 3 presents climatology maps of accumulated November–May precipitation from both WRF 2 km (climatology from 2001/02, 2003/04, 2005/06, 2007/08) and PRISM (climatology from 1971 to 2000), as well as a difference map between the two. Despite the fact that the WRF climatology is limited to a 4-yr period at present, it still represents a reasonable average because it contains a relatively dry year (2001/02), a wet year (2007/08), and two normal years (2003/04, 2005/06). Both precipitation maps show the same broad spatial patterns representing the influence of the underlying topography. The difference map does not necessarily represent errors in either product, as both have their own errors. The spatially averaged difference between WRF and PRISM over the subdomain is −23 mm with a rms difference of 72 mm. The spatially averaged difference between the two for the low-resolution monthly PRISM data is −47 mm, rms difference of 81 mm. The fact that the high-resolution climatology is closer to the WRF average precipitation than the lower-resolution PRISM data from the same four years as the WRF data suggests that the trade-off of higher resolution for nonoverlapping time periods is beneficial.

Cross sections of topography and accumulated precipitation are presented in Fig. 4. Figure 4 shows that, over broad mountain ranges, both PRISM and WRF predict enhanced precipitation on the upwind (west) side and less precipitation relative to local elevation on the downwind (east) side. This effect varies between models and locations, and local differences between WRF and PRISM exceed 300 mm in places. In particular, mountain ranges with few high-elevation observational stations show relatively large differences between the two products (100–300 mm, Fig. 4a), while ranges with better distributed observational stations show only minimal differences between WRF and PRISM (<25 mm, Fig. 4b).

Results over narrow, isolated mountain ranges, such as the Sangre de Cristo Mountains (Fig. 4c), illustrate another difference between the two products. Over narrow mountain ranges, WRF puts the precipitation maximum downwind (east) of the topographic peak, with regard to prevailing westerlies. This results in WRF estimating less precipitation than PRISM (100 mm) on the windward side and slightly more precipitation in the lee (20–30 mm). Near the peak of the Sangre de Cristo Mountains, PRISM predicts a precipitation maximum just under 500 mm, while WRF predicts precipitation at the peak of only 430–375 mm. The precipitation maximum in WRF occurs in the first grid cell downwind of the peak (450 mm). However, in this mountain range, WRF and PRISM agree quite well where PRISM is incorporating observations (differences less than 10 mm). Further elaboration of the processes controlling precipitation across the Sangre de Cristo mountain range is provided in section 6.

b. Spatial and interannual variability with scale

Figure 5 provides a comparison between WRF 2-km and NARCCAP WRF 50-km precipitation for the period November 2003–May 2004. When the WRF 2-km precipitation is aggregated to the 50-km scale (Fig. 5a), the two datasets are correlated with $r^2 = 0.31$ ($p < 0.01$). It is interesting to note that the aggregated high-resolution model does not yield a strong correlation with a coarse-resolution model. The spread about the one-to-one line is representative of the nonlinear effects of topography and microphysical interactions in a high-resolution model as well as other unresolved processes in the coarse model. This difference between the average of high-resolution grid cells and a single low-resolution grid cell is often termed the gridcell-by-gridcell bias in an RCM (e.g., Wood et al. 2004).

Not surprisingly, the original 50-km NARCCAP precipitation is only weakly correlated with the raw 2-km WRF precipitation ($r^2 = 0.15$, $p < 0.01$), it has a bias of −13 mm, and an rms error of 170 mm (Fig. 5b). When the 50-km dataset is downscaled using the ratio developed with the 2001/02 model runs, the correlation improves dramatically ($r^2 = 0.78$, $p < 0.01$), the magnitude of the bias increases to −41 mm, and the rms error is reduced to 92 mm (Fig. 5c). The increase in the magnitude of the bias can be related to the interannual variations; because the downscaling is only based on a single six-month period, it is fit too strongly to precipitation amounts and patterns specific to that year. As a result, part of the spread around the one-to-one line in Fig. 5c may be a result of the real interannual variability. Without additional model runs, it is impossible to know how much of this spread may be an artifact of the training period.

In addition, the raw 50-km model is particularly bad at capturing the extremes of precipitation in the area (Fig. 5b). The maximum precipitation in the NARCCAP model is just over 600 mm, while the maximum precipitation in the 2-km WRF model is almost 1000 mm. This result has implications for the representation of spatial extremes in RCMs. The downscaled model output adequately captures the highest precipitation totals (Fig. 5c). These plots illustrate the major spatial differences between a coarse and fine resolution model, as well as the interannual variations that may not be well captured by a statistical downscaling. Although interannual differences may be magnified here by the limited duration of the six-month downscaling calibration period, calibration on a multiyear—if not multidecadal—record would likely provide a better starting point.

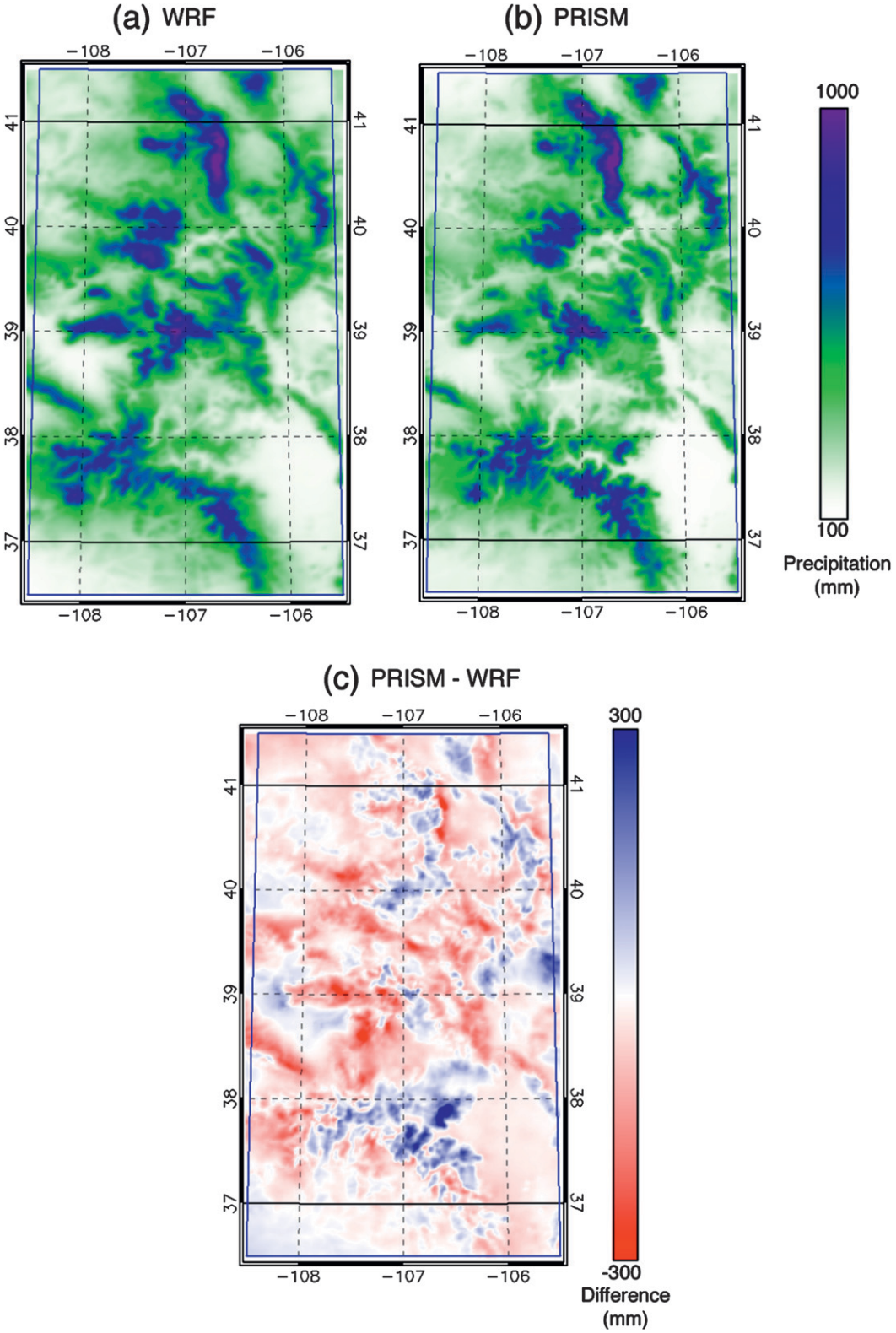

FIG. 3. (a) WRF precipitation climatology (2001/02, 2003/04, 2005/06, 2007/08), (b) PRISM precipitation climatology (1971–2000), and (c) difference between the two. Red (blue) indicates PRISM has less (more) precipitation than WRF. The two datasets are highly correlated in space [(a),(b)]; however, differences of 300 mm are evident in some locations [(c)]. In addition, PRISM often has more precipitation than WRF [blue in (c)] west (upwind) of mountain crests [mountain crests are visible in (a) and (b) as maxima in precipitation].

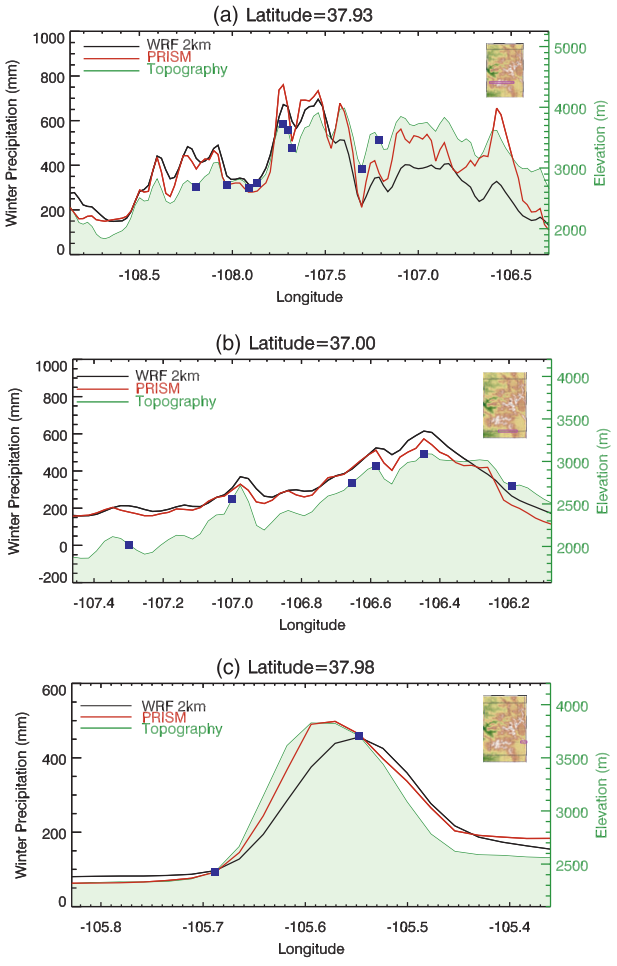

FIG. 4. Cross sections of winter precipitation for PRISM (red) and WRF2-km (black), with topography plotted on the right axis in green; observational stations that are used in the PRISM dataset and are within 10 km north or south of the transect are marked with blue squares at the relevant longitude and elevation. Inset maps show transect locations. In areas with sufficient observations, WRF and PRISM agree on precipitation totals to within 50–100 mm [e.g., from −108.5 to −107.5 in (a), all of (c), most of (b)]. However, on the east side of (a), WRF and PRISM precipitation differ by ~300 mm in places. In all profiles, the maximum precipitation in PRISM is placed on the topographic peak, while the maximum precipitation in WRF is slightly east (downwind) of the topographic peak; this is most evident in (c).

c. *SNOTEL–NARCCAP*

We present a comparison between model-estimated precipitation and SNOTEL-observed precipitation in Figs. 6–8. Figure 6 compares the 2-km WRF simulation with SNOTEL observations accumulated over all four WRF simulations. WRF is strongly correlated with SNOTEL precipitation ($r^2 = 0.73$, $p < 0.01$), the rms error is 92 mm, and the bias is 46 mm. The PRISM dataset, which incorporates these same SNOTEL data, is correlated with the SNOTEL data with $r^2 = 0.9$, $p < 0.01$, a bias of 2.3 mm, and rms error of 49 mm. This provides further validation of the WRF output. It should be noted that the 46-mm bias is comparable to the likely 10% gauge undercatch expected for snow gauges installed at the SNOTEL sites.

The six NARCCAP models are compared with SNOTEL for the period 1988–2000 in Fig. 7 and Table 2, and the downscaled NARCCAP data are compared with SNOTEL in Fig. 8 and Table 2. In all cases, the 50-km models are poorly correlated with the SNOTEL observations with statistically significant correlations ($p < 0.05$) in only one of the six models (CRCM), and two are even negatively correlated (Fig. 7). Additionally, the 50-km model runs were biased low from −262 to −37 mm, and the rms errors ranged from 207 to 301 mm. The low bias is to be expected from models run with a low resolution, and as a result relatively low elevation and topography. In addition, some models have smaller biases (HadRM3), while others have lower RMS errors (RegCM), and others still have higher r^2 values (CRCM). This stresses the importance of understanding the tendencies of any individual model and using either the best model for a specific application or ideally using a model ensemble.

After downscaling the 50-km NARCCAP models to 2 km with the WRF 2-km model output, the downscaled precipitation agreed much more closely with the SNOTEL observations (Fig. 8). The correlation between downscaled precipitation and SNOTEL precipitation ranged from $r^2 = 0.45$ to $r^2 = 0.72$ (in all cases $p < 0.01$). While this is not as strong as the correlation between 2-km WRF and SNOTEL ($r^2 = 0.75$, $p < 0.01$), it is much stronger than the original 50-km data ($r^2 < 0.08$). Furthermore, the rms errors ranged from 87 to 156 mm; this is a little less than half the 50-km rms errors (207–301 mm) and close to the 2-km WRF rms error (98 mm). Finally, the downscaled bias ranged from −123 to −34 mm. The changes in the bias and RMS errors in all cases are statistically significant ($p < 0.01$) based on a paired Student's t test. While a similar improvement might be seen if the PRISM dataset were used to downscale the NARCCAP product, it would not be an independent verification because the PRISM dataset itself incorporates these same SNOTEL data. Figures 6–8 illustrate that such a simple statistical downscaled product can be valid when trained on as little as a single 6-month period.

d. *Future climate*

We present the mean precipitation and differences in precipitation for PGW and the current climate in Fig. 9 and Table 3. The WRF 2-km model precipitation increases by 51 mm ($p < 0.01$) from 369 to 420 mm. In

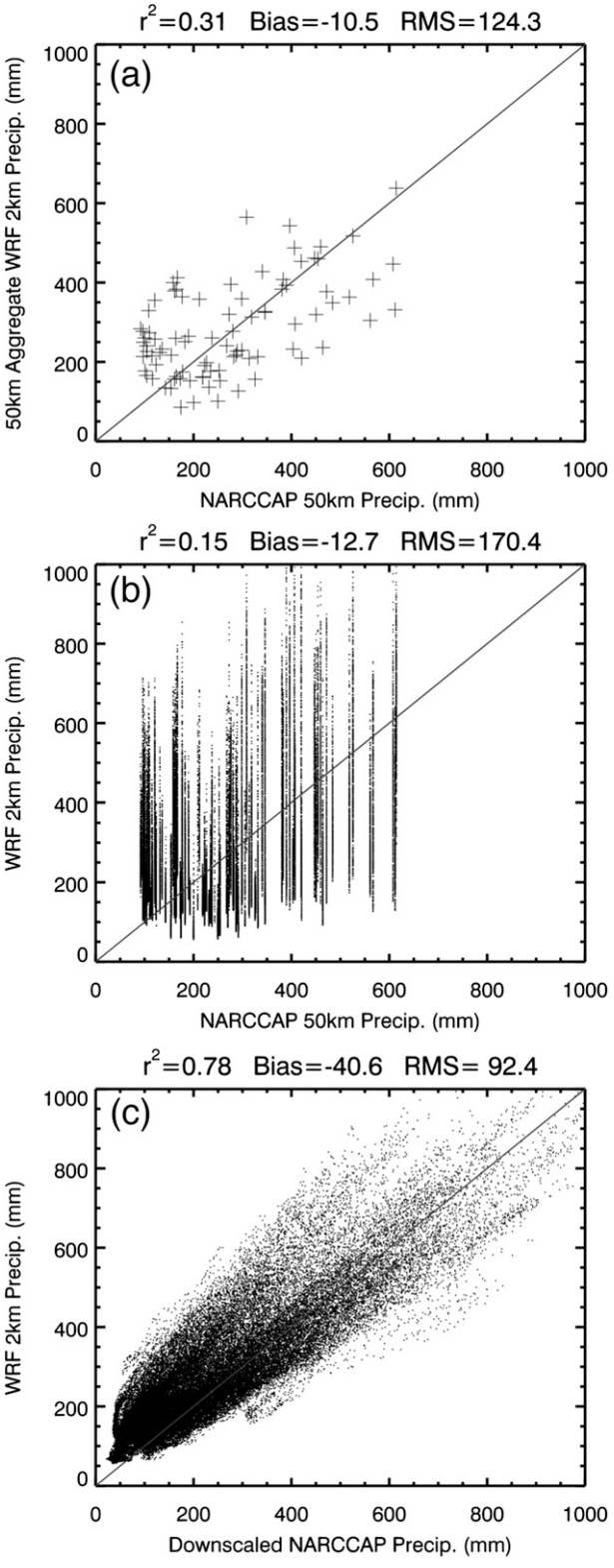

FIG. 5. NARCCAP precipitation vs WRF 2-km precipitation with (a) WRF 2 km aggregated to the NARCCAP 50-km grid, (b) raw NARCCAP precipitation vs WRF 2-km precipitation, and (c) downscaled NARCCAP precipitation vs WRF 2-km precipitation. Each point represents the winter season total (2001/02) for a single grid cell. A one-to-one line is overlaid on all plots for reference. When the 2-km model is aggregated to the 50-km model or the 50-km model is downscaled to the 2-km model, there is some correlation between the two; however, when the raw 50-km model is plotted directly against the 2-km output, there is very little correlation. Surprisingly, the 2-km model aggregated to the 50-km model is only weakly correlated with the 50-km model.

contrast, the WRF 36-km model only increases by 35 mm ($p < 0.01$) from 325 to 360 mm. However, when the downscaled PGW product is compared with the 2-km current climate, the increase in precipitation (43 mm) is somewhat closer to the dynamically predicted increase.

In Fig. 9, the spatial differences between the 36-km WRF, statistically downscaled, and 2-km WRF changes in precipitation are presented. While all products show increases in precipitation over the entire domain, the 36-km simulation shows smaller and less spatially variable changes. The two high-resolution products show larger increases, often focused on the mountain peaks, but the areas with the largest increases in precipitation differ between the two. The 2-km WRF PGW run shows the largest increases in precipitation on the southwestern side of the San Juan Mountains (the southwest corner of the subdomain.) In contrast, the downscaled PGW climate shows the greatest increases along the Colorado Front Range (the northeast corner of the subdomain), although both show large increases in precipitation over the Sangre de Cristo Mountains (southeast edge of the subdomain.)

The spatial distribution of downscaled PGW precipitation is, in some ways, more similar to the current climate WRF 2-km run on which it was trained than it is to the 2-km WRF PGW model that it is supposed to simulate. The rms difference between the downscaled PGW climate and the 2-km WRF current climate (40 mm) is less than the difference between the downscaled PGW climate and the 2-km WRF PGW climate (54 mm). For reference, the difference between the 2-km WRF PGW and the 2-km WRF current climate precipitation is 58 mm. Despite this similarity to the current climate, the mean of the downscaled future climate is more similar to the future climate than it is to the current climate (Table 3). In addition, the histogram of precipitation totals (Fig. 9d) shows the peak of the downscaled precipitation between the WRF current climate and the WRF PGW climate simulations, although the tails of the downscaled histogram are both closer to the WRF PGW distribution. It is also worth noting that the downscaled future climate is highly correlated in space

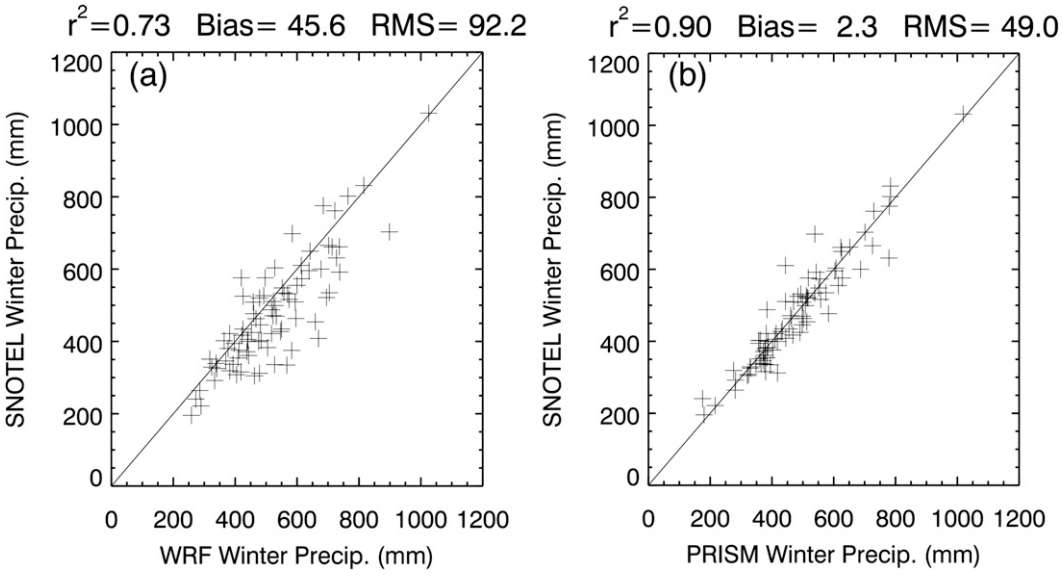

FIG. 6. Average SNOTEL precipitation for the November–May seasons 2001/02, 2003/04, 2005/06, and 2007/08 compared with corresponding data from (a) WRF 2 km and (b) PRISM. The one-to-one line is plotted for reference; each point represents the winter season average precipitation for a single SNOTEL station (model on x axis; observed on y axis). Both models are correlated with the observations; however, the PRISM model is an interpolation that includes these same observations, while the WRF model is an independent dataset.

with both current climate ($r^2 = 0.98$) and future climate ($r^2 = 0.97$). The extremely good correlation in both cases is partially because the downscaling was trained on the WRF 2-km current climate run, and the changes imposed in the PGW simulation are minor relative to the magnitude of the normal spatial variability.

The skill of statistical downscaling at capturing changes in the spatial distribution of precipitation is measured by the correlation between the change predicted by the downscaled PGW simulation and the change predicted by the 2-km WRF PGW simulation ($r^2 = 0.05, p < 0.01$). This extremely low correlation is partially driven by the extremely poor representation of the northeast corner of the domain. If the subdomain is restricted to areas west of -106.5, the correlation increases to $r^2 = 0.17$ ($p < 0.01$). In more localized areas the downscaling performs reasonably well. For example, around the Sangre de Cristo Mountains in the southeastern part of the subdomain, the correlation increases to $r^2 = 0.59$ ($p < 0.01$), and in the San Juan Mountains in the southwestern part of the subdomain the correlation is $r^2 = 0.29$ ($p < 0.01$).

6. Discussion

We first broadly discuss the benefits and problems with the downscaling technique presented in this study. We then provide a specific discussion comparing the dataset used in our technique to the more widely used PRISM-type dataset. Next, we discuss the results from downscaling the NARCCAP data in comparison with the SNOTEL observations, and finally we return to the comparison of statistical and dynamical downscaling in a perturbed climate scenario.

a. Statistical downscaling technique

The downscaling approach used here is simple but sufficient for the experiments presented here and is functionally similar to the BCSD method, without requiring a long time series of data to develop a statistical relationship. However, our approach requires running a high-resolution model with the same boundary conditions as the low-resolution model for every low-resolution model to be used. Because the high-resolution model runs of Rasmussen et al. (2011) were done using NARR observational forcing, it is impossible to run a model that will be used for future simulations as forced by a free-running climate model in a direct comparison mode. To be able to apply the existing high-resolution model runs to future climate simulations, it will be necessary to use an approach such as the BCSD or constructed analogs (Hidalgo et al. 2008), but that is not tested here.

The major drawback to our approach and the BCSD method is related to the distribution of precipitation event intensities. Many studies have focused on more sophisticated methods to accurately reproduce a time series of precipitation, including nonhomogenous hidden Markov models (Hughes and Guttorp 1994), nonhomogenous

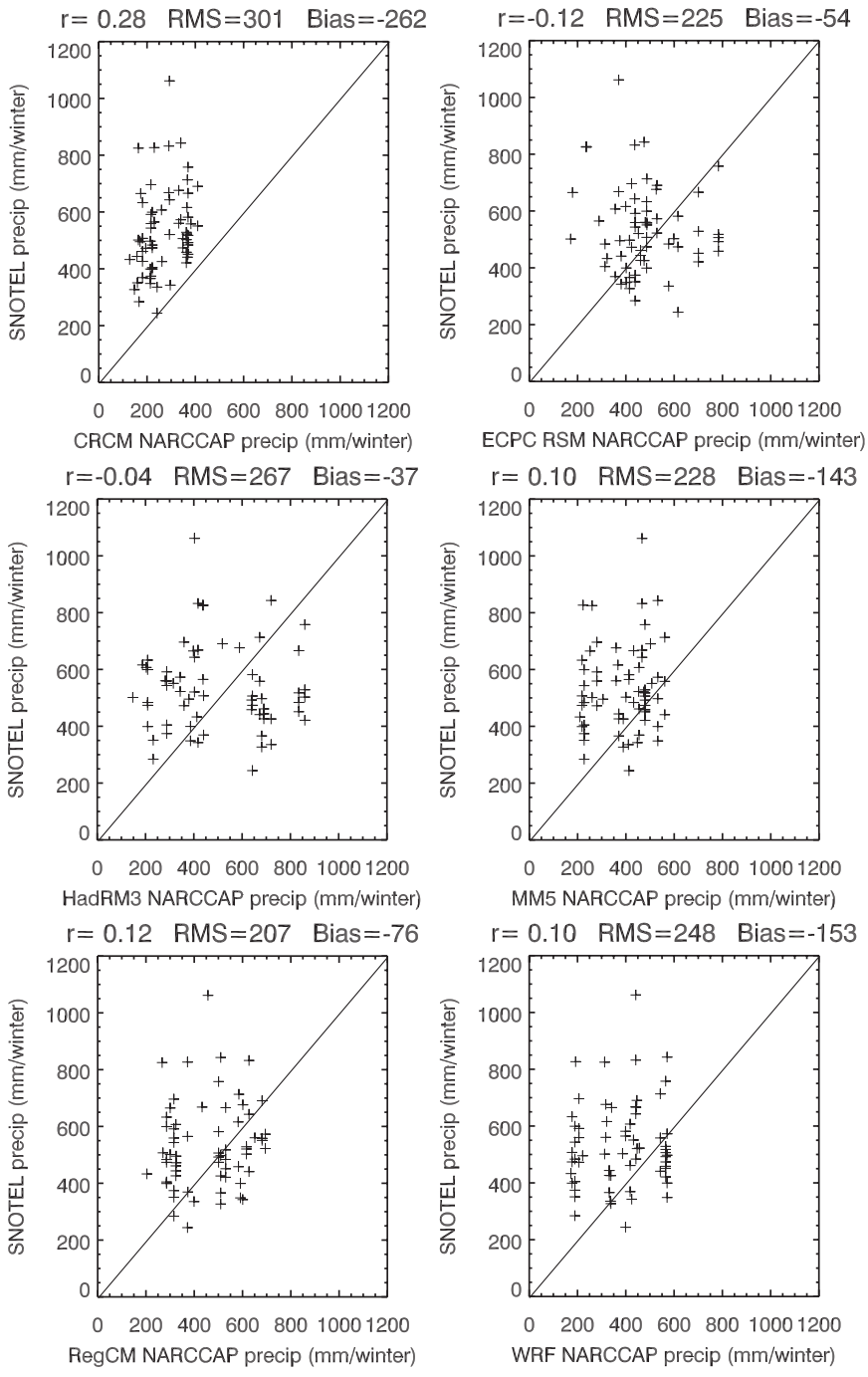

FIG. 7. Comparison between NARCCAP-simulated precipitation and SNOTEL-observed precipitation for the raw 50-km data from the six NARCCAP models. Each point represents the winter season average precipitation (1988–2000) for a single SNOTEL station (model on *x* axis; observed on *y* axis). Correlations are weak to nonexistent and even inverse in two out of six cases.

(nonhidden) Markov models (Vrac et al. 2007), constructed analogs (Hidalgo et al. 2008), and *K*-nearest-neighbor analogs (Gangopadhyay et al. 2005). These statistical downscaling techniques could also be applied to our dataset to produce a better spatiotemporal distribution of precipitation. The Markov models would be slightly more complicated to apply and verify because they generate a new random time series of weather

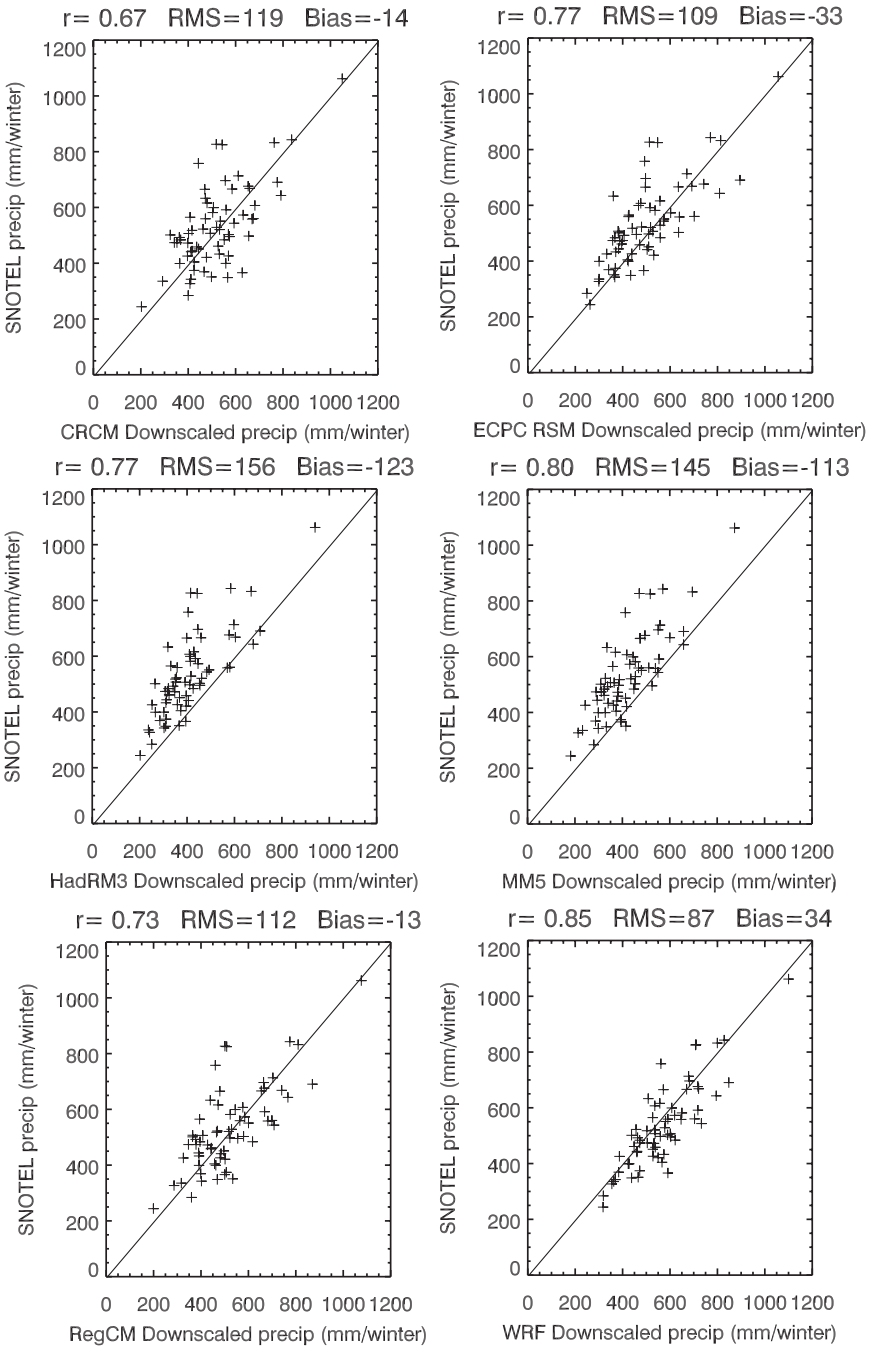

FIG. 8. Comparison between downscaled NARCCAP precipitation and SNOTEL-observed precipitation from the six NARCCAP models. Each point represents the winter season average precipitation (1988–2000) for a single SNOTEL station (model on *x* axis; observed on *y* axis). Correlations, rms errors, and biases are improved over those in Fig. 7 in all cases except the HadRM3 bias.

parameters. The constructed analog and *K*-nearest-neighbor approaches could be applied to our datasets; however, both rely on a long time series of observations to develop adequate statistics. Our current model data, four winter seasons, is probably too short to use these methods.

b. PRISM comparison

Comparisons between WRF and PRISM are difficult because there are few—if any—observation stations that can be used to verify which estimate is correct because

TABLE 2. Differences and correlations in winter season average precipitation (1988–2000) for 50-km NARCCAP (left) and 2-km downscaled NARCCAP (right) relative to SNOTEL. Downscaling substantially improves rms error, bias, and correlation for all cases except the HadRM3 bias.

	50-km model			2-km downscaling		
Model	Rms	Bias	*r*	Rms	Bias	*r*
CRCM	301	−262	0.28	119	−14	0.67
ECPC RSM	225	−54	−0.12	109	−33	0.77
HadRM3	267	−37	−0.04	156	−123	0.77
MM5	228	−143	0.10	145	−113	0.80
RegCM	207	−76	0.12	112	−13	0.73
WRFP	248	−153	0.10	87	34	0.85

data from most stations are incorporated into the PRISM dataset itself. However, it is clear from the simple comparison presented here that the two datasets provide a very similar overall map of precipitation, and the differences are highly localized. The similarity in the two estimates is encouraging from a regional climate modeling perspective as it demonstrates a level of model simulation skill that has heretofore been lacking. Additionally, the fact that these differences between the 2-km WRF and PRISM commonly occur in locations that do not have observational data raises additional issues, such as the possibility that that a highly resolved regional climate model may be capable of providing better estimates of the spatial distribution and intensity of winter precipitation in some regions than an empirically tuned regression model, such as PRISM. However, there is no clear answer to this at present.

Exploring some of these differences further, we revisit the two examples presented in which WRF differs substantially from PRISM in an attempt to provide a physical justification for why WRF may be producing a more credible result. In Fig. 4a, both WRF and PRISM capture the relatively high precipitation on the dominantly upwind (western) slope, but WRF predicts a substantial decrease further downwind (eastern San Juan Mountains). When the larger picture of both WRF and PRISM are examined, both models agree that the mountain range to the north, the Sawatch Range, exhibits a substantial decrease in precipitation relative to local elevation and that the mountains to the south, the southern San Juan Mountains, exhibit greater precipitation relative to local elevation. Because of their proximity, it is the southern mountain observations that are presumably influencing the PRISM regression over the eastern San Juan Mountains; however, the southern mountains are not downwind of another mountain range in the way that the eastern San Juan Mountains, and the Sawatch Range, are. Unfortunately, because the Sawatch Range is farther away, PRISM apparently does not provide as strong a weight to stations in this region when predicting precipitation over the eastern San Juan Mountains.

This theory can be tested using a recent addition to the SNOTEL network in the eastern San Juan Mountains at Moon Pass (37.96°N, −106.56°W; 3400 m). This station was installed in October 2008, thus is it not used in the PRISM climatology. At this site, the PRISM climatology estimates a winter precipitation total of 600 mm, while WRF estimates a winter precipitation total of 280 mm. The SNOTEL winter precipitation from 2008 to 2011 ranges from 193 to 292 mm with an average of 232 mm. This is 48 mm lower than WRF, approximately the same as the WRF bias across all the SNOTEL sites; however, it is dramatically below the PRISM climatology for this region (bias of 368 mm). This shows one area where a statistical downscaling based on PRISM would do very poorly, even though this region has one of the best observing networks in the world.

Similarly, the differences between WRF and PRISM in the Sangre de Cristo Mountains (Fig. 4c) make sense physically, but we have no station observations to verify if WRF is correct. Over these mountains, one would expect a strong updraft on the upwind side (Medina et al. 2005; Garvert et al. 2007; Colle et al. 2000) leading to an increase in precipitation production but a decrease in snowfall on the ground. This precipitation would be advected downwind until it encountered the strong downdraft on the east side of the mountain range, where it would fall to the ground. This occurs because snow terminal velocity is typically around 1 m s^{-1}, while strong updrafts are 1–2 m s^{-1} in this region (cf. Rasmussen et al. 2011). The strong updraft leads to substantial condensate loading, but the aerodynamics of snowflakes may prevent them from reaching the ground. While this phenomenon has been documented in previous studies (Medina et al. 2005; Garvert et al. 2007; Colle et al. 2000), measurements of precipitation close to mountain peaks have been sparse and only hinted at this phenomena, and have not clearly confirmed it (Garvert et al. 2005; Medina et al. 2005). Based on these other studies, we hypothesize WRF is correct here, but it will take further measurements in the field to confirm or reject this. The Sangre de Cristo Mountains are a relatively simple, linear mountain ridge that would make an ideal location to study the local-scale (1–10 km) distribution of precipitation resulting from interactions between dynamics and microphysics.

c. *Spatial and interannual variability*

The correlation between downscaled NARCCAP and 2-km WRF output (Fig. 5) illustrates the improvements possible with statistical downscaling and the limitations in estimating changes in interannual spatial distributions. Because precipitation is strongly tied to topography, it is

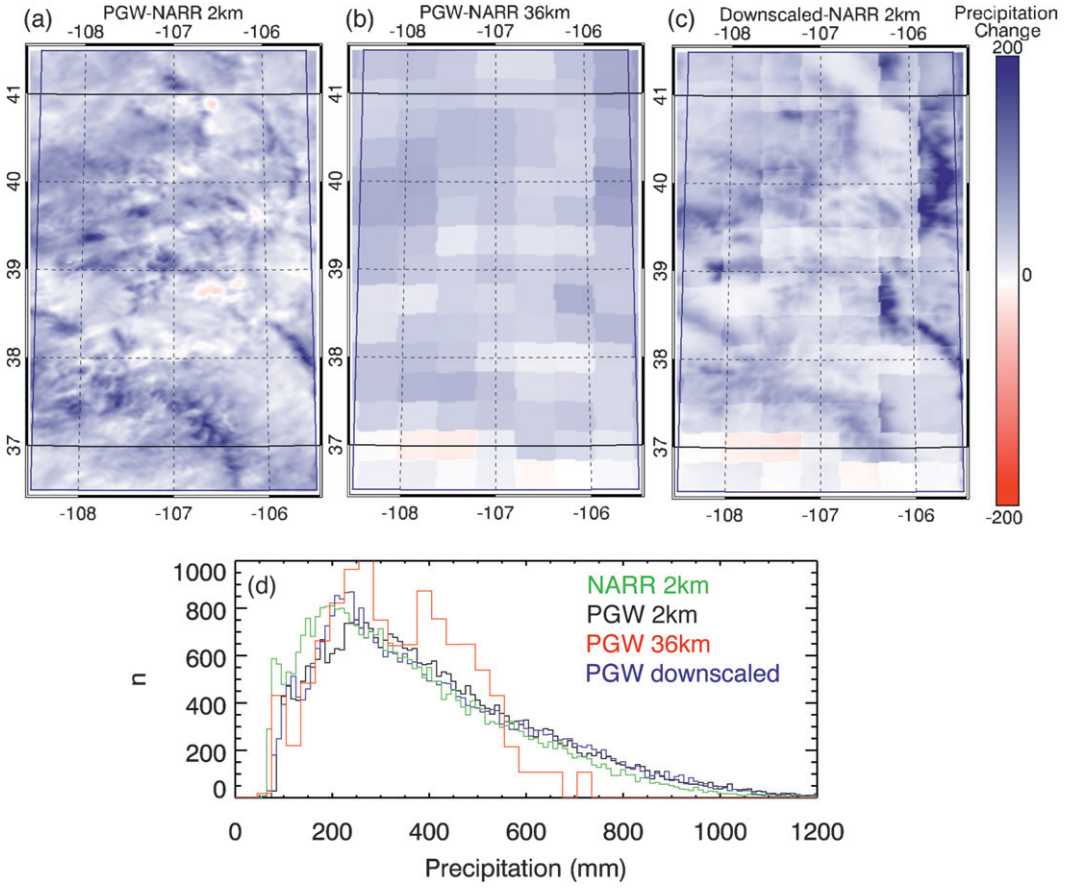

FIG. 9. Difference maps between current and PGW precipitation for (a) WRF 2 km, (b) WRF 36 km, and (c) downscaled WRF 36 km. Blue scale represents increases in precipitation in a future climate up to 200 mm; red scale represents decreases in precipitation in the future down to −200 mm. (d) Histograms of precipitation for the WRF 2-km NARR forced simulation used to train the downscaling, the PGW WRF 2 km, the PGW WRF 36 km, and the PGW downscaled simulations. All models predict an increase in future precipitation, but the pattern of change in the 2-km model differs from both the 36-km model and the downscaled product. Locally (e.g., the Sangre de Cristo and the San Juan Mountains), the downscaled change is correlated with the 2-km change; however, over the entire region, areas of poor performance (e.g., the front range) substantially decrease the total skill of the downscaled product at predicting changes.

not surprising that applying a simple downscaling developed in one year to a second year would work reasonably well. However, the fact that the downscaled product does not match the high-resolution model exactly shows that there is substantial interannual variation in spatial distribution of precipitation that is not properly depicted at the coarse 50-km scale. The r^2 of 0.78 implies that nearly 20% of the interannual variation is not predicted by the simple statistical downscaling presented here, but it is worth noting that the r^2 between the two 2-km WRF runs for these two years is only 0.75. Thus, there may be a small improvement (~3%) gained from the spatial variability in the NARCCAP model. When one looks at all pairs of 2-km WRF runs, the interannual r^2 values range from 0.73 to 0.90 with a mean of 0.80, suggesting that the two years compared here are relatively

TABLE 3. Mean and change in subdomain precipitation for each current (NARR) and future (PGW) high- (2 km) and low- (36 km) resolution WRF models and the downscaled 36-km PGW simulation. The 36-km PGW simulation underpredicts the change in precipitation relative to the 2-km PGW simulations; the downscaled PGW comes closer to the 2-km PGW simulations.

Model	2-km NARR	2-km PGW	36-km NARR	36-km PGW	Downscaled PGW
Mean precipitation (mm)	369	420	325	360	412
Change in PGW precipitation (mm)	N/A	51	n/a	35	43

different from one another and that statistical downscaling built on two or more years would better predict interannual variability on average than the current r^2 of 0.78 would suggest.

Further improvements to downscaled interannual predictions could be made by increasing the sophistication of the downscaling method. Winter precipitation in Colorado is dominated by two primary synoptic features. The first is driven by dominantly westerly winds and occurs throughout the winter, while the second is dominated by northeasterly winds and commonly occurs in the spring. The second, referred to as an ''upslope'' storm in the greater Rocky Mountain Front Range region, is often associated with a high pressure system to the north or a low pressure system to the south driving northly and easterly winds over Colorado; see Rasmussen et al. (1995) for more detailed descriptions. In particular, these storms represent cases in which topography can be a controlling factor not directly tied to orographic-lifting-enhanced precipitation. Rasmussen et al. describe a 1990 upslope storm in which regional snowfall was greatly enhanced because the movement of the storm system was retarded by elevated terrain to the south and west. Some of the interannual variability is simply due to the varying influences of these two important synoptic features and, if a statistical downscaling can properly treat large-scale features such as these independently, then substantial improvements could be made. Some work on this type of downscaling has been done by Vrac et al. (2007), and a simple measure of synoptic conditions from slightly above ridgetop height could improve downscaling predictions substantially.

To analyze the potential to improve statistical downscaling based on wind direction, we looked at WRF simulations of accumulated precipitation over the Sangre de Cristo Mountains, binned into four 650-mb wind directions (−45°–45°, 45°–135°, 135°–225°, and 225°–315°). Figure 10 shows significant differences in the spatial distribution of precipitation as a function of wind direction. Total precipitation reaches a local maximum just east of the topographic peak. When winds are from the north, the precipitation maximum is shifted farther east, as might be expected in a typical front range upslope storm event. When winds are from the west, the precipitation maximum is closer to the topographic peak. More of the precipitation in this region comes when winds are from the south than from any other direction, and the distribution is close to that of the total precipitation, though the precipitation maximum is slightly broader. Finally, relatively little precipitation falls when winds are from the east; however, what does fall is more homogenously distributed than for any other wind direction. This behavior suggests that at least some of the spatial variance

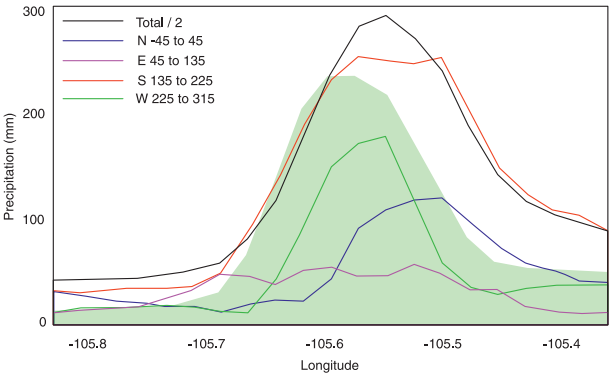

FIG. 10. WRF 2-km accumulated precipitation across the Sangre de Cristo Mountains binned into four 650-mb wind directions, with total accumulation scaled by 0.5 in black; topography is shaded in light green. The variation in precipitation patterns as a function of wind direction could be used to improve statistical downscaling techniques.

in precipitation in this region can be explained by synoptic wind conditions and therefore serve as a useful predictor in a downscaling model (e.g., Guan et al. 2005).

Similar improvements to statistical downscaling methods could be made by taking into account regional parameters, such as the air temperature and humidity. Temperature and moisture will affect not only the total regional precipitation but also the spatial distribution as a result of microphysical feedbacks between atmospheric dynamics and hydrometeors [e.g., cooling (warming) caused by melting (freezing) hydrometeors]. While a high-resolution simulation can explicitly account for these microphysical feedbacks, most statistical downscaling techniques do not include them.

d. Climate change scenario

This study presents a unique opportunity to assess the ability of statistical downscaling methods, which were trained in the current climate, to operate properly in a modified climate. While the generally good correlation between downscaled PGW precipitation and 2-km WRF PGW precipitation ($r^2 = 0.97$) suggests that statistical downscaling will work exceptionally well in a changing climate, it is important to point out that the downscaling did exceedingly poorly at predicting the spatial variation of the changes in precipitation ($r^2 = 0.05$, locally 0.17–0.59) that were predicted by a dynamically downscaled model. In areas in which the primary change in precipitation predicted by the 2-km WRF PGW model is a simple percentage of the current precipitation (e.g., the Sangre de Cristo Mountains), the statistical downscaling performs well in comparison to the 2-km WRF PGW model ($r^2 = 0.59$). In contrast, regions in which the regional climate model predicts substantial changes in flow

patterns that are different from the high-resolution model (e.g., the Colorado Front Range), the statistical downscaling does extremely poorly relative to the dynamical downscaling. While it is not clear which model is necessarily ''correct,'' the difference between the two should provide a note of caution in the application of statistical downscaling to climate change scenarios.

The broad spatial pattern of precipitation in the Colorado Rockies is unlikely to change substantially in a future climate scenario (e.g., precipitation will remain concentrated on mountain peaks); thus, a simple statistical downscaling may be able to capture most of the real spatial variability (e.g., $r^2 = 0.97$ here). However, statistical downscaling may not capture the finer-scale spatial changes in precipitation resulting from subtle changes in storm patterns and the microphysical response to a warming climate. This microphysical response is expected to make a significant contribution to increases in future precipitation (Rasmussen et al. 2011) and, as a result, even the statistically downscaled future climate does not capture the full change in precipitation expected from high-resolution dynamical downscaling (43 vs 51 mm). Thus, while statistical downscaling does a reasonable job at predicting the overall spatial distribution of precipitation in any climate, it does not capture some of the changes in precipitation that are important for a water-use planning scenario. For example, the downscaled simulation predicts a large change in precipitation for the Colorado Front Range that is not predicted by the 2-km WRF model, while the downscaled model does not predict the same magnitude increase in precipitation in the San Juan Mountains that the 2-km WRF model does. Even smaller-scale differences than this can be important because a shift in peak snowfall of only a few kilometers can make substantially more water available in one river basin and less in another. This illustrates some of the uncertainties in predicting future water resource availability using statistical downscaling.

With the availability of the high-resolution dynamically downscaled future climate predictions, it may be possible to further improve statistical downscaling beyond the improvements described in section 6c by assessing the changes in the dynamic model in a warmer climate. In particular, future changes in precipitation resulting from increases in temperature and humidity could be quantified using this high-resolution model, as could the expected precipitation change in a warmer climate as a function of wind speed and direction or moisture flux convergence fields. By applying this information to a regional climate model, it is expected that a high-resolution (~2 km grid) hourly or daily climate time series could be developed that would have a greater fidelity to the underlying physics than a simple downscaling based on a PRISM-type model. Work on this is ongoing.

7. Conclusions

Using a dynamical model to develop a statistical downscaling methodology may capture important features related to but not directly correlated with topography, thereby potentially improving upon PRISM-type datasets that have been used as the basis for previous statistical downscaling methods. While we have not verified that the dynamically downscaled product is better than PRISM in all areas, we have shown locations where there are significant differences between the two. These differences can be reasonably explained by physical processes that are included in the dynamical downscaling, but not in the PRISM dataset. In the case of the eastern San Juan Mountains, recent observations have shown that the WRF estimate of 280 mm of winter precipitation is much closer to the observed 230 mm than the PRISM estimate of 600 mm. As such, any statistical downscaling technique that uses a PRISM-type interpolated dataset (e.g., Maurer et al. 2002; Daly et al. 2008; Thornton et al. 1997) for a basemap will likely exhibit important differences from a dynamical downscaling in these locations.

Applying this statistical downscaling to six coarse-resolution regional climate models for a 12-yr period greatly improves the comparison between these models and observations. The average rms error across models goes from 246 to 121 mm, average bias goes from −120 to −43 mm, and average r^2 goes from 0.02 to 0.58. Future work will use a more sophisticated downscaling technique to apply the dynamically downscaled information to all of the NARCCAP simulations, thus providing a high-spatial-resolution (2 km) and high-temporal-resolution (3 h) statistically downscaled dataset that can be used to drive hydrologic models. In addition, this dataset would include an ensemble of regional and global climate models, thus providing some sense of the variability expected from current modeling frameworks. Such a downscaling will require work to ensure that statistical features of the climate system (e.g., extremes) are reasonably reproduced.

Overall, while individual models fair better or worse, the net gain in correlation to SNOTEL sites and the decrease in the rms error and bias show that statistical downscaling based on a dynamic model works reasonably well. While it is not completely clear how this downscaling compares to a product such as PRISM under a purely independent validation, it is noteworthy that the downscaling approaches presented here required no observations of precipitation on the ground. As such, it may be considered especially useful for regions with few local precipitation observations (e.g., the Himalayas or the Andes), or for future climate scenarios in which no measurements can be made a priori.

Developing this downscaling in the current climate and applying it to an imposed climate change scenario makes the low-resolution model better match a dynamically downscaled model on average (decreased bias and rms error), but it does not capture finescale features that arise from the interactions of a more moist atmosphere with the underlying topography. As a result, changes in the spatial distribution of precipitation predicted by the dynamically downscaled model are only captured by the statistical downscaling with an r^2 of 0.05–0.19, though locally this can be as high as 0.6. This result is expected to hold true for most—if not all—current statistical downscaling techniques, and it should be considered in any assessment of future precipitation that is reliant on statistical downscaling methods. Statistical downscaling will improve the mean water availability and the general spatial distribution of that precipitation, but it will not significantly improve our understanding of the changes in the spatial distribution of precipitation that will occur without substantial improvements to statistical downscaling techniques.

Dynamical downscaling methods, such as those presented in Rasmussen et al. (2011), offer a method to identify expected changes in the spatial distribution of precipitation and mechanisms behind the precipitation changes as well, and future research will address ways to combine these two approaches. For example, it may be that specific slopes will see increases in precipitation that are a function of the large-scale wind field, temperature, humidity, or 500-mb height, and changes in these variables might be reasonably predicted by a low-resolution simulation. Thus, by combining these factors with a statistical downscaling method, better estimates of the spatial distribution of future precipitation can be achieved. Furthermore, because dynamical downscaling outputs all of the relevant forcing variables for a land surface or hydrologic model (e.g., wind, pressure, humidity, radiation), it is expected that it will be possible to develop statistical downscaling functions for other such variables, not just precipitation.

Acknowledgments. The NCAR Water Systems Program provided the WRF 2-km simulations. We thank the North American Regional Climate Change Assessment Program (NARCCAP) for providing the data used in this paper. NARCCAP is funded by the National Science Foundation (NSF), the U.S. Department of Energy (DOE), the National Oceanic and Atmospheric Administration (NOAA), and the U.S. Environmental Protection Agency (EPA) Office of Research and Development. SNOTEL data were provided by the National Resources Conservation Service (NRCS) branch of the U.S. Department of Agriculture (USDA). The PRISM precipitation climatology was provided by the PRISM Climate Group, Oregon State University (www.prismclimate.org). In addition, we thank the editors and the two anonymous reviewers for providing valuable feedback.

REFERENCES

Chen, F., and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.,* **129,** 569–585.

Colle, B., C. Mass, and K. Westrick, 2000: MM5 precipitation verification over the Pacific Northwest during the 1997–99 cool seasons. *Wea. Forecasting,* **15,** 730–744.

Collins, W. D., and Coauthors, 2004: Description of the NCAR Community Atmosphere Model (CAM 3.0). NCAR Tech. Note NCAR/TN-464+STR, 214 pp.

——, and Coauthors, 2006: Radiative forcing by well-mixed greenhouse gases: Estimates from climate models in the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4). *J. Geophys. Res.,* **111,** D14317, doi:10.1029/2005JD006713.

Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical–topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.,* **33,** 140–158.

——, M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, 2008: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *Int. J. Climatol.,* **28,** 2031–2064.

Gangopadhyay, S., M. Clark, and B. Rajagopalan, 2005: Statistical downscaling using *K*-nearest neighbors. *Water Resour. Res.,* **41,** W02024, doi:10.1029/2004WR003444.

Garvert, M. F., C. P. Woods, B. A. Colle, C. F. Mass, P. V. Hobbs, M. T. Stoelinga, and J. B. Wolfe, 2005: The 13–14 December 2001 IMPROVE-2 event. Part II: Comparisons of MM5 model simulations of clouds and precipitation with observations. *J. Atmos. Sci.,* **62,** 3520–3534.

——, B. Smull, and C. Mass, 2007: Multiscale mountain waves influencing a major orographic precipitation event. *J. Atmos. Sci.,* **64,** 711–737.

Guan, H., J. Wilson, and O. Makhnin, 2005: Geostatistical mapping of mountain precipitation incorporating autosearched effects of terrain and climatic characteristics. *J. Hydrometeor.,* **6,** 1018–1031.

Hara, M., T. Yoshikane, H. Kawase, and F. Kimura, 2008: Estimation of the impact of global warming on snow depth in Japan by the pseudo-global-warming method. *Hydrol. Res. Lett.,* **2,** 61–64.

Hayhoe, K., and Coauthors, 2004: Emissions pathways, climate change, and impacts on California. *Proc. Natl. Acad. Sci. USA,* **101,** 12 422–12 427.

Hewitson, B., and R. Crane, 1996: Climate downscaling: Techniques and application. *Climate Res.,* **7,** 85–95.

Hewitt, C. D., and D. J. Griggs, 2004: Ensembles-based predictions of climate changes and their impacts (ENSEMBLES). *Eos, Trans. Amer. Geophys. Union,* **85,** doi:10.1029/2004EO520005.

Hidalgo, H. G., M. D. Dettinger, and D. R. Cayan, 2008: Downscaling with constructed analogues: Daily precipitation and temperature fields over the United States. PIER Final Project Rep., California Energy Commission Rep. CEC-500-2007-123, 62 pp.

Hughes, J., and P. Guttorp, 1994: A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resour. Res.,* **30,** 1535–1546.

Hurrell, J. W., G. J. Holland, and W. G. Large, 2008: The nested regional climate model: An approach toward prediction across scales. *Eos, Trans. Amer. Geophys. Union,* **89** (Fall Meeting Suppl.), Abstract PA13C-1350.

Ikeda, K., and Coauthors, 2010: Simulation of seasonal snowfall over Colorado. *Atmos. Res.,* **97,** 462–477.

Kain, J. S., 2004: The Kain–Fritsch convective parameterization: An update. *J. Appl. Meteor.,* **43,** 170–181.

Kawase, H., T. Yoshikane, M. Hara, B. Ailikun, F. Kimura, and T. Yasunari, 2008: Downscaling of the climatic change in the mei-yu rainband in East Asia by a pseudo climate simulation method. *SOLA,* **4,** 73–76.

Klemp, J. B., W. C. Skamarock, and J. Dudhia, 2007: Conservative split-explicit time integration methods for the compressible nonhydrostatic equations. *Mon. Wea. Rev.,* **135,** 2897–2913.

Liu, C., K. Ikeda, G. Thompson, R. Rasmussen, and J. Dudhia, 2011: High-resolution simulations of wintertime precipitation in the Colorado Headwaters region: Sensitivity to physics parameterizations. *Mon. Wea. Rev.,* **139,** 3533–3553.

Maurer, E. P., A. Wood, J. Adam, D. Lettenmaier, and B. Nijssen, 2002: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States. *J. Climate,* **15,** 3237–3251.

——, L. Brekke, T. Pruitt, and P. B. Duffy, 2007: Fine-resolution climate projections enhance regional climate change impact studies. *Eos, Trans. Amer. Geophys. Union,* **88,** doi:10.1029/2007EO470006.

Mearns, L. O., W. Gutowski, R. Jones, R. Leung, S. McGinnis, A. Nunes, and Y. Qian, 2009: A regional climate change assessment program for North America. *Eos, Trans. Amer. Geophys. Union,* **90,** doi:10.1029/2009EO360002.

Medina, S., B. Smull, R. Houze, and M. Steiner, 2005: Cross-barrier flow during orographic precipitation events: Results from MAP and IMPROVE. *J. Atmos. Sci.,* **62,** 3580–3598.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.,* **87,** 343–360.

Rasmussen, R. M., B. C. Bernstein, M. Murakami, G. Stossmeister, J. Reisner, and B. Stankov, 1995: The 1990 Valentine's Day arctic outbreak. Part I: Mesoscale and microscale structure and evolution of a Colorado Front Range shallow upslope cloud. *J. Appl. Meteor.,* **34,** 1481–1511.

——, and Coauthors, 2011: High-resolution coupled climate runoff simulations of seasonal snowfall over Colorado: A process study of current and warmer climate. *J. Climate,* **24,** 3015–3048.

Schar, C., C. Frei, D. Luthi, and H. Davies, 1996: Surrogate climate-change scenarios for regional climate models. *Geophys. Res. Lett.,* **23,** 669–672.

Serreze, M., M. Clark, R. Armstrong, D. McGinnis, and R. Pulwarty, 1999: Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resour. Res.,* **35,** 2145–2160.

Skamarock, W. C., and J. B. Klemp, 2008: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.,* **227,** 3465–3485.

Stern, P. C., and W. E. Easterling, Eds., 1999: *Making Climate Forecasts Matter*. National Academy Press, 175 pp.

Thompson, G., R. M. Rasmussen, and K. Manning, 2004: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon. Wea. Rev.,* **132,** 519–542.

——, P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.,* **136,** 5095–5115.

Thornton, P., S. Running, and M. White, 1997: Generating surfaces of daily meteorological variables over large regions of complex terrain. *J. Hydrol.,* **190,** 214–251.

Vrac, M., M. Stein, and K. Hayhoe, 2007: Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing. *Climate Res.,* **34,** 169–184.

Wang, S.-Y., R. R. Gillies, E. S. Takle, and W. J. Gutowski Jr., 2009: Evaluation of precipitation in the intermountain region as simulated by the NARCCAP regional climate models. *Geophys. Res. Lett.,* **36,** L11704, doi:10.1029/2009GL037930.

Wilby, R. L., T. M. L. Wigley, D. Conway, P. D. Jones, B. C. Hewitson, J. Main, and D. S. Wilks, 1998: Statistical downscaling of general circulation model output: A comparison of methods. *Water Resour. Res.,* **34,** 2995–3008.

Wood, A., L. Leung, V. Sridhar, and D. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change,* **62,** 189–216.

Yang, D., B. Goodison, S. Ishida, and C. Benson, 1998: Adjustment of daily precipitation data at 10 climate stations in Alaska: Application of World Meteorological Organization intercomparison results. *Water Resour. Res.,* **34,** 241–256.