Extreme Precipitation and Temperature over the U.S. Pacific Northwest: A Comparison between Observations, Reanalysis Data, and Regional Models*

VALÉRIE DULIÈRE

Climate Impacts Group, Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, Washington, and Management Unit of the North Sea Mathematical Models, Department VI of the Royal Belgian Institute of Natural Sciences, Brussels, Belgium

YONGXIN ZHANG

National Center for Atmospheric Research, Boulder, Colorado

ERIC P. SALATHÉ JR.

Science and Technology Program, University of Washington—Bothell, Bothell, and Climate Impacts Group, Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, Washington

(Manuscript received 27 April 2009, in final form 3 December 2010)

ABSTRACT

Extreme precipitation and temperature indices in reanalysis data and regional climate models are compared to station observations. The regional models represent most indices of extreme temperature well. For extreme precipitation, finer grid spacing considerably improves the match to observations. Three regional models, the Weather Research and Forecasting (WRF) at 12- and 36-km grid spacing and the Hadley Centre Regional Model (HadRM) at 25-km grid spacing, are forced with global reanalysis fields over the U.S. Pacific Northwest during 2003–07. The reanalysis data represent the timing of rain-bearing storms over the Pacific Northwest well; however, the reanalysis has the worst performance at simulating both extreme precipitation indices and extreme temperature indices when compared to the WRF and HadRM simulations. These results suggest that the reanalysis data and, by extension, global climate model simulations are not sufficient for examining local extreme precipitations and temperatures owing to their coarse resolutions. Nevertheless, the large-scale forcing is adequately represented by the reanalysis so that regional models may simulate the terrain interactions and mesoscale processes that generate the observed local extremes and frequencies of extreme temperature and precipitation.

1. Introduction

Extreme weather events such as heat waves, floods, droughts, or storms can lead to severe societal and economical impacts. Over recent decades, the cost of extreme events has increased dramatically (United Nations Environment Programme 2002). Global climate models simulate a link between a warmer climate and changes in extreme weather events, and extreme events are expected to change in frequency or intensity in a warming climate (Solomon et al. 2007; Tebaldi et al. 2006). More recently, there have been some observational evidence of this connection between a warmer climate and extreme events. For instance, Allan and Soden (2008) demonstrated a direct link between a warmer climate and an amplification of precipitation extremes in tropical areas using satellite observations. Thus, simulating the effects of climate on extreme events at the local scale is of great importance for assessing the impacts of projected climate change.

Global models are powerful tools to investigate climate change on large scales. However, such models do not represent local terrain and mesoscale weather systems well owing to their coarse horizontal resolution (~150–300 km). Therefore, they face difficulties in adequately

* National Oceanic and Atmospheric Administration Contribution Number 1752.

Corresponding author address: Valérie Dulière, University of Washington, Box 355672, Seattle, WA 98195-5672.
E-mail: v.duliere@mumm.ac.be

DOI: 10.1175/2010JCLI3224.1

© 2011 American Meteorological Society

resolving the interactions of large-scale weather systems with local terrain and mesoscale processes that are important for causing localized extreme weather events, which have the greatest impacts. Note that the U.S. Pacific Northwest is especially challenging for global models since this region is characterized by complex terrain that includes mountainous ranges and land–sea contrasts (see Fig. 1).

To capture the finescale features such as orographic precipitation, land–sea breeze, rain shadows, and wind storms, regional climate models (RCMs) with a more realistic representation of the complex terrain and heterogeneous land surfaces are needed (Mass et al. 2002; Leung et al. 2003a,b). High-resolution simulations with a fifth-generation Pennsylvania State University–National Center for Atmospheric Research Mesoscale Model (MM5)-based regional climate simulation at 15-km grid spacing show markedly different trends in temperature and precipitation over the Pacific Northwest compared to the driving global models, presumably due to mesoscale processes not being resolved at coarse resolution (Salathé et al. 2008).

To evaluate regional climate model performance over the Pacific Northwest, Zhang et al. (2009) analyzed simulations from two limited-area coupled land–atmosphere models, the Weather Research and Forecasting (WRF) model and the Hadley Centre Regional Model (HadRM), forced at their lateral boundaries with data from the National Centers for Environmental Prediction (NCEP)–NCAR reanalysis 2 (R2 hereafter). Zhang et al. noted improvement of the regional climate models performance over the large-scale driving data in simulating the observed climatology of precipitation and temperature. However, relatively few studies (e.g., Fowler et al. 2005; Fowler and Ekström 2009; May 2008; Räisänen et al. 2004) have been dedicated to regional model performance in terms of extreme weather events despite the tremendous interest in such phenomena. In this study, we build on the analysis in Zhang et al. (2009) to focus on extreme precipitation and temperature in the two regional climate simulations. Here we compare the R2 and the regional model simulations with observed indices of extreme temperature and precipitation. This model evaluation is a necessary step toward confident interpretation of the projections of future changes in extreme weather events at the local scale, which we take to be tens of square kilometers.

The goal of this paper is to evaluate whether the regional climate models, when forced by reanalysis, can reproduce local daily temperature and precipitation statistics observed at a station. The motivation for this approach is that, for climate impacts applications, the station-level observations provide the closest representation of extreme

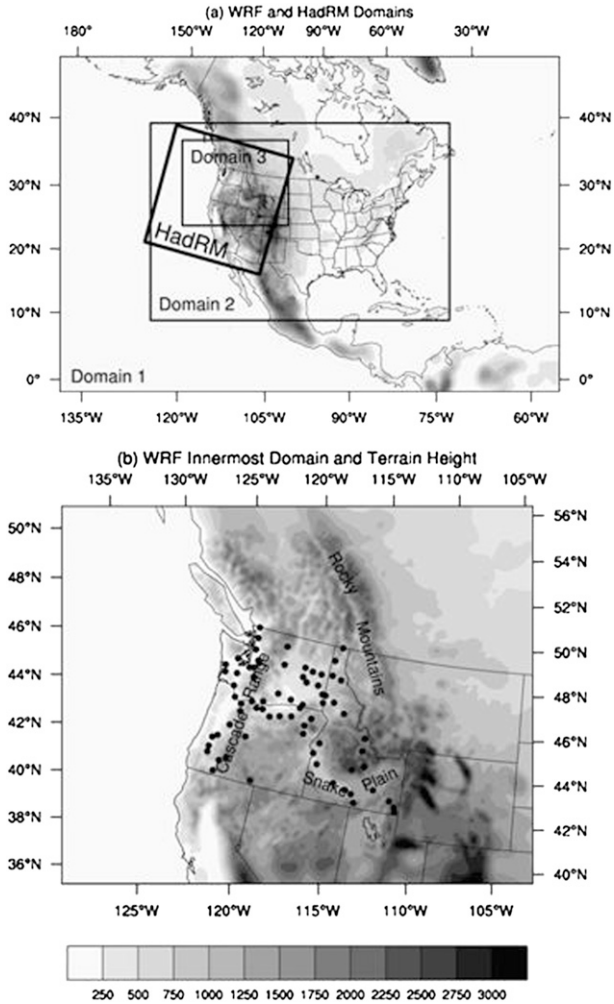

FIG. 1. (a) WRF model domains with two nests (Domain 1, 2, 3) and HadRM domain and (b) WRF innermost domain (Domain 3) and terrain height (m). Shadings represent terrain height (m) for the corresponding WRF domain. Grid spacing for each domain is WRF Domain 1: 108 km, WRF Domain 2: 36 km, WRF Domain 3: 12 km, and HadRM Domain: 25 km. HCN stations in the states of Washington, Oregon, and Idaho are represented by filled black circles in (b).

events of interest. However, since model gridcell values represent the average over the area covered by the grid cell (1296 to 144 km^2 in this study), comparing a gridcell value to a point observation is potentially inconsistent due to heterogeneity within a grid cell. Similar issues have been encountered in other studies comparing model output to station data, for example, in verifying weather forecasts (Mass et al. 2002) or comparing Atmospheric Radiation Measurement project data to models (Hinkelman et al. 1999). As in previous studies, we mitigate the effects of subgrid-scale variability by applying a temperature lapse rate correction (Mass et al. 2002) and using time averaging (Hinkelman et al. 1999). As a field is averaged over time,

the spatial heterogeneity within a grid cell would be reduced and, for longer averaging periods, the grid cell and point observation would tend to converge. For this study, we have selected 1-day and 5-day precipitation accumulation and compared annual statistics.

To be useful in representing local impacts of heavy precipitation, we expect that at some grid spacing, practical for regional climate modeling, the simulated daily precipitation will reflect the extreme values observed at stations reasonably well. In fact, this is the hypothesis explored in this paper. Our motivation is to ascertain whether regional climate model results are adequate for assessing impacts such as flooding of urban areas or small mountainous water sheds, which require simulating extreme daily intensities comparable to station observations. We do not require the daily time series from the regional model to correlate well with the observed time series, which would test the ability of the forcing reanalysis and regional simulation to match the observed timing of events at the station. This correlation could be reduced by minor errors in the simulated location of mesoscale precipitation systems, which would not necessarily imply poor model performance in the climatological sense. By comparing only annual statistics of heavy precipitation, the test is less demanding yet evaluates the statistical results most important to impacts studies.

This work is organized as follows. The models, experimental design, and observations are briefly described in sections 2, 3, and 4, respectively. A comparison of models simulations with observations is discussed in section 5. Major conclusions and discussions are presented in section 6.

2. Models description

a. WRF model

The WRF model is a mesoscale numerical weather system designed for short-term weather forecast as well as long-term climate simulation (http://www.wrf-model.org). It is a nonhydrostatic model with many different choices for physical parameterizations suitable for a broad spectrum of applications across scales ranging from meters to thousands of kilometers. The physics package includes microphysics, cumulus parameterization, planetary boundary layer, land surface models (LSM), and longwave and shortwave radiation (Skamarock et al. 2006).

In this work, WRF version 2.2 was used. The microphysics and convective parameterizations were the WRF single-moment 5-class (WSM5) scheme (Hong et al. 2004) and the Kain–Fritsch scheme (Kain and Fritsch 1993), respectively. The land surface model used was the Noah (NCEP, Oregon State University, Air Force, and Hydrologic Research Laboratory) LSM four-layer soil temperature and moisture model with canopy moisture and snow cover prediction (Chen and Dudhia 2001).

b. HadRM model

HadRM (Jones et al. 2004) is the third-generation regional climate model (HadRM3P) developed at the U.K. Met Office Hadley Centre. It is a limited-area, high-resolution version of the atmospheric general circulation model HadAM3P, which is itself a high-resolution version of the atmospheric component of the atmosphere–ocean coupled third climate configuration of the Met Office Unified Model (HadCM3) (Gordon et al. 2000; Johns et al. 2003).

HadRM is a hydrostatic version of the fully primitive equations. Model parameterizations include dynamical flow, horizontal diffusion, clouds and precipitation, radiative processes, gravity wave drag, land surface, and deep soil (Jones et al. 2004).

The horizontal resolution of the HadRM model grid is $0.22° \times 0.22°$ (although a resolution of $0.44° \times 0.44°$ is also available). The HadRM latitude–longitude grid is rotated in a way that the equator lays inside the region of interest. This permits a quasi-uniform gridbox area over the region of interest with a minimum horizontal resolution of ~25 km at the rotated equator.

HadRM was released as part of the Providing Regional Climates for Impacts Studies (PRECIS) package (http://precis.metoffice.com). This package also includes software to allow processing and display of the model output data. The PRECIS package is flexible, user friendly, and computationally inexpensive. It can easily be applied over any regions of the globe to provide detailed climate information for regional climate studies and climate impacts assessment.

3. Experimental design

The experimental design follows Zhang et al. (2009) and is briefly described here. WRF was set up by using multiple nests (Fig. 1a). The outermost domain at 108-km resolution covers nearly the entire North American continent and much of the eastern Pacific Ocean and the western Atlantic Ocean. The second domain at 36-km resolution (WRF-36) encompasses the continental United States and parts of Canada and Mexico. The innermost domain at 12-km resolution (WRF-12) covers the U.S. Pacific Northwest (Fig. 1b). Thirty-one vertical levels were used in the model spanning from the surface to 10 mb with the highest resolution (~20–100 m) in the boundary layer. One-way nesting was applied in this study.

We chose the highest available resolution (~25 km) for the domain of HadRM (Fig. 1a). The HadRM model domain includes a large part of the eastern Pacific Ocean, the western United States, and parts of Mexico and Canada to better represent the synoptic weather systems that affect the Pacific Northwest. There are 19 vertical hybrid levels in HadRM spanning from the surface to 0.5 mb.

The WRF and HadRM runs were initialized at 0000 UTC 1 December 2002 and ended at 0000 UTC 31 December 2007. The first one-month simulations by WRF and HadRM were regarded as model spinup. Such a short spinup is not ideal for all applications. However, for the current study, we found no significant difference (in terms of extreme event statistics and comparison to the observations) between the first year of simulation and the following years; thus, the full five years will be considered in this analysis.

The initial and lateral boundary conditions were interpolated from the NCEP–Department of Energy (DOE) Atmospheric Model Intercomparison Project (AMIP-II) reanalysis (R-2) data (Kanamitsu et al. 2002). The lateral boundary conditions were updated every six hours for both models. The use of reanalysis fields, where observations are assimilated into an atmospheric model, for evaluating the regional models has two advantages over using simulations forced by a free-running global climate model. First, errors in the large-scale climatology from the reanalysis are small, thereby isolating deficiencies in the forcing fields and regional model. Second, the reanalysis represents the observed interannual and seasonal variability, which is then incorporated into the regional simulation. The climate of the western United States is characterized by substantial climate variability at interannual to decadal time scales, and the timing of this variability is not coincident between observations and a free-running global model, even if variability is simulated well. However, the reanalysis represents the observed large-scale climate variability both in timing and magnitude. Thus, we do not need to average over several cycles of natural climate variability (several decades) to achieve reliable statistics, and a relatively short simulation is adequate for evaluation.

Sea surface temperature (SST) was updated every six hours in WRF using the real-time, global, sea surface temperature (RTG_SST) analysis (ftp://polar.ncep.noaa.gov/pub/history/sst) developed and archived at NCEP. In HadRM, SST was taken from a combination of the monthly Hadley Centre Sea Ice and SST dataset (HadISST) (http://badc.nerc.ac.uk/data/hadisst) and weekly NCEP observed datasets (http://www.cdc.noaa.gov/cdc/reanalysis/reanalysis.shtml). WRF and HadRM SST forcing are nearly identical over the entire model domains, so the use of different datasets for SST is not likely to have any significant effects on the results. The simulations from both WRF and HadRM models were output every hour.

4. Observations

Model simulations from WRF and HadRM are compared with observations at 72 U.S. Historical Climatology Network (HCN) stations (Karl et al. 1990) in the states of Washington, Oregon, and Idaho, following the methods in Zhang et al. (2009). These observations have been subject to a suite of quality assurance checks such as tests to detect duplicated data, climatological outliers, and various inconsistencies (internal, temporal, and spatial). Data and detailed information on quality tests can be found at http://www.ncdc.noaa.gov/oa/climate/ghcn-daily/.

We select only the stations from which 80% or more of the daily precipitation and temperature measurements are available during 2003–07 and the stations whose corresponding model grid points are land grid points in the WRF and HadRM model domains. The locations of these HCN stations are indicated in Fig. 1b. A lapse rate correction for terrain differences between station locations and model grid cells is applied as described below. This correction reduces the discrepancy between the area average represented by the regional model grid cell and the point value represented by the station observation.

The precipitation at a single station is not well represented by the model value, which reflects the aggregate precipitation over the grid box. Local terrain and mesoscale effects can produce unresolved heterogeneity within a grid cell. For coarse-resolution models, many stations may be combined, but for regional models the grid spacing is generally finer than the station network. One alternative would be to aggregate the regional model grid cells to coarser resolution, on the order of 100 km, and compare the model against the average over many stations. This approach would evaluate the regional climate simulations without regard to differences in their grid spacing and would test whether the models simulated the regional-scale precipitation statistics without regard to poorly resolved local effects. The objective of the current paper, however, is to ascertain whether—and at what spatial resolution—regional models can represent the intensity and frequency of heavy precipitation as observed at a station. This is an important objective since regional climate models are applied to studies of flooding impacts over small urban and mountain watersheds (e.g., Rosenberg et al. 2010).

TABLE 1. Definitions list of extreme precipitation and temperature indices. The indices and their definition follow primarily the recommendations from the Climate Variability and Predictability (CLIVAR) Expert Team on Climate Change Detection and Indices.

1	Number of days with daily precipitation greater than 10 mm
2	Number of days with daily precipitation greater than 20 mm
3	Number of days with daily precipitation greater than 40 mm
4	Number of wet days when daily precipitation greater than 1 mm
5	Maximum number of consecutive wet days with daily precipitation greater than 1 mm
6	Maximum number of consecutive dry days with daily precipitation less than 1 mm
7	Simple daily precipitation index (the average daily precipitation on a wet day)
8	Total precipitation in wet days
9	Monthly maximum 1-day precipitation
10	Monthly maximum 5-day precipitation
11	Number of frost days with daily minimum temperature less than 0°C
12	Number of summer days with daily maximum temperature greater than 25°C
13	Number of days with daily maximum temperature greater than 30°C
14	Number of days with daily maximum temperature greater than 35°C
15	Maximum value of daily maximum temperature
16	Minimum value of daily minimum temperature

5. Results

We will focus on indices of extreme weather using a variety of intensity thresholds to assure adequate statistics for the 5-yr period (see below). For this analysis, we compare observed and simulated annual statistics of extreme events to test how well the regional models reproduce the magnitude and frequency of extreme events observed over that period of time over the U.S. Pacific Northwest. Definition of these extreme weather indices based on daily maximum and minimum temperatures and precipitation are described in Table 1.

The daily maximum and minimum temperatures ($T_{\max}$ and $T_{\min}$) are obtained from simulated hourly temperature with terrain adjustment performed in the same manner as in Zhang et al. (2009) and are the values for the single-model grid cell containing the corresponding station location. The terrain adjustment is performed to account for differences in altitude between the station and the model grid point. To summarize, we compute the local daily lapse rate (λ_r) at each station location using the following formula:

$$\lambda_r = \frac{1}{n}\sum_{i=1}^{n}\frac{|T_r - T_i|}{|h_r - h_i|},$$

where T represents the surface air temperature from the model simulation and h represents the terrain elevation of the grid cell. The r and i subscripts stand for the grid cell of reference (for which the lapse rate is computed in the WRF and HadRM domains) and one of the four closest grid cells, respectively. Note that among these four neighborhood grid cells, only the ones with an elevation at least 100 m higher or lower than the grid cell of reference were used in this formula. Otherwise, the standard lapse rate of 6.5°C km^{-1} was used. Finally, the computed lapse rates were constrained to the interval from 2° to 7°C km^{-1} in agreement with what has been observed (Minder et al. 2009). We noticed a rather small difference between using this computed lapse rate and the standard one. For R2, we used the standard lapse rate (6.5°C km^{-1}).

Simulated precipitation values are the daily accumulated total precipitation and are taken from the single grid cell containing the station coordinates. No lapse rate was applied to precipitation since a lapse rate over complex terrain depends on several factors such as mountain width, buoyancy, and moisture fields (Smith

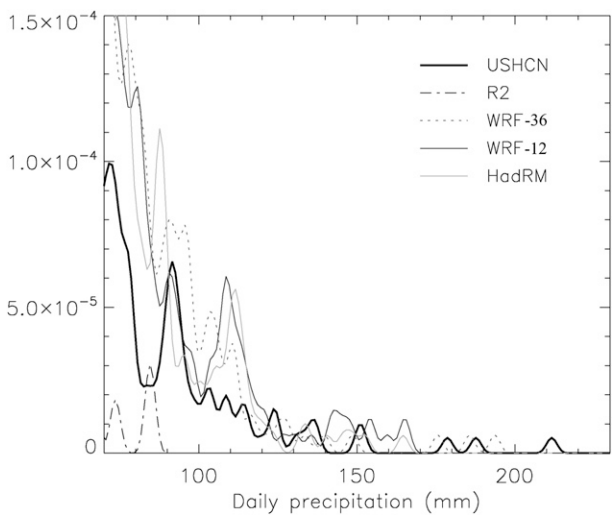

FIG. 2. Normalized probability distribution function of daily precipitation (mm) at HCN stations location from observations (thick black), R2 reanalysis (dash dotted), and WRF-36 (dotted gray), WRF-12 (black) and HadRM (gray) simulations.

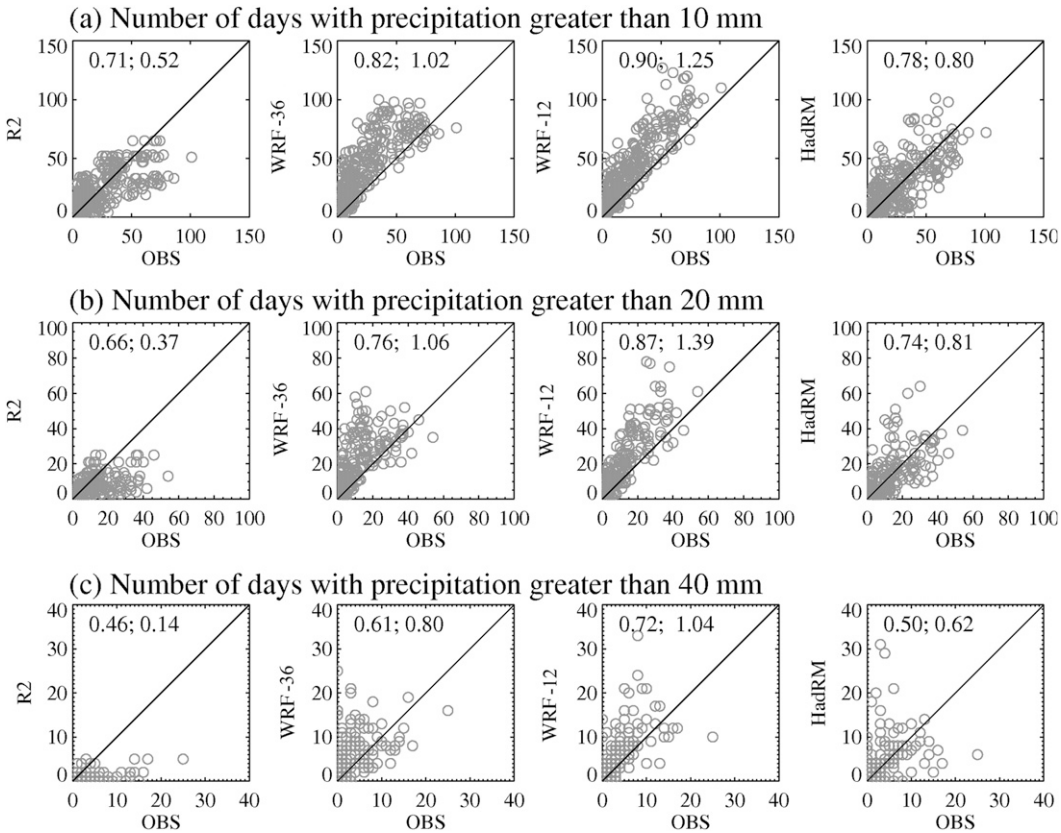

FIG. 3. Scatterplots of annual number of days with daily precipitation greater than (a) 10, (b) 20, and (c) 40 mm between observations and (first column) R2 reanalysis data and between observations and (second column) WRF-36, (third column) WRF-12, and (fourth column) HadRM simulations. Each point reflects the extreme index at one HCN station for each year over the 5-yr period. The two numbers in each scatterplot correspond to the correlation coefficient and linear regression slope, respectively. Except for the number of days with precipitation greater than 40 mm in R2, all correlation coefficients are significant at a significance level of 0.05 based on *t* statistics.

and Barstad 2004) as well as winds (Esteban and Chen 2008).

a. Extreme precipitation indices

For extreme precipitation, we mainly focus on annual values of 1) number of days with precipitation over certain thresholds (10, 20, and 40 mm), 2) number of wet days, 3) maximum number of consecutive wet or dry days, 4) simple daily precipitation index, 5) total precipitation in wet days, and 6) monthly values of maximum precipitation within 1 day and 5 consecutive days.

Figure 2 shows the tail of the normalized probability density function (pdf) of daily precipitation greater than 70 mm over the 5-yr period with all stations combined. Station observations show 26 days with daily precipitation greater than 100 mm; the regional climate models simulate frequencies of 32, 41, and 34 days for WRF-36, WRF-12, and HadRM, respectively, combining all grid cells containing the HCN station locations. However, the reanalysis data never generate precipitation greater than 100 mm and severely underestimates the number of days with precipitation between 70 and 100 mm. Notice in Fig. 2 that the regional climate models generally overestimate the number of days with daily precipitation lower than 130 mm.

Scatterplots for the annual number of days with daily precipitation greater than 10, 20, and 40 mm are presented in Fig. 3; each point represents the results for a single HCN station/regional model gridcell pair and for a single year over the 5-yr period. Correlation coefficients and linear regression slopes are given in each panel and indicate how well the simulations capture the variation in extreme precipitation across the region. All correlation coefficients presented in this figure are significant at 95% confidence based on *t* statistics, except for the number of days with precipitation greater than 40 mm in R2. There is a clear tendency of decreasing correlation coefficient with increasing threshold, which suggests that models have difficulty in resolving the increasingly heavy precipitation. The R2 reanalysis consistently shows the

lowest correlation coefficients and slopes. For instance, the correlation coefficient is merely 0.46 with a slope of 0.14 for the number of days with precipitation greater than 40 mm day^{-1}. Among the regional models, WRF-12 always shows the highest correlation coefficients, which reflects the better representation of terrain effects and spatial heterogeneity due to its higher resolution. Note also that WRF tends to overestimate the frequency of extreme precipitation.

Hence, the regional models show improvement over the R2 reanalysis data in simulating the observed magnitude of extreme precipitations. However, it is important to know whether or not the models also simulate the timing of extreme precipitation in response to the large-scale forcing from the reanalysis. To evaluate the timing of heavy precipitation events, we examined the probability of occurrence of a wet day (daily precipitation greater than 1 mm) in the reanalysis data and regional models on days when the observed daily precipitation is greater than a certain threshold (Fig. 4). More than 80% of the time, when precipitation is observed, the reanalysis data also indicate precipitation (Fig. 4). This percentage increases to more than 93% when the observed daily precipitation is greater than 10 mm. Similar probabilities are noted in the regional models (Fig. 4), confirming that timing of precipitation in the regional models is dictated by the reanalysis data.

Next, we examine the frequency and duration of wet and dry days. Simulated gridcell values for wet and dry days for large grid spacing would tend to poorly represent station values. When precipitation is scattered within a grid cell, the gridcell value will indicate a wet day even if no precipitation is observed at a particular point within the cell. Thus, we expect the models will overrepresent wet days and underrepresent dry days. Scatterplots of the annual number of wet days (daily precipitation greater than 1 mm, Fig. 5a) simulated and observed at each station show overestimation not only in the R2 reanalysis but also in the regional models. Correlation coefficient and slope for R2 are 0.52 and 0.39, respectively. Correlations for the regional models are higher than for R2, with the WRF-36 showing the highest correlation coefficient (0.78) while HadRM shows the best slope (0.64) among the regional models. Scatterplots of the annual maximum number of consecutive wet days (maximum number of consecutive days with daily precipitation greater than 1 mm, Fig. 5b) show overestimation in R2 and regional models. The correlation coefficients between R2 and the regional climate models are similar. The slopes for the R2 reanalysis and HadRM model are comparable and better than the slopes for the WRF model. Similarly, the annual maximum number of consecutive dry days is underestimated by R2 and regional

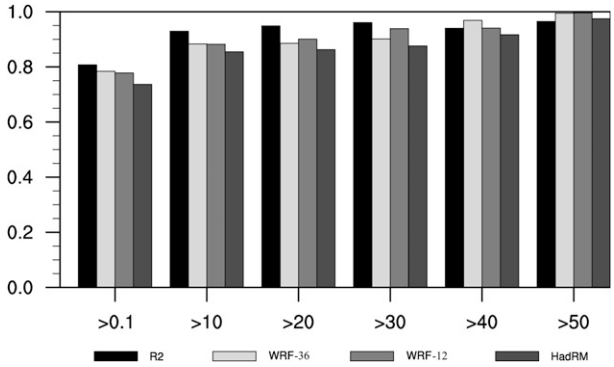

FIG. 4. Probability of a wet day in the reanalysis data (black), WRF-36 (light gray), WRF-12 (medium gray), and HadRM (dark gray) given that the corresponding observed daily precipitation is greater than a certain threshold (0.1, 10, 20, 30, 40, and 50 mm). The probability is computed at each station location, and then spatially averaged.

models (Fig. 5c). This result is consistent with the overestimation of the annual number of wet days (Fig. 5a) as well as the maximum number of consecutive wet days (Fig. 5b). The correlation coefficients and the slopes for the maximum number of consecutive dry days are rather small (between 0.40 and 0.56 for correlations and between 0.30 and 0.46 for slopes). However, among the regional models WRF-12 shows slightly better correlation coefficient and slope. For the annual values of the number of wet days, maximum number of consecutive wet days, and maximum number of consecutive dry days we do not find substantial improvement from the regional climate models over the driving data.

The simple daily precipitation index, defined as the average precipitation during wet days, provides a useful metric for evaluating the simulated precipitation from the models. While this index does not focus on heavy events exclusively, it does separate the frequency of precipitation (i.e., the number of wet days, discussed above) from the intensity of precipitation. Scatterplots of simple daily precipitation index in mm day^{-1} are displayed in Fig. 5d. Correlation coefficients are higher and slopes are closer to 1 for regional models than for the R2 reanalysis with the highest correlation coefficient and the best slope noted for WRF-12 among the regional domains. As mentioned before, the timing of rain-bearing storms in regional models is determined by the large-scale driving data; however, the magnitude strongly depends on the interactions of the local terrain with the large-scale weather systems. Regional models, with their improved representation of local terrain, better resolve orographic precipitation and yield precipitation intensities closer to the observations. This result can also be seen in the spatial distribution of a simple daily

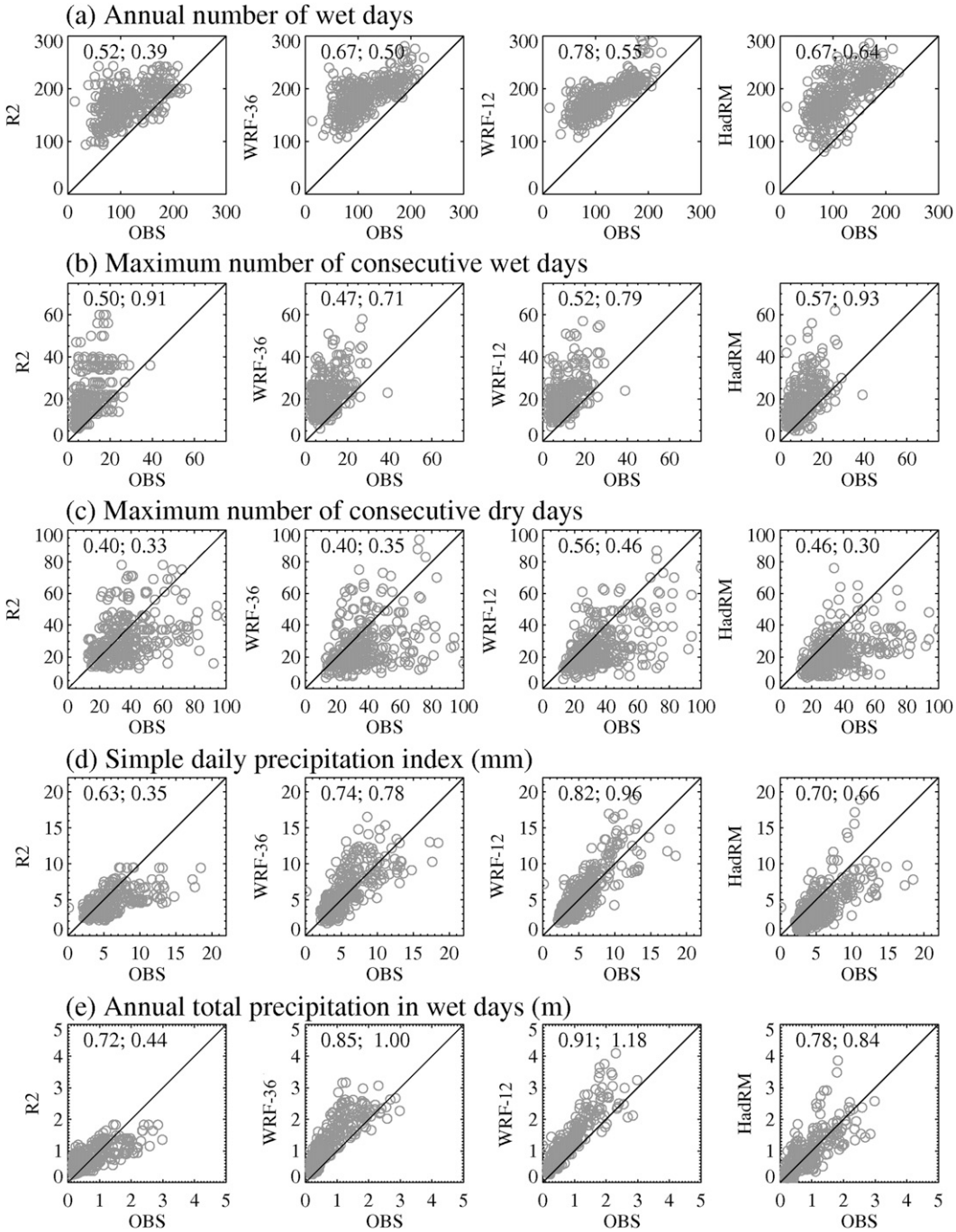

FIG. 5. Scatterplots of various annual precipitation indices between observations and (first column) R2 reanalysis and between observations and (second column) WRF-36, (third column) WRF-12, and (fourth column) HadRM simulations. Each point reflects the extreme index at one HCN station for each year over the 5-yr period. The two numbers in each scatterplot correspond to the correlation coefficient and linear regression slope, respectively. All correlation coefficients are significant at a significance level of 0.05.

precipitation index (Fig. 6) where larger (smaller) precipitation intensities occur on the windward (leeward) side of the Cascade Range in observations and regional model simulations, while the R2 reanalysis data display a relatively homogeneous pattern and a smaller gradient between windward and leeward regions.

Figure 5e presents the scatterplots of annual total precipitation; this measure is the product of the precipitation

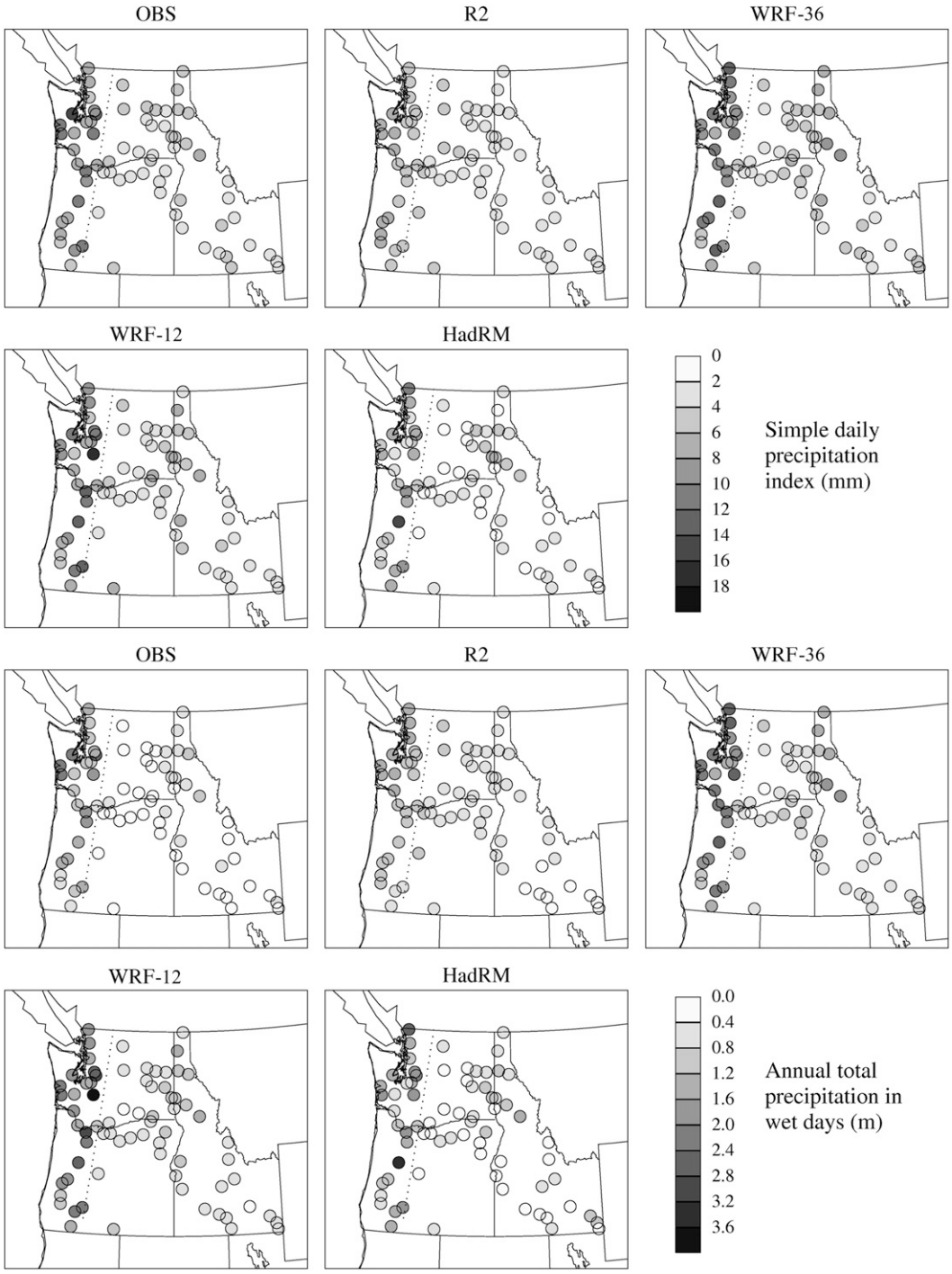

FIG. 6. (top two rows) Geographical distributions of simple daily precipitation index and (bottom two rows) annual total precipitation in wet days in average over the 5-yr period for HCN observations, R2 reanalysis, and WRF-36, WRF-12, and HadRM simulations at stations location. The Cascade crest is represented by the dotted line.

index and the number of wet days. The benefit of regional climate models over the R2 reanalysis data is clear in this figure as evidenced in higher correlation coefficients and slopes closer to 1. This is also clearly indicated by the geographical distribution of the annual total precipitation in wet days (Fig. 6). Although the correlation coefficient of WRF-12 is highest among the regional domains, note that this domain also overestimates the annual total

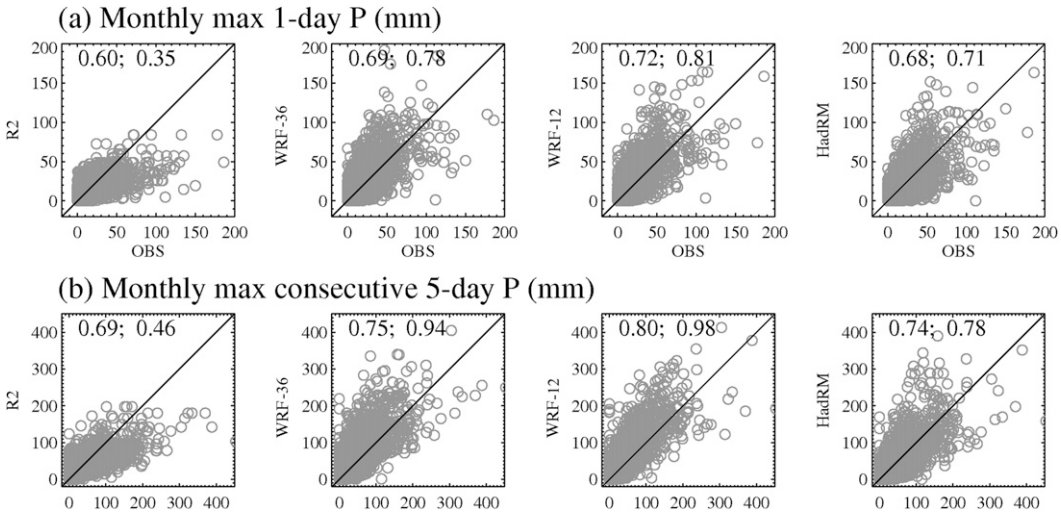

FIG. 7. Scatterplots of monthly maximum (a) 1-day and (b) 5-day accumulated precipitation between observations and (first column) R2 reanalysis and between observations and (second column) WRF-36, (third column) WRF-12, and (fourth column) HadRM simulations. Each point reflects the extreme index at one HCN station for each calendar month over the 5-yr period. Precipitation is given in mm. The two numbers in each scatterplot correspond to the correlation coefficient and linear regression slope, respectively. All correlation coefficients are significant at a significance level of 0.05.

precipitation in wet days due to the already discussed overestimation of the annual number of wet days (Fig. 5a).

A common metric for the magnitude of heavy precipitation used in infrastructure design and planning are the maximum 1-day and 5-day accumulated precipitation observed over a period at a location. Figure 7 presents scatterplots of maximum 1-day and 5-day accumulated precipitation for each calendar month over the 5-yr period. The correlation coefficients between the R2 reanalysis and the observations for maximum 1-day and 5-day precipitation are 0.60 and 0.69, respectively. The regional models show substantially improved correlations, of 0.72 and 0.80 for the WRF-12 simulation. Additionally, the slopes for the regional models are considerably better and closer to 1 than those for the reanalysis data. Severe underestimation of the observed precipitation is noted in the R2 reanalysis data especially for maximum 1-day precipitation greater than 100 mm and 5-day accumulations greater than 200 mm. Here again, WRF-12 with its highest resolution exhibits the highest correlation coefficients and the best slopes among the regional domains. For 5-day accumulations, the WRF-12 simulation represents the full range of intensities that are observed by the stations.

Table 2 presents the statistical values of the correlation coefficient and linear regression slope computed from scatterplots of two different precipitation indices (viz., the simple daily precipitation index and number of days with daily precipitation greater than 10 mm) between observations and R2 reanalysis and between observations and model simulations. Statistical values are computed for each year over the 2003–07 period. Year-by-year values are consistent with their corresponding 5-yr values (cf. Figs. 3 and 5). Also, for each model, the year-by-year statistical values sit fairly close to each other suggesting model consistency at simulating extreme precipitation events for each year of the simulation. Finally, note that values of correlation coefficients and slopes for year 2003 sit within the range of the corresponding values for the subsequent years, supporting our inclusion of 2003 in the analysis despite the one-month spinup.

To summarize, the above analysis suggests that the R2 reanalysis data resolve the timing of the rain-bearing storms relatively well; however, R2 reanalysis data show poor performance in capturing extreme precipitation events. This suggests that the reanalysis data adequately represent the well-resolved fields such as moisture flux and synoptic storms. Because of this, extreme precipitation can be adequately simulated in regional models given boundary conditions from the reanalysis. That is, the large-scale conditions that control the spatial distribution of heavy precipitation are well represented by the reanalysis, and the regional models can simulate the local effects (such as orographic enhancement and mesoscale weather patterns) that produce heavy precipitation. This is especially true for WRF-12, which shows the best statistical performance among the regional domains, indicating the importance of model resolution in simulating extreme precipitation. This conclusion should

TABLE 2. Precipitation correlation coefficient (ρ) and linear regression slope (s) of scatterplots between observations and R2 reanalysis and between observations and WRF-36, WRF-12, and HadRM simulations.

	Simple daily precipitation index								No. of days with daily precipitation greater than 10 mm							
	R2		WRF-36		WRF-12		HadRM		R2		WRF-36		WRF-12		HadRM	
Year	ρ	s	ρ	s	ρ	s	ρ	s	ρ	s	ρ	s	ρ	s	ρ	s
2003	0.65	0.34	0.79	0.78	0.87	0.95	0.74	0.69	0.75	0.53	0.82	1.00	0.92	1.22	0.78	0.81
2004	0.52	0.28	0.76	0.88	0.86	1.15	0.73	0.72	0.62	0.37	0.82	1.08	0.92	1.29	0.78	0.78
2005	0.57	0.31	0.65	0.71	0.77	0.94	0.61	0.58	0.70	0.53	0.83	1.03	0.90	1.29	0.77	0.85
2006	0.67	0.38	0.74	0.71	0.80	0.84	0.69	0.61	0.75	0.55	0.81	0.92	0.89	1.11	0.77	0.72
2007	0.62	0.34	0.75	0.93	0.84	1.11	0.72	0.78	0.68	0.54	0.84	1.20	0.92	1.51	0.82	0.94

likewise apply when using regional models to downscale global climate models with similar resolution to the R2 reanalysis, provided the climatology of large-scale circulation and moisture transport is well represented by the global model.

b. Extreme temperature indices

For extreme temperatures we examined the annual number of 1) frost days, 2) summer days, 3) days with T_{max} greater than 30° and 35°C, and 4) the monthly extreme values of T_{max} and T_{min}. The results are presented in Figs. 8 and 9.

R2 is comparable to regional climate models in terms of correlation coefficients for the number of frost days but with a better slope (Fig. 8a). This might be related to large deficiencies in the regional climate models during nighttime (Zhang et al. 2009). As also noted in Zhang et al. (2009), during the same 5-yr simulation period, a warm bias of T_{min} on the order of 2°C is identified over the Pacific Northwest in both WRF and HadRM simulations. This warm bias tends to reduce the number of frost days in the regional models. Furthermore, a cold bias of T_{min} on the order of 1°C is noted in the R2 reanalysis, which would result in an excess number of frost days. These biases are reflected in the scatterplots with the R2 reanalysis showing the majority of the points above the 1:1 line and the regional model results falling below the line.

For the annual number of summer days and the annual number of days with T_{max} greater than 30° and 35°C (Figs. 8b–d), the regional models consistently show higher correlation coefficients and significantly better slopes compared to the R2 reanalysis. Severe underestimation identified in these categories in the R2 reanalysis might be related to the cold bias of T_{max} on the order of 3°C (Zhang et al. 2009). As pointed out in the same paper, the regional models with higher resolution tend to partially reduce this large bias to values less than 1°C.

All scatterplots in Fig. 8 display a larger spread in R2 than in regional models, suggesting that regional models better represent the spatial pattern of extreme temperatures compared to R2, even with a lapse rate correction applied, which accounts for the effects of topography. Temperature variability is primarily dictated by large-scale weather systems with finescale terrain features (e.g., land cover, albedo, soil moisture, and cloudiness) playing a secondary role in modulating the local temperatures. While the R2 reanalysis depicts the large-scale weather systems well, the improved resolution of the regional models is better able to represent mesoscale processes (Salathé et al. 2008) and land surface characteristics, which yields the narrower spreads in the scatterplots.

The correlation coefficients and slopes for the extreme temperature indices (Fig. 8) do not differ significantly between the regional models in contrast to the extreme precipitation results. This suggests that higher resolution does not lead to better model performance beyond a certain threshold. This may be in part due to the lapse rate correction, which accounts for variations in topographic relief among the models. As opposed to topography, variations in land cover characteristics are likely sufficiently well resolved even at 36-km grid spacing to simulate the metrics evaluated in this study.

Figure 9 shows the annual extreme T_{max} and T_{min} for each year. Note that these annual maximum and minimum extremes basically refer to summer and winter extremes, respectively. For the annual maximum value of daily T_{max} (Fig. 9a), the correlation coefficient and slope corresponding to the R2 reanalysis are 0.59 and 0.61, respectively. The regional models show considerably higher correlation coefficients (~0.80) and much better slopes (~0.90). In terms of annual minimum values of T_{min} (Fig. 9b), the correlation coefficients and slopes corresponding to the regional models do not differ appreciably from those for the R2 reanalysis except for HadRM regional model, which shows a rather low slope (0.53). The R2 reanalysis strongly underestimates the annual minimum value of T_{min} by as much as about 8°C (Fig. 9b). The WRF domains both show small bias in the annual minimum value of T_{min}, while HadRM shows a warm bias on the order of ~3°C. Zhang et al. (2009) suggest that the

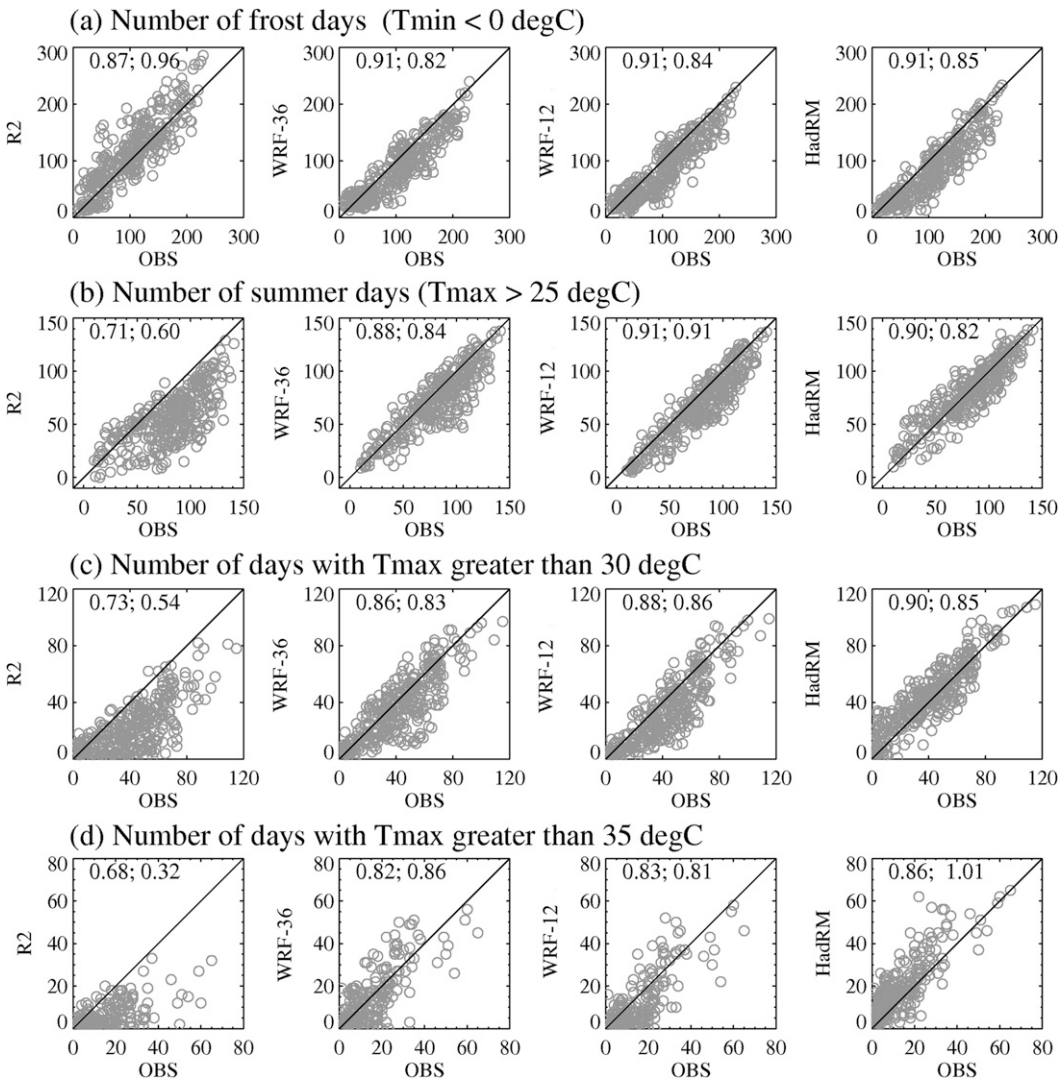

FIG. 8. Scatterplots of annual temperature indices (°C) between observations and (first column) R2 reanalysis and between observations and (second column) WRF-36, (third column) WRF-12, and (fourth column) HadRM simulations. Each point reflects the extreme index at one HCN station for each year over the 5-yr period. The two numbers in each scatterplot correspond to the correlation coefficient and linear regression slope, respectively. All correlation coefficients are significant at a significance level of 0.05.

regional models have warm biases that might offset the large cold biases of R2 in winter.

Table 3 shows statistical values of the correlation coefficient and linear regression slope computed from scatterplots of two different extreme temperature indices (viz., the annual maximum of maximum daily temperature and the annual minimum of minimum daily temperature) between observations and R2 reanalysis and between observations and model simulations. Statistical values are computed for each year over the 2003–07 period. Values found in Table 3 are consistent with the corresponding 5-yr values presented in Fig. 8. Moreover, each model presents values among the five years that are close to each other, showing model consistency at simulating extreme temperatures.

6. Conclusions and discussion

This work examines the performance of two regional models, WRF and HadRM, in simulating station observations of several indices of extreme temperature and precipitation for the U.S. Pacific Northwest during a 5-yr period (2003–07). Our goal is to establish whether regional climate models at typical grid spacings (between 36 and 12 km) can represent the intensities and frequencies of extreme events at the scale important for

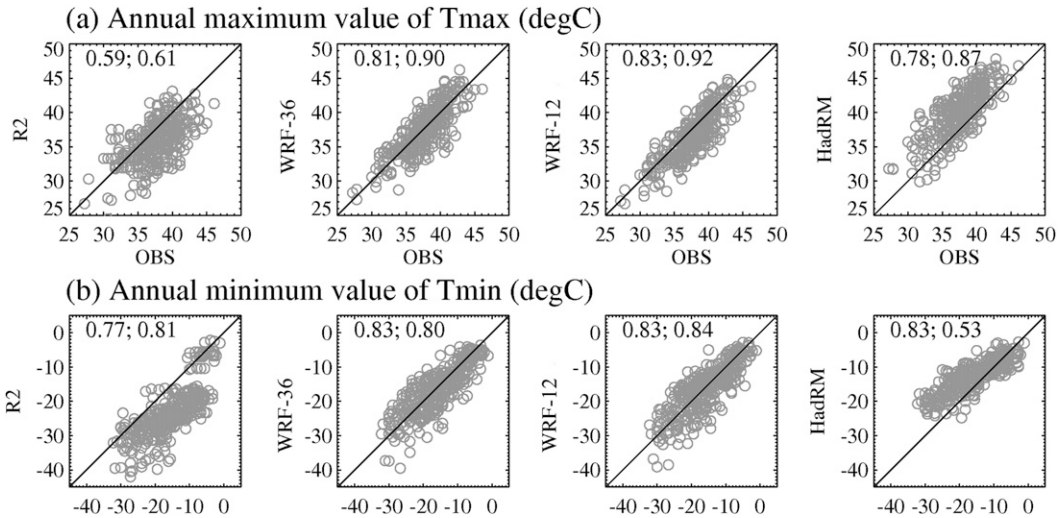

FIG. 9. Scatterplots of (a) annual maximum value of T_{max} and (b) annual minimum value of T_{min} (°C) between observations and (first column) R2 reanalysis and between observations and (second column) WRF-36, (third column) WRF-12, and (fourth column) HadRM simulations. Each point reflects the extreme index at one HCN station for each year over the 5-yr period. The two numbers in each scatterplot correspond to the correlation coefficient and linear regression slope, respectively. All correlation coefficients are significant at a significance level of 0.05.

climate impacts assessment, and thus we evaluate the models against observations at single-point stations. Since models represent the areal average over individual grid cells, they cannot represent events whose spatial extent is small compared to the grid cell. This problem is mitigated in two ways in this study. For temperature, we apply a lapse-rate correction to the gridcell value to account for the elevation disparity between the station and the grid cell. Other finescale terrain and weather processes would still contribute to a disparity between the gridcell average and a point observation. For precipitation, it is not feasible to remove the effect of elevation as for temperature. In this case, we consider precipitation accumulations over 1- and 5-day periods. Averaging in time essentially smoothes over instantaneous and highly localized extremes that would be observed at a station.

a. Precipitation

Our analysis indicates that, while the R2 reanalysis data represent the timing and intensities of rain-bearing storms over the Pacific Northwest well, they cannot represent the observed spatial distribution of extreme precipitation indices when compared to the WRF and HadRM simulations. This is explained by the rather coarse resolution of the R2 reanalysis system that cannot simulate the magnitude of locally intense precipitation, which depends on the influence of local complex terrain and mesoscale weather systems. Thus, the R2 reanalysis data provide realistic large-scale boundary conditions necessary for driving regional climate models and allow the regional models to simulate locally intense precipitation events that are not captured in the reanalysis. This conclusion may also hold true for dynamically downscaling global climate models provided they simulate realistic large-scale patterns. Comparing regional simulations at multiple grid spacing and two models illustrates the importance of fine grid spacing in simulating extreme precipitation.

The WRF and HadRM simulations resolve the observed extreme precipitation indices reasonably well as

TABLE 3. As in Table 2, but for temperature.

	Maximum value of daily maximum temperature								Maximum value of daily maximum temperature							
	R2		WRF-36		WRF-12		HadRM		R2		WRF-36		WRF-12		HadRM	
Year	ρ	s	ρ	s	ρ	s	ρ	s	ρ	s	ρ	s	ρ	s	ρ	s
2003	0.66	0.72	0.77	0.80	0.81	0.87	0.73	0.76	0.70	0.78	0.86	0.78	0.82	0.81	0.77	0.41
2004	0.47	0.41	0.78	0.80	0.83	0.85	0.79	0.97	0.81	0.97	0.87	0.82	0.87	0.83	0.86	0.55
2005	0.67	0.65	0.82	0.90	0.81	0.82	0.82	0.76	0.78	0.83	0.84	0.74	0.86	0.83	0.84	0.57
2006	0.47	0.40	0.84	0.92	0.86	0.92	0.77	0.90	0.80	0.85	0.69	0.54	0.80	0.71	0.78	0.53
2007	0.58	0.72	0.78	0.79	0.81	0.83	0.67	0.73	0.84	0.64	0.90	0.83	0.91	0.76	0.85	0.52

reflected by high correlation coefficients and slopes close to 1 in terms of the extreme precipitation indices. The improvement of the regional models over the R2 reanalysis data in simulating the extreme precipitation events are primarily related to the better representation of the local complex terrain and mesoscale processes by the regional models. The WRF-12 with its highest resolution (~12 km) always shows the best statistical performance when compared to the WRF-36 (~36 km) and HadRM (~25 km).

Note that, by comparing model gridcell values to station observations, we are not evaluating the aggregated precipitation over the grid cell, which would require much finer spatial observations than are available. The results indicate how well the model can simulate the extremes observed at a station location. Finer model grid spacing should converge to the station observation as precipitation heterogeneity is better resolved. This approach is successful in that the regional models are able to approach the full range of extreme events observed across the region with the finer grid models showing better correspondence with observations for most precipitation parameters. The primary exception is the simulation of wet and dry day frequencies when even the simulation at 12-km grid spacing tends to overestimate the number of wet days compared to the station observation. These results underscore the importance of grid spacing for simulating local frequencies and intensities of heavy precipitation in a region of complex terrain.

b. Temperature

Appreciable improvement in the extreme temperature indices is also noted for WRF and HadRM when compared to the R2 reanalysis data. This is likely related to the capability of the regional models in resolving mesoscale processes associated with complex terrain. Daily temperature variability is primarily controlled by large-scale weather systems; however, mesoscale processes can modulate the temperature fields in a nontrivial way, especially over complex terrain. An improvement in simulating extreme temperature indices at finer grid spacing is found even with a lapse rate correction applied to model results, so elevation is not the only issue producing better local temperature simulations. Extremes in temperature depend also on radiative transfer, boundary layer dynamics, and latent and sensible heat transfer with the surface. These in turn depend on surface properties such as vegetation, snow cover, surface albedo, and soil moisture and temperature. The fine grid spacing of the regional models is necessary to adequately resolve local variations in these surface properties and provide realistic simulations of extreme temperatures.

We find that all simulations improve on the coarse large-scale forcing, suggesting that the models can better resolve the mechanisms of extreme temperature. However, the performance of the models is fairly uniform, with little difference between 36- and 12-km grid spacing. This result suggests that, once elevation is accounted for, even at 36-km spacing, the surface and mesoscale weather processes that affect extreme temperature are well resolved.

c. WRF and HadRM

HadRM and WRF are generally comparable in their performance in resolving the observed precipitation and temperature extremes at horizontal resolutions on the order of tens of kilometers. The higher vertical resolution in WRF than in HadRM may contribute to the slightly better performance by the WRF-12; however, the WRF-36 does not always outperform the 25-km HadRM even though the WRF-36 vertical resolution is better. For the WRF-36 and WRF-12 simulations, only the horizontal grid spacing differs, so the better resolution of terrain effects and other mesoscale processes accounts for the better performance. The HadRM model was run at 25-km grid spacing, yet shows performance comparable to the 36-km WRF simulation, especially for extreme precipitation. Thus, other aspects of the models or experiment design likely contribute to the differences in performance. These differences would include higher vertical resolution in WRF, more advanced numeric and dynamics in WRF, and more sophisticated schemes for the land surface, cloud microphysics, and convection. Nevertheless, the HadRM model is significantly less demanding computationally than the WRF model, and provides comparable performance to the 36-km WRF.

Acknowledgments. This work is funded by an Environmental Protection Agency STAR Grant, by the National Science Foundation (ATM0709856), and by a Microsoft Corporation gift to the Climate Impacts Group. We thank the PRECIS team from the U.K. Met Office, especially Richard Jones, David Hein, David Hassell, Simon Wilson, and Wilfram Moufouma-Okia for supplying the PRECIS package and helping us to use it. We thank Prof. Clifford Mass and Rick Steed for their insights in running the WRF model. The WRF simulations were performed at the National Center for Atmospheric Research (NCAR) Computational and Information System Laboratory (CISL). NCAR is sponsored by the National Science Foundation. This publication is partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement No. NA17RJ1232.

REFERENCES

Allan, R. P., and B. J. Soden, 2008: Atmospheric warming and the amplification of precipitation extremes. *Science,* **321,** 1481–1484.

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

Esteban, M. A., and Y.-L. Chen, 2008: The impact of trade wind strength on precipitation over the windward side of the Island of Hawaii. *Mon. Wea. Rev.,* **136,** 913–928.

Fowler, H. J., and M. Ekström, 2009: Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes. *Int. J. Climatol.,* **29,** 385–416.

——, ——, C. G. Kilsby, and P. D. Jones, 2005: New estimates of future changes in extreme rainfall across the UK using regional climate model integrations. 1. Assessment of control climate. *J. Hydrol.,* **300,** 212–233.

Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.,* **16,** 147–168.

Hinkelman, L. M., T. P. Ackerman, and R. T. Marchand, 1999: An evaluation of NCEP Eta model predictions of surface energy budget and cloud properties by comparison with measured ARM data. *J. Geophys. Res.,* **104,** 19 535–19 549.

Hong, S.-Y., J. Dudhia, and S.-H. Chen, 2004: A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. *Mon. Wea. Rev.,* **132,** 103–120.

Johns, T. C., and Coauthors, 2003: Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. *Climate Dyn.,* **20,** 583–612.

Jones, R. G., M. Noguer, D. C. Hassel, D. Hudson, S. S. Wilson, G. J. Jenkins, and J. F. B. Mitchell, 2004: Generating high resolution climate change scenarios using PRECIS. Met Office Hadley Centre Rep. 40 pp.

Kain, J. S., and J. M. Fritsch, 1993: Convective parameterization for mesoscale models: The Kain–Fritsch scheme. *The Representation of Cumulus Convection in Numerical Models, Meteor. Monogr.,* No. 46, Amer. Meteor. Soc., 165–170.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.,* **83,** 1631–1643.

Karl, T. R., J. C. N. Williams, F. T. Quinlan, and T. A. Boden, 1990: United States Historical Climatology Network (HCN) serial temperature and precipitation data. Carbon Dioxide Information and Analysis Center, Oak Ridge National Laboratory, Environmental Science Division Publ. 3404, 389 pp.

Leung, L. R., Y. Qian, and X. Bian, 2003a: Hydroclimate of the western United States based on observations and regional climate simulation of 1981–2000. Part I: Seasonal statistics. *J. Climate,* **16,** 1892–1911.

——, ——, ——, and A. Hunt, 2003b: Hydroclimate of the western United States based on observations and regional climate simulation of 1981–2000. Part II: Mesoscale ENSO anomalies. *J. Climate,* **16,** 1912–1928.

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc.,* **83,** 407–430.

May, W., 2008: Potential future changes in the characteristics of daily precipitation in Europe simulated by the HIRHAM regional climate model. *Climate Dyn.,* **30,** 581–603.

Minder, J. R., P. W. Mote, and J. D. Lundquist, 2010: Surface temperature lapse rates over complex terrain: Lessons from the Cascade Mountains. *J. Geophys. Res.,* **115,** D14122, doi:10.1029/2009JD013493.

Räisänen, J., and Coauthors, 2004: European climate in the late twenty-first century: Regional simulations with two driving global models and two forcing scenarios. *Climate Dyn.,* **22,** 13–31.

Rosenberg, E. A., P. W. Keys, D. B. Booth, D. Hartley, J. Burkey, A. C. Steinemann, and D. P. Lettenmaier, 2010: Precipitation extremes and the impacts of climate change on stormwater infrastructure in Washington State. *Climatic Change,* **102,** 319–349.

Salathé, E. P., R. Steed, C. F. Mass, and P. H. Zahn, 2008: A high-resolution climate model for the United States Pacific Northwest: Mesoscale feedbacks and local responses to climate change. *J. Climate,* **21,** 5708–5726.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, 2006: A description of the Advanced Research WRF version 2. NCAR Tech. Note NCAR/TN-468+STR, 88 pp.

Solomon, S., D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. B. Tignor, H. L. Miller Jr., and Z. Chen, Eds., 2007: *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, 996 pp.

Smith, R. B., and I. Barstad, 2004: A linear theory of orographic precipitation. *J. Atmos. Sci.,* **61,** 1377–1391.

Tebaldi, C., K. Hayhoe, J. M. Arblaster, and G. A. Meehl, 2006: Going to the extremes: An intercomparison of model-simulated historical and future changes in extreme events. *Climatic Change,* **79,** 185–211.

United Nations Environment Programme, 2002: State of the environment and policy retrospective: 1972–2002. *Global Environment Outlook 3 (GEO-3),* Earthscan Publications Ltd.

Zhang, Y., V. Dulière, P. Mote, and E. P. Salathé Jr., 2009: Evaluation of WRF and HadRM mesoscale climate simulations over the United States Pacific Northwest. *J. Climate,* **22,** 5511–5526.