Effects of Different Regional Climate Model Resolution and Forcing Scales on Projected Hydrologic Changes

Pablo A. Mendoza[1,2,3,*], Naoki Mizukami[3], Kyoko Ikeda[3], Martyn P. Clark[3], Ethan D. Gutmann[3], Jeffrey R. Arnold[4], Levi D. Brekke[5] and Balaji Rajagopalan[1,2]

[1]Department of Civil, Environmental, and Architectural Engineering, University of Colorado at Boulder, Boulder, Colorado, USA
[2]Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, Colorado, USA
[3]Hydrometeorological Applications Program, National Center for Atmospheric Research, Boulder, Colorado, USA
[4]Climate Preparedness and Resilience Programs, U.S. Army Corps of Engineers, Seattle, USA
[5]Bureau of Reclamation, Denver, USA

Running head: Effects of RCM resolution on hydrologic changes

Keywords: climate change; regional climate model; spatial aggregation; horizontal resolution; hydrologic model structure

Manuscript submitted to Journal of Hydrology

*Corresponding author. Email address: pmendoza@ucar.edu (Pablo A. Mendoza)

Abstract

We examine the effects of regional climate model (RCM) horizontal resolution and forcing scaling (i.e., spatial aggregation of meteorological datasets) on the portrayal of climate change impacts. Specifically, we assess how the above decisions affect: (i) historical simulation of signature measures of hydrologic behavior, and (ii) projected changes in terms of annual water balance and hydrologic signature measures. To this end, we conduct our study in three catchments located in the headwaters of the Colorado River basin. Meteorological forcings for current and a future climate projection are obtained at three spatial resolutions (4-, 12- and 36-km) from dynamical downscaling with the Weather Research and Forecasting (WRF) regional climate model, and hydrologic changes are computed using four different hydrologic model structures. These projected changes are compared to those obtained from running hydrologic simulations with current and future 4-km WRF climate outputs re-scaled to 12- and 36-km.

The results show that the horizontal resolution of WRF simulations heavily affects basin-averaged precipitation amounts, propagating into large differences in simulated signature measures across model structures. The implications of re-scaled forcing datasets on historical performance were primarily observed on simulated runoff seasonality. We also found that the effects of WRF grid resolution on projected changes in mean annual runoff and evapotranspiration may be larger than the effects of hydrologic model choice, which surpasses the effects from re-scaled forcings. Scaling effects on projected variations in hydrologic signature measures were found to be generally smaller than those coming from WRF resolution; however, forcing aggregation in many cases reversed the direction of projected changes in hydrologic behavior.

1. Introduction

Although global climate models (GCMs) are widely used for generating information on future climate scenarios, their native grid size (~100-200 km on a side) is a serious limitation for characterizing climate projections at the basin scale, where features such as elevation and aspect become relevant. To reconcile differences between coarse resolution GCM outputs and regional or local scale climate processes, Regional Climate Models (RCMs) are run with lateral boundary conditions from GCMs to force fine-scale climate simulations, a process typically referred to as dynamical downscaling (Xu 1999; Fowler et al. 2007). Teutschbein and Seibert (2010) presented a detailed review of approaches that make use of RCMs for quantifying climate change impacts on hydrologic processes, and a plethora of additional example applications can be found in the literature (e.g., Wood et al. 2004; Steele-Dunne et al. 2008; Suklitsch et al. 2008; Kay et al. 2009; Prudhomme and Davies 2009; Gao et al. 2011; Vicuña et al. 2011; Majone et al. 2012; Wi et al. 2012; Lauer et al. 2013; Velázquez et al. 2013).

However, a key aspect rarely explored is the choice of RCM horizontal resolution, which determines how precipitation – in particular snowfall – and other hydrologic variables are represented in highly heterogeneous regions (Rasmussen et al. 2011). For example, Kleinn et al. (2005) compared hydrologic model simulations forced with 14-km and 56-km RCM outputs in the Rhine basin in Central Europe, finding that although the finer resolution provided more realistic precipitation fields, improvements in streamflow simulation skill were small. Contrarily, Dankers et al. (2007) showed that 12-km simulations conducted with the HIRHAM RCM provided a better representation of orographic patterns and extreme precipitation events in the Upper Danube basin in Central Europe, and better simulations of hydrologic extreme events at

67 the sub-basin scale in comparison to coarser (50-km) RCM outputs. Graham et al. (2007)
68 concluded that a 25-km resolution provided more systematic and less spatially variable biases in
69 RCM precipitation and temperature fields when compared to 50-km resolution. Van Roosmalen
70 et al. (2010) evaluated the implications of choosing different RCM resolutions (12-, 25- and 50-
71 km) on delta change factors (e.g., additive perturbation for temperature, multiplier for
72 precipitation) – from a control and future climate scenario – computed at a monthly basis for
73 Denmark, finding that the added value of increasing resolution was almost negligible. A set of
74 studies conducted in the Colorado Headwaters Region (Ikeda et al., 2010; Rasmussen et al.,
75 2011, 2014) explored the effects of horizontal resolution using the Weather Research and
76 Forecasting (WRF) regional climate model (Skamarock et al. 2008). Specifically, they showed
77 that the use of horizontal resolution of 6 km or less in RCMs allowed accurate estimations of
78 vertical motions driven by topography without the need to include a convective parameterization
79 scheme, improving the representation of seasonal snowfall and snowpack. Along these lines,
80 Prein et al. (2013) compared the effects of different horizontal resolutions (4-, 12- and 36-km) on
81 daily heavy precipitation events simulated by WRF over the same domain, finding that only the
82 4-km simulation was able to reproduce heavy summertime events, and that both 4-km and 12-km
83 outputs were comparable and superior to the 36-km simulation when looking at winter events.
84 More recently, Olsson et al. (2015) obtained similar findings – i.e., better simulation of summer
85 extremes and summer wet spells – when moving from 50-km to 6-km horizontal grids.

86 The choice of the RCM resolution is typically determined by climate modelers to optimize
87 some constraints including available computer (i.e., time to compute the solution) and the need
88 to represent selected important atmospheric processes as explicitly as possible, but the domains
89 of these solutions are nearly always rectilinear for the Eulerian grid. However, hydrologic

90 modelers choose any shaped spatial element named Hydrologic Response Unit (HRU) at which
91 the hydrologic model is run. The HRUs can be the entire catchment, a grid box, or
92 hydrologically similar areas (e.g., similar soil-vegetation areas) and those are at different scale
93 than RCM resolution. Accordingly, scaling or spatial aggregation of RCM outputs is nearly
94 always required to obtain HRU averaged meteorological forcing. Several studies have examined
95 the hydrologic implications of spatially aggregating meteorological fields from finer scales (e.g.,
96 Finnerty et al. 1997; Koren et al. 1999; Bell and Moore 2000; Arnaud et al. 2002; Liang et al.
97 2004; Shrestha et al. 2006, 2007; Tramblay et al. 2011; Rasmussen et al. 2012), showing mixed
98 conclusions. Lobligeois et al. (2014) conducted a detailed review of previous efforts, and
99 analyzed the benefits of using high-resolution rainfall fields for flood simulation, including a
100 large sample of flood events (3620) in a large number of catchments (181); although they
101 concluded that these effects are "scale-dependent and event-specific-dependent", they also found
102 that regions with high spatial rainfall variability obtained the greatest benefits from high-
103 resolution precipitation inputs. Importantly, none of the above studies assessed the sensitivity of
104 hydrologic changes to the spatial scale at which historical and future climate datasets are used.

105 Given the evidence showing that RCM resolution affects climate outputs, a natural question
106 that arises is how the effects of RCM horizontal resolution on hydrologic portrayals of climate
107 change compare to those of scaled RCM at the same horizontal resolution. This paper examines
108 how the grid spacing adopted in a RCM for dynamical downscaling affects hydrologic change
109 estimates. In particular, we aim to characterize these effects on: (i) historical simulation of
110 signature measures of hydrologic behavior (e.g., runoff ratio, seasonality, log-term baseflow),
111 and (ii) projected hydrologic change in terms of annual water balance and hydrologic signature
112 measures. Further, we compare the implications of choosing different horizontal grid sizes to

113 those associated with spatial aggregation of high resolution RCM output. Given the increasing
114 awareness of the importance of hydrologic model structural uncertainty to climate change impact
115 studies (e.g., Boorman and Sefton 1997; Jones et al. 2006; Jiang et al. 2007; Kay et al. 2009;
116 Ludwig et al. 2009; Bae et al. 2011; Bastola et al. 2011; Najafi et al. 2011; Poulin et al. 2011;
117 Miller et al. 2012; Vano et al. 2012; Surfleet et al. 2012; Addor et al. 2014; Mendoza et al. 2015,
118 2016; Mizukami et al. 2016), we include four different hydrologic/land surface models for two
119 reasons: to examine the robustness of RCM resolution and forcing aggregation effects, and to
120 obtain insights on the relative importance of forcing-related decisions versus hydrologic model
121 choice.

122 The remainder of this paper is organized as follows. Section 2 provides a description of the
123 study domain. Section 3 describes the meteorological forcing data, hydrologic models and the
124 experimental design adopted in this study. Section 4 illustrates how the choice of RCM
125 horizontal resolution and forcing aggregation affect hydrologic portrayals – obtained from
126 different hydrologic models – under historical and modified climatic conditions. Finally, section
127 5 summarizes our main findings.

128 2. Study Area

129 The Colorado River basin (CRB) is one of the major water sources for consumption,
130 irrigation and hydropower in the western United States, draining parts of seven states and
131 Mexico, and covering the needs of more than 30 million people. Given its strategic relevance,
132 several studies have been conducted to quantify the potential effects of changes in precipitation
133 and temperature on the hydrology of this area (e.g., Milly et al. 2005; Christensen and
134 Lettenmaier 2007; Hoerling and Eischeid 2007; Ray et al. 2008; Rasmussen et al. 2011, 2014;

135 Miller et al. 2011, 2012; Bureau of Reclamation 2012; Vano et al. 2012; Vano and Lettenmaier
136 2014). Much of the water for this region comes from the high-elevation area – the Colorado
137 Headwaters – that act as a natural reservoir during the winter, storing precipitation as snowpack.
138 Hence, we select three basins in the Colorado Headwaters with outlets at streamflow stations
139 managed by the United States Geological Survey (USGS) – Yampa River at Steamboat Springs,
140 East River at Almont and Animas River at Durango – whose location and elevation ranges are
141 shown in Figure 1.

142 Table 1 summarizes the main hydroclimatic characteristics of the three basins for which
143 historical data are available, over an 8-year period (Oct/2000 - Sep/2008). Mean basin
144 precipitation ranges between 700 mm/year to 900 mm/year, while mean basin elevation is above
145 2500 m.a.s.l. Among these basins, the East River at Almont has the largest runoff ratio (0.42),
146 and the Yampa at Steamboat Springs has the lowest runoff ratio (0.32, with the lowest runoff and
147 precipitation amounts). The land surface of the Yampa and Animas River basins is
148 predominantly covered by deciduous forests (26 % at Yampa and 23 % at Animas) and
149 evergreen forests (37 % at Yampa and 39 % at Animas), while the land surface of the East River
150 basin is mainly covered by evergreen forests (29 %) and grassland/herbaceous (26 %).

151 # 3. Data and methods

152 ## *3.1 Climate datasets*

153 We use dynamically downscaled climate datasets obtained with the WRF model to force
154 hydrologic simulations and compute hydrologic changes due to a climate perturbation. These
155 datasets consist of historical (control run, CTRL) and pseudo global warming (PGW) outputs at

three different horizontal resolutions (4-, 12- and 36-km). These WRF simulations are fully described in Rasmussen et al. (2014), but briefly reviewed below. The initial and 3-hourly lateral boundary conditions were taken from the North American Regional Reanalysis (NARR; Mesinger et al. 2006) coarse resolution dataset (~32 km). The model physics options used in that study included the Noah Land Surface Model (Noah-LSM) version 3.2 with upgraded snow physics (Chen and Dudhia, 2001; Barlage et al., 2010), the Thompson mixed-phase cloud microphysics scheme (Thompson et al. 2008), the Yonsei University planetary boundary layer (Hong et al. 2006) and the Community Atmosphere Model's (CAM) longwave and shortwave radiation schemes (Collins et al. 2006). Because the use of a horizontal resolution of 6-km or less is able to accurately estimate vertical motions driven by topography (Ikeda et al. 2010; Rasmussen et al. 2011), a convective parameterization was included for the 12- and 36-km simulations – using the Betts-Miller-Janjić scheme (Janjić 1994)– but not for the 4-km simulation (Rasmussen et al. 2014).

The PGW approach (Schär et al., 1996; Hara et al., 2008; Kawase et al., 2009; Rasmussen et al., 2011) consists of adding monthly mean climate perturbations to the initial and 3-hourly boundary conditions taken from NARR at each WRF vertical level. The climate perturbation was obtained from the Community Climate System Model Version 3 (CCSM3) runs performed by the National Center for Atmospheric Research's (NCAR) Climate and Global Dynamics Division (Collins et al. 2006) under the A1B scenario (Nakicenovic et al. 2000; Meehl et al. 2007). This perturbation is generated by subtracting the current 10-yr (1995-2005) monthly climatology from a future 10-yr (2045-2055) monthly climatology.

Meteorological data from WRF simulations is available for all horizontal resolutions at hourly time steps, for both historical and modified climatic conditions. The 12-km and 36-km

179 grids are perfectly aligned to each other, but not with the high-resolution 4-km grid. The output
180 variables and temporal discretization used depend on specific hydrologic model requirements
181 (Table 2). To compare the effects of WRF horizontal resolution – referred as *experiment 1* – with
182 those from re-scaling high-resolution outputs, we create two additional forcing datasets by re-
183 gridding WRF outputs obtained at 4-km to the 12- and 36-km grid cells used by Rasmussen et al.
184 (2014) – referred as *experiment 2*. This is done in two steps: (1) identification of all the 4-km
185 grid points contained in WRF grid cells at 12- and 36-km resolutions, and (2) computation of the
186 new forcing data from the simple average of the 4-km points contained in the 12- and 36-km grid
187 cells.

188 Figure 2 includes basin-averaged monthly precipitation values computed from: (a) WRF
189 outputs obtained by Rasmussen et al. (2014) with three different horizontal resolutions (4-, 12-
190 and 36-km), and (b) 4-km WRF outputs, re-scaled to 12- and 36-km. These results correspond to
191 the period October/2002 – September/2008, for current (CTRL, dashed lines) and future (PGW,
192 solid lines) climate scenarios. As a reference, observed basin-averaged monthly precipitation
193 values, also displayed with black symbols in Figure 2, were obtained by applying the Thiessen
194 interpolation method (Thiessen 1911) on SNOTEL observations, including 8, 4 and 11 stations
195 for the Yampa, East and Animas River basins, respectively. The results show that a closer match
196 is achieved using high-resolution (i.e. 4-km) WRF outputs compared to 12- and 36-km,
197 especially at the East and Animas River basins. Although 4-km WRF underestimates basin-
198 averaged precipitation at Yampa, especially during late winter and spring, additional analyses
199 (not shown) suggested that a single station – located at the northeastern edge of the basin –
200 recorded much greater precipitation amounts than the others, enhancing the mismatch between
201 observed and simulated precipitation.

Figure 2a shows that WRF horizontal resolution affects precipitation amounts considerably. Indeed, 4-km WRF-CTRL simulations generate basin-averaged annual precipitation amounts of 735 mm, 824 mm and 937 mm in Yampa, East at Animas River basins, respectively, while the 12-km (36-km) WRF-CTRL simulation produces 473 mm (371 mm) at Yampa, 596 mm (580 mm) at East, and 650 mm (611 mm) at the Animas River basin. These relative differences are augmented over summer (June-September), when the 4-km WRF-CTRL outputs are 168 mm, 205 mm and 195 mm for Yampa, East and Animas, respectively, while 12-km (36-km) outputs are 68 mm (39 mm) at Yampa, 95 mm (77 mm) at East, and 111 mm (80 mm) at Animas. Moreover, PGW simulations project an increase in precipitation over fall/winter and a decrease during summer months at all basins, regardless of the horizontal resolution adopted in WRF. By contrast, scaling effects (Figure 2b) on monthly precipitation amounts are minor compared to those coming from WRF resolution, with 12-km (36-km) annual totals of 715 mm (698 mm) at Yampa, 808 mm (759 mm) at East and 934 mm (960 mm) at the Animas River basin. Datasets spatially aggregated at 12-km (36-km) provide accumulated October-March differences of -14 mm (-25 mm) at Yampa, and -17 mm (-61 mm) at East with respect to the 4-km WRF-CTRL simulation, and the 36-km aggregated dataset generated a +10 mm February-April precipitation difference at Animas when compared to the same benchmark.

Figure 3 displays basin-averaged monthly temperatures computed for experiments 1 (Figure 3a) and 2 (Figure 3b). The 4-km WRF historical simulation provided annual mean temperatures of 2.1º C for Yampa, -0.5º C for East and 1.2º C for Animas, while the 12-km (36-km) WRF run resulted in 1.8º C (1.5º C) at Yampa, -1.1º C (-0.4º C) at East and 0.5º C (0.7º C) at Animas These differences can be mostly explained by discrepancies over December-February (DJF). Similar relative differences (i.e. higher DJF temperatures) between the 4-km WRF

225 simulation and the 12-km/36-km runs were found under the PGW climate scenario, with annual
226 increases (i.e. future – current) in basin-averaged mean annual temperature ranging from +2.3º C
227 (Animas River basin) to +2.5º C (Yampa River basin). On the other hand, Figure 3b shows that
228 scaling effects on monthly temperatures were relatively smaller, with mean annual temperatures
229 of 2.2º C (2.1º C) for Yampa, -0.5º C (0.1º C) for East and 1.1º C (0.8º C) for Animas from the
230 re-scaled 12-km (36-km) dataset. Moreover, the two coarse resolution datasets provided almost
231 identical projected changes in mean annual temperature than the original 4-km WRF simulation:
232 +2.4º C in the Yampa and East River basins, and +2.3º C in the Animas River basin. However,
233 there are still differences in basin-averaged monthly temperature values among the three datasets
234 in Experiment 2, which arise from daily temperature discrepancies and are so small that they are
235 not distinguishable at the scale of Figure 3.

236 ### *3.2 Hydrologic/land surface models*

237 To explore the interplay between WRF horizontal resolutions and the choice of hydrologic
238 model structure, we include four hydrologic/land surface models: the US Geological Survey's
239 Precipitation Runoff Modeling System (PRMS; Leavesley et al., 1983; Leavesley and Stannard,
240 1995), the Variable Infiltration Capacity model (VIC; Wood et al., 1992; Liang et al., 1994,
241 1996) the Noah Land Surface Model (Noah-LSM; Ek 2003; Mitchell et al. 2004), and the Noah
242 Land Surface Model with Multiple Parameterizations (Noah-MP; Niu et al. 2011; Yang et al.
243 2011). Our choice builds on the different degrees of complexity spanned by these models in
244 terms of conceptualization of vegetation, soil and seasonal snowpack (see Table 3 for further
245 details), and also different parameterizations for some hydrologic processes (e.g., different model
246 equations for canopy storage, baseflow, etc.). It should be noted that these models are not

ordered by degree of complexity, which can vary depending on the process representation of interest (Table 3). Simulation time steps, forcing variables and land cover data used for a priori parameter estimates vary depending on specific model requirements (see Table 2 for further details).

In this study we use a single suite of physics options for Noah-MP, including a Ball-Berry type model for canopy stomatal resistance, the Community Land Model (CLM; Oleson et al. 2010) soil stress function to control stomatal resistance, the SIMTOP model for runoff and groundwater (Niu et al. 2005), a Monin-Obukhov similarity theory-based drag coefficient, supercooled liquid water and frozen soil permeability based on Niu and Yang (2006), a two-stream radiation transfer scheme applied only to the vegetated fraction, a snow surface albedo parameterization based on the Canadian Land Surface Scheme (CLASS; Verseghy 1991), partitioning of precipitation into snowfall and rainfall based on Jordan (1991) and a Noah-type lower boundary of soil temperature. Readers are referred to Niu et al. (2011) for a full description of each model component.

3.3 Experimental design

3.3.1 *Hydrologic model simulations*

To assess the effects of WRF horizontal resolution and spatial scaling on the portrayal of climate change impacts, we perform offline hydrologic model simulations under historical (CTRL) and modified climate (PGW) scenarios for the following cases:

- Experiment 1: hydrologic models are forced with 4-, 12- and 36-km WRF outputs produced by Rasmussen et al. (2014).

- Experiment 2: hydrologic models are forced with 4-km WRF outputs, and two additional datasets obtained from re-scaling 4-km outputs to 12- and 36-km resolutions.

All model simulations are conducted for the period between October 1, 2000 and September 30, 2008, using the first two years as spin up (not processed) to initialize model states. We compute hydrologic changes using the same parameter values obtained by Mendoza et al. (2015), i.e. calibrated by minimizing the root mean squared error (RMSE) between observed and simulated daily streamflow (period October 1, 2002 to September 30, 2008) with the Shuffled Complex Evolution (SCE-UA) algorithm (Duan et al., 1992, 1993), using 4-km resolution WRF historical datasets. Setting WRF-CTRL output as a historical climate baseline fits with the aim to better understand how different methodological choices affect hydrologic portrayals using the climate datasets obtained by Rasmussen et al. (2014); therefore, tuning hydrologic model parameters to WRF-CTRL allows a direct examination of how the climate change signal (given by the differences between WRF-CTRL and WRF-PGW) propagates into hydrologic changes. Although there are still some biases in the WRF-CTRL output (Figure 2), this is a problem common to all historical forcing datasets (Mizukami et al. 2014), including those based on interpolated station data and empirical algorithms (e.g., Maurer et al. 2002; Livneh et al. 2013), and differences between WRF and such gridded datasets are often attributable to errors in the gridded datasets and not to errors in WRF (Gutmann et al. 2012). It is important that hydrologic model parameters are consistent with the features of meteorological fields used, and the high-resolution WRF simulations are most consistent with both the observations, and with other WRF simulations.

Because the purpose of these experiments is to examine the implications of forcing datasets developed at different spatial resolutions, we fix the grid spacing of hydrological models to 4-

km – identical to the regular grid used in the 4-km WRF simulations performed by Rasmussen et al. (2014)– to isolate effects of forcing scale from those of hydrologic model grid size. Hence, when hydrologic model simulations are forced with 12- and 36-km meteorological datasets, meteorological variables are distributed to hydrologic model grid cells using a nearest neighbor interpolation method as in Shrestha et al. (2006). Finally, hydrologic changes are computed for the period Oct/2002 - Sep/2008 by forcing all hydrologic models with the same current (CTRL) and future (PGW) WRF datasets.

3.3.2 *Process-based evaluation metrics*

We use four hydrologic signature measures (Yilmaz et al. 2008; Stewart et al. 2005) to quantify model performance and projected changes in catchment behavior. These metrics are intended to represent different hydrologic processes – ranging from overall precipitation partitioning into ET and runoff to vertical redistribution of soil moisture –, and they are derived from daily runoff time series. The notation, short description, mathematical formulation and physical process associated with each signature measure are detailed in Table 4. Similar diagnostic evaluation metrics have been used in past studies with multiple purposes, such as model evaluation (e.g., Herbst et al., 2009; Majone et al., 2012; Pfannerstill et al., 2014), catchment classification (e.g., Oudin et al., 2010; Carrillo et al., 2011; Ley et al., 2011; Sawicz et al., 2011), sensitivity analysis (e.g., van Werkhoven et al., 2008; Wagener et al., 2009), hydrologic model structure identification (e.g., Hartmann et al. 2013; Hrachowitz et al. 2014), analysis of spatial distribution of hydrologic processes (e.g., McMillan et al. 2014) and the choice of realistic model parameter values in terms of process representations (e.g., Pokhrel and Gupta, 2009; van Werkhoven et al., 2009; Kollat et al., 2012; Pokhrel et al., 2012).

4. Results and discussion

4.1 Model performance

We first analyze how hydrologic model performance is affected by WRF horizontal resolution (Figure 4a) and forcing re-scaling (Figure 4b) over the period October/2002 - September/2008. To this end, we computed the differences between simulated (control, CTRL) and observed (Obs) values of signature measures of hydrologic behavior (Table 4). In Figure 4, each evaluation metric is displayed in a different row, hydrologic model structures are represented by different symbols, and different colors depict different WRF horizontal resolutions (Figure 4a) or spatial forcing scales (Figure 4b). Therefore, differences between symbols of different colors indicate the magnitudes of effects of RCM horizontal resolution and/or scaling on hydrologic model performance. In Figure 4, the dispersion of red symbols (4-km) indicates inter-model differences in calibrated model performance, which are still considerable when compared to those from uncalibrated model parameters (Mendoza et al. 2015). Finally, the spread provided by the small multi-model ensemble when using a single meteorological forcing dataset is quantified by the sample standard deviation:

$$s_M = \sqrt{\frac{\sum_{i=1}^{N} (M_i - \bar{M})}{N-1}} \qquad (1)$$

where M_i is the metric value (e.g., runoff ratio) obtained by model *i*, *N* is the number of hydrologic model structures ($N = 4$), and $\bar{M}$ is the mean of the metric obtained from the multi-model ensemble.

The results from experiment 1 (Figure 4a) clearly show the impact of WRF horizontal resolution on water balance simulations. Due to the decreased precipitation amounts from

334 coarse-resolution WRF simulations (see 12- and 36-km in Figure 2a), all the hydrologic models
335 produced less runoff in all basins, leading to increased biases (i.e. underestimation) in simulated
336 runoff ratios (RR). When looking at the center of time of runoff (CTR), however, there is no
337 dependence between the selected WRF horizontal resolution and performance, although runoff
338 seasonality is likely to be affected by differences in winter and summer precipitation amounts –
339 together with differences in DJF temperatures controlling the simulation of snowpack processes
340 – between 4-km and 12- or 36-km WRF simulations. Using 12- and 36-km WRF outputs
341 increases inter-model differences when comparing to 4-km WRF simulations for all the
342 signatures (Table 5). For example, the sample standard deviations of errors in simulated runoff
343 seasonality (i.e. CTRL - Obs) – computed from the multi-model ensemble forced with the 4-km
344 WRF output – are 2.3, 5.9 and 3.6 days for Yampa, East, and Animas, respectively; however,
345 when using 12-km (36-km) WRF outputs the spread from the multi-model ensemble increases up
346 to 9.4 (18.5) days at Yampa, 8.0 (8.3) days at East, and 7.0 (7.5) days at Animas. Note that the
347 baseflow processes (FLV) result obtained from VIC and the 36-km WRF for the Yampa River
348 basin (i.e. blue triangle) has been omitted in Figure 4a to allow a better visualization and
349 comparison between experiments 1 and 2 (since CTRL – Obs = 4579.4 log(m^3/s)). With the
350 exception of runoff ratio (RR), the degree of improvement or degradation in hydrologic model
351 performance obtained from switching WRF horizontal resolution depends on the combination of
352 hydrologic model and basin.

353 For the particular case of experiment 2, errors in signature measures reflect how much
354 information is lost when hydrologic model simulations are forced with coarse-scale
355 meteorological fields (Zhao et al. 2009). As shown in Figure 4b, the effects of forcing scaling on
356 model performance exhibit similar patterns to those from WRF horizontal resolution, but

propagating into generally smaller inter-model differences (Table 5) in runoff ratio (RR), flashiness of runoff (FMS), and baseflow processes (FLV). However, the representation of runoff seasonality (CTR) is considerably affected in some combinations of hydrologic model and basin. For instance, errors in simulated CTR at the Yampa River basin span from 5 day with Noah-LSM and 8.2 days with Noah-MP using 4-km WRF outputs, to -4.2 days with Noah-LSM and -5.6 days with Noah-MP when using the 12-km aggregated dataset. Large changes in simulations of CTR are also observed with PRMS, Noah-LSM and Noah-MP at the East River basin. In summary, forcing scaling can translate into similar or even larger inter-model differences in simulations of runoff seasonality. This might be attributed to the smoothing effect when spatially aggregating high precipitation/snowmelt fields from 4-km WRF output, affecting the historical simulation of high daily runoff events and therefore the computation of center of timing (CTR).

Figure 5 illustrates how hydrologic signature measures obtained from coarse resolution datasets (12- and 36-km) differ from those computed using high-resolution WRF outputs under current climate. The results obtained in experiments 1 and 2 are displayed in Figures 5a and 5b, respectively, where each column represents a specific signature measure, and each row contains signature measures computed with 12- and 36-km horizontal resolutions, versus metrics obtained with the 4-km WRF output (i.e. baseline climate dataset). Results from experiment 1 (Figure 5a) demonstrate that while a coarser horizontal resolution propagates into decreased simulated runoff ratios (RR) and a flashier catchment response (FMS) in comparison to the baseline dataset using any hydrologic model, the effects on other metrics depend on the model structure and/or the basin analyzed. For example, 12-km WRF outputs increase the center of time of runoff (CTR) – with respect to the 4-km WRF output – from 229 to 235 days at Yampa, from 222 to 231 days at

East, and from 214 to 227 days at Animas when using PRMS, and decreases the same metric from 227 to 214 days at Yampa, and from 227 to 225 days at East if the model is Noah-MP. Similar effects are observed if hydrologic simulations are forced with 36-km WRF outputs. The implications of WRF horizontal resolution on long-term baseflow (FLV) are basin-dependent for each model with the exception of VIC, for which signature values (all of which have units of log(m^3/s)) increase from 523 at Yampa, 292 at East and 225 at Animas – obtained with 4-km WRF output – to 826 (4793) at Yampa, 432 (458) at East, and 376 (464) when using 12-km (36-km) WRF outputs.

The results displayed in Figure 5b show that the effects of re-scaling forcing datasets are less pronounced than those from WRF horizontal resolution, but still are important for some signature measures. Overall, re-gridding 4-km WRF outputs to coarser horizontal resolutions generates a reduction in simulated runoff ratio (RR) at Yampa (except 36-km with PRMS) and East (except 12-km and Noah-LSM), and shifts to earlier center of time of runoff (CTR) at Yampa with all hydrologic models, and decreases in CTR at East and Animas with all models except VIC. Scaling effects translate into generally small variations in FMS (except at the Yampa River basin when forcing VIC with 12-km and 36-km datasets, and East when forcing VIC with 36-km WRF), with increases or decreases depending on the combination of hydrologic model and basin. For instance, the 12-km (36-km) dataset shows decreases in FMS (all values with units of log(m^3/s)) from 1.21 to 1.17 (1.15) when running PRMS at the Animas River basin, but increases the same metric from 1.23 to 1.28 (1.33) if the hydrologic model is VIC. Further, the 12-km forcing dataset increases FMS from 1.59 to 1.61 at the Yampa River basin, and from 1.35 to 1.37 at the East River basin when hydrologic simulations are performed with PRMS. Finally, scaling effects on long-term baseflow (FLV) are more pronounced when the

403 forcing grid size is 36-km, with increase or decrease in signature values depending on the
404 combination of forcing scale and hydrologic model structure at each basin.

405 ### *4.2 Changes in annual water balance*

406 In this section, we examine and compare the effects of WRF horizontal resolution
407 (experiment 1) and spatial forcing aggregation (experiment 2) on the partitioning of precipitation
408 into ET and runoff under current and future climate scenarios. In each panel of Figure 6, the
409 diagonal lines represent basin-averaged mean annual precipitation for current and future climate
410 scenarios over a 6-year average period (Oct/2002 - Sep/2008). The intersection of these lines
411 with the x-axis indicates where all precipitation becomes runoff, while the intersection with the
412 y-axis indicates where the system converts all precipitation into ET. In each panel, different
413 symbols depict outputs coming from different hydrologic model structures for current climate
414 (unfilled) and future climate (solid) with symbol colors showing the spatial resolutions. A
415 symbol located exactly on the 1:1 lines represents a simulation with negligible changes in storage
416 over the 6-year simulation period (i.e. P = ET + R), whereas symbols located below the 1:1 line
417 imply increases in storage, and those above denote decreases in storage. Inter-model differences
418 in precipitation partitioning are represented by the distance between different symbols (unfilled
419 or solid), while the distance between a particular symbol (e.g., star for Noah-MP) for current
420 (unfilled) and future (solid) climate scenarios represents the hydrologic change signal. The
421 uncertainty arising from model choice, represented by the dispersion of symbols holding the
422 same color along the precipitation (diagonal) line, is quantified in Table 6 by the following
423 metric:

424 $$s_{R,ET} = \sqrt{\sigma_R^2 + \sigma_{ET}^2} \qquad (2)$$

where σ_R and σ_{ET} are the sample standard deviations obtained with equation (1) from the multi-model outputs for mean annual runoff and mean annual ET, respectively.

The results obtained from experiment 1 (Figure 6a and Figure 2a) show that 4-km WRF simulations generate the largest precipitation amounts under current (CTRL) and future (PGW) climate scenarios at all basins, followed by 12- and 36-km WRF outputs. On the other hand, the effects of forcing re-scaling (Figure 6b) on basin-averaged annual precipitation are reduced compared to those of WRF horizontal resolution. Interestingly, forcing re-scaling can increase inter-model differences (i.e. larger dispersion of symbols when 4-km WRF are re-gridded to 12- and 36-km) in precipitation partitioning under current and future climate (Table 6). For example, $s_{R,ET}$ increases from 40.5 mm/year with 4-km WRF-CTRL output, to 45.5 mm/year with the 12-km aggregated dataset in the East River basin. Further, scaling effects on inter-model differences can be larger than those from WRF horizontal resolution. For instance, $s_{R,ET}$ = 41.3 mm/year when using 36-km WRF-CTRL output at the Yampa River basin, but the same metric goes up to 51.7 mm/year with the 36-km aggregated dataset.

To understand how WRF horizontal resolution and forcing re-scaling affect the portrayal of climate change impacts in annual water balance, we compute projected changes in basin-averaged mean annual runoff and ET (Figure 7) for each basin (displayed in different columns). In each panel, the dispersion of the same symbol (e.g., triangle for VIC) holding different colors across the ΔRunoff – ΔET space (with Δ representing the difference between future and current climate scenarios) represents the uncertainty introduced by the choice of WRF horizontal resolution (Figure 7a) or spatial resolution in forcing datasets (Figure 7b). Similarly, the dispersion of different symbols holding the same color (e.g., red for 4-km WRF datasets) illustrates the uncertainty associated with hydrologic model choice. We quantify inter-model

448 differences in projected changes in annual water balance using equation (2), together with inter-
449 forcing differences when a single model structure is applied (Table 7).

450 Figure 7a shows that the choice of WRF horizontal resolution has large effects on
451 hydrologic changes projected through different hydrologic model structures. These effects are
452 reflected in the *magnitude* (i.e. location or distance from each symbol to the point ΔRunoff = 0
453 mm/yr, ΔET = 0 mm/yr) and *direction* (i.e. quadrant in which symbols are located, indicating
454 increase/decrease of mean annual runoff and ET) of projected changes in mean annual runoff and
455 mean annual ET obtained with each hydrologic model. Further, the dispersion provided by
456 different WRF horizontal resolutions – represented by different colors and quantified in Table 7b
457 – through a single model structure –represented by a single symbol in Figure 7a – may be
458 comparable or even larger than that obtained from multiple model structures forced with a
459 unique WRF dataset (Table 7a). For example, inter-forcing differences in projected hydrologic
460 changes at the East River basin from experiment 1 (i.e. different WRF horizontal resolutions) are
461 $s_{\Delta R,\Delta ET}$ = 28.5 mm/year using PRMS, and $s_{\Delta R,\Delta ET}$ = 22.6 mm/year with Noah-MP. Both values are
462 larger than inter-model differences obtained using 12-km WRF outputs, for which $s_{\Delta R,\Delta ET}$ = 16.7
463 mm/year. Further, in most cases inter-forcing differences from experiment 2 (i.e. different
464 forcing scales) are smaller than inter-model differences summarized in Table 7a.

465 In summary, the results presented Figure 7 in Table 7 indicate that the effects of forcing
466 scaling on projected changes in the annual water balance are smaller than those coming from
467 WRF horizontal resolution. Indeed, the direction of hydrologic change is mostly preserved when
468 forcing the same hydrologic model with re-scaled datasets (Figure 7b). In opposition to the
469 results obtained from experiment 1, the uncertainty coming from hydrologic model choice is
470 much larger than the uncertainty from the choice of dataset. Additionally, re-scaling forcing

inputs to coarser resolutions can enhance inter-model differences in hydrologic change shown by the increased dispersion of symbols of the same color and $s_{\Delta R,\Delta ET}$ values (Table 7).

4.3 Projected changes in catchment behavior

Finally, we compare the effects of WRF horizontal resolution and forcing re-scaling on projected changes in hydrologic signature measures across multiple model structures (Figure 8). Again, differences between symbols of different colors indicate the magnitudes of effects of RCM horizontal resolution and/or scaling on projected changes in basin behavior. The results from experiment 1 (Figure 8a) show that the only consistent change obtained with all WRF horizontal resolutions is a decrease in the center of time of runoff (CTR), or earlier annual peak flow, under the future climate scenario, although the magnitudes obtained can be different depending on the hydrologic model selected. These inter-model differences – computed with equation (1) for each forcing dataset – are presented in Table 8. One can note that – as opposed to model performance results (Table 5) – there is not a defined relation between the choice of WRF horizontal resolution, and the spread in projected changes from different hydrologic model structures.

Selecting coarser grid sizes (i.e. 12- and 36-km) with a convective parameterization in WRF generally translates into increased projected changes (i.e. PGW – CTRL) in runoff ratio (RR), changing in some cases the sign (i.e. from negative to positive values) of projected variations. For instance, outputs from the 4-km WRF solution show a decrease in RR of -0.01 with PRMS at the Yampa River basin, but outputs from the 12-km (36-km) WRF solution show a change of +0.02 (+0.03) for the same model/basin. When looking at flashiness of runoff (FMS), the effects of WRF horizontal resolution on projected changes obtained with a specific

493 hydrologic model structure depend on the basin analyzed. For example, the choice of grid
494 spacing has relatively small effects on projected changes in FMS obtained with Noah-LSM at
495 Yampa (+0.27 to +0.36 log(m^3/s)), Noah-MP at East (+0.03 to +0.08 log(m^3/s)), and PRMS at
496 Animas (-0.18 to -0.27 log(m^3/s)), but larger implications in future projections for the rest of
497 models/basins. The results for low flow volumes (FLV) show that WRF horizontal resolution
498 mostly affects the direction and magnitude of projections obtained with VIC, especially in the
499 Yampa River basin, for which projected changes go from +87.1 log(m^3/s) with 4-km WRF
500 output, to -65.3 log(m^3/s) using 36-km WRF datasets. This behavior can be explained by the
501 parameter values found from the calibration of VIC at Yampa, which compensated a good match
502 of high flows at the expense of very poor baseflow simulations during July-September (not
503 shown).

504 According to Figure 8b, while the effects of forcing re-scaling on projected changes in
505 runoff ratio (RR) are smaller than those from WRF horizontal resolution (Figure 8a), they can
506 still change the sign (e.g., VIC simulations at the Yampa River basin) and magnitude (e.g.,
507 PRMS and VIC at East, Noah-MP at Animas) of projections. Although forcing scaling has very
508 minor effects on changes in runoff seasonality (CTR) across all models, it can affect both the
509 magnitude and direction of projections in flashiness of runoff (FMS). For example, projected
510 changes at the East River basin with VIC span from +0.34 log(m^3/s) with 4-km WRF output, to –
511 0.15 log(m^3/s) using 36-km re-scaled datasets; and projected changes in FMS at the Animas
512 River basin with Noah-MP range from -0.04 log(m^3/s) with 4-km WRF output, to +0.02
513 log(m^3/s) using 36-km re-scaled datasets. Finally, scaling effects on projected changes in low
514 flow volumes (FLV) are generally smaller than those from WRF horizontal resolution, and are
515 mostly reflected in VIC simulations. Nevertheless, other hydrologic model structures can also

516 reverse the signal of projected changes of FLV in some basins; for instance, Noah-LSM projects
517 a variation of +33.4 log(m^3/s) at the Animas River basin using 4-km WRF outputs, switching to -
518 31.2 log(m^3/s) if 36-km aggregated datasets are used.

519 # 5. Summary and Conclusions

520 Hydrologists encounter many subjective decisions when configuring hydrologic models for
521 climate change impact studies. One of the choices is how forcing inputs are prepared to drive
522 hydrologic model simulations. Scaling (up-scaling) high-resolution climate model output is
523 typically done to match up with the resolutions of other hydrologic model inputs, such as
524 parameter fields. Furthermore, climate modelers often use different microphysics schemes (e.g.,
525 convection parameterization or explicit solution of physical equations) depending on spatial
526 resolution, affecting precipitation inputs. We investigated the implications that RCM horizontal
527 resolution and re-scaling of RCM outputs may have on the portrayal of climate change impacts.
528 Specifically, we assessed the effects of the above decisions on: (i) historical performance in
529 terms of hydrologic signature measures, and (ii) hydrologic changes due to a climate
530 perturbation, with focus on the annual water balance and catchment processes. The analyses
531 were conducted in three catchments located in the headwaters of the Colorado River basin. To
532 explore the interplay between forcing effects and hydrologic model choice, we include four
533 model structures, whose parameters were calibrated against observed runoff using 4-km WRF
534 historical datasets (Mendoza et al. 2015).

535 As illustrated by Rasmussen et al. (2014), the choice of WRF horizontal resolution (i.e.
536 model grid size and inclusion of convective parameterization) has large effects on simulated
537 precipitation amounts. Specifically, the use of 12- and 36-km resolutions and a convective

538 parameterization results in the underestimation of basin-averaged annual precipitation totals with
539 respect to 4-km WRF simulations. Therefore, it was found that the choice of WRF spatial
540 resolution has larger effects on the historical simulation of hydrologic signature measures in
541 comparison to those provided by up-scaling forcing datasets. However, up-scaling still affects
542 runoff seasonality (CTR) considerably due to the smoothing effects on high spatial variability in
543 precipitation and temperature over the mountainous, changing snowmelt timing. This is in line
544 with the results of Lobligeois et al. (2014), who found that using high-resolution precipitation
545 fields mostly benefited hydrologic simulations in areas with highly heterogeneous rainfall
546 patterns.

547 The water balance analysis revealed that WRF horizontal resolution has tremendous effects
548 on the portrayal of hydrologic change at an annual basis (i.e. variations in mean annual runoff
549 and ET), regardless of the hydrologic model structure selected. Moreover, the effects of WRF
550 horizontal resolution on hydrologic change may overwhelm the uncertainty from model choice,
551 which surpasses the uncertainty from re-scaled forcings. It was also found that re-scaling forcing
552 datasets to coarser resolutions may augment inter-model differences in precipitation partitioning
553 and projected changes in runoff and ET. The same conclusions can be drawn from hydrologic
554 model simulations conducted with uncalibrated parameter values (not shown here).

555 Forcing scaling effects on projected changes in hydrologic signature measures were found to
556 be generally smaller than those coming from WRF horizontal resolution. However, using coarser
557 forcing resolutions may create an artificial switch in the sign of changes projected by a particular
558 hydrologic model structure (e.g., runoff ratio, flashiness of runoff). Even more, we found that
559 scaling effects can exceed those associated with WRF spatial resolution when projecting
560 variations in hydrologic behavior (e.g., flashiness of runoff).

561 This study has shown that the horizontal grid spacing used in RCMs can have important
562 consequences on the magnitude and direction of the hydrologic change signal; however, this
563 specific uncertainty source is still only one part of the entire uncertainty envelope of simulated
564 hydrologic projections (Clark et al. 2016). For example, a relevant component excluded from our
565 analyses is hydrologic parameter uncertainty (e.g., Cameron et al. 1999; Wilby 2005; Steele-
566 Dunne et al. 2008; Surfleet and Tullos 2013; Mendoza et al. 2016). Earlier work in our group and
567 by others has shown that the choice of forcing scale may have large effects on calibrated
568 parameters – especially for precipitation (e.g., Liang et al. 2004; Bárdossy and Das 2008;
569 Tramblay et al. 2011) –, and that forcing generation methods may have substantial implications
570 on hydrological portrayals (Wayand et al. 2013; Mizukami et al. 2014; Elsner et al. 2014),
571 meaning that parameter values could be considerably affected when calibrating hydrologic
572 models against re-scaled datasets. In order to avoid an over-confident portrayal of climate
573 change impacts, future studies should incorporate an integrated characterization and
574 quantification of the different sources of uncertainty in hydrologic modeling, with particular
575 emphasis on meteorological forcings, model structure, and parameters.

576 ## Acknowledgments

577 We thank Marketa Elsner, Andy Newman, Michael Barlage, Roland Viger, Adam Carheden
578 and Tor Mohling for their feedback and assistance for setting up the models. This work was
579 supported through a contract with the U.S. Army Corps of Engineers, through a Cooperative
580 Agreement with the U.S. Bureau Of Reclamation, and through a graduate fellowship from the
581 Cooperative Institute for Research and Environmental Sciences (CIRES). The National Center
582 for Atmospheric Research is sponsored by the National Science Foundation. Finally, we thank

583 two anonymous reviewers for their highly constructive comments, which contributed to improve
584 this manuscript considerably.

585 References

586 Addor, N., O. Rössler, N. Köplin, M. Huss, R. Weingartner, and J. Seibert, 2014: Robust
587 changes and sources of uncertainty in the projected hydrological regimes of Swiss
588 catchments. *Water Resour. Res.*, **50**, 7541–7562, doi:10.1002/2014WR015549.

589 Arnaud, P., C. Bouvier, L. Cisneros, and R. Dominguez, 2002: Influence of rainfall spatial
590 variability on flood prediction. *J. Hydrol.*, **260**, 216–230, doi:10.1016/S0022-
591 1694(01)00611-4.

592 Bae, D.-H., I.-W. Jung, and D. P. Lettenmaier, 2011: Hydrologic uncertainties in climate change
593 from IPCC AR4 GCM simulations of the Chungju Basin, Korea. *J. Hydrol.*, **401**, 90–105,
594 doi:10.1016/j.jhydrol.2011.02.012.

595 Bárdossy, A., and T. Das, 2008: Influence of rainfall observation network on model calibration
596 and application. *Hydrol. Earth Syst. Sci.*, **12**, 77–89, doi:10.5194/hessd-3-3691-2006.

597 Barlage, M., and Coauthors, 2010: Noah land surface model modifications to improve snowpack
598 prediction in the Colorado Rocky Mountains. *J. Geophys. Res.*, **115**, D22101,
599 doi:10.1029/2009JD013470.

600 Bastola, S., C. Murphy, and J. Sweeney, 2011: The role of hydrological modelling uncertainties
601 in climate change impact assessments of Irish river catchments. *Adv. Water Resour.*, **34**,
602 562–576, doi:10.1016/j.advwatres.2011.01.008.

603 Bell, V. A., and R. J. Moore, 2000: The sensitivity of catchment runoff models to rainfall data at
604 different spatial scales. *Hydrol. Earth Syst. Sci.*, **4**, 653–667, doi:10.5194/hess-4-653-2000.

605 Boorman, D., and C. Sefton, 1997: Recognising the uncertainty in the quantification of the
606 effects of climate change on hydrological response. *Clim. Change*, **35**, 415–434.

607 Bureau of Reclamation, 2012: *Colorado River Basin Water Supply and Demand Study:*
608 *Technical Report B – Water Supply Assessment*.

609 Cameron, D., K. Beven, and P. Naden, 1999: Flood frequency estimation by continuous
610 simulation under climate change (with uncertainty). *Hydrol. Earth Syst. Sci.*, **4**, 393–405.

611 Carrillo, G., P. A. Troch, M. Sivapalan, T. Wagener, C. Harman, and K. Sawicz, 2011:
612 Catchment classification: hydrological analysis of catchment behavior through process-
613 based modeling along a climate gradient. *Hydrol. Earth Syst. Sci.*, **15**, 3411–3430,

614 doi:10.5194/hess-15-3411-2011.

615 Chen, F., and J. Dudhia, 2001: Coupling an Advanced Land Surface–Hydrology Model with the
616 Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity.
617 *Mon. Weather Rev.*, **129**, 569–585, doi:10.1175/1520-
618 0493(2001)129<0569:CAALSH>2.0.CO;2.

619 Christensen, N. S., and D. P. Lettenmaier, 2007: A multimodel ensemble approach to assessment
620 of climate change impacts on the hydrology and water resources of the Colorado River
621 Basin. *Hydrol. Earth Syst. Sci.*, **11**, 1417–1434, doi:10.5194/hess-11-1417-2007.

622 Clark, M. P., and Coauthors, 2016: Characterizing Uncertainty of the Hydrologic Impacts of
623 Climate Change. *Curr. Clim. Chang. Reports*, **2**, 55–64, doi:10.1007/s40641-016-0034-x.

624 Collins, W. D., and Coauthors, 2006: The Community Climate System Model Version 3
625 (CCSM3). *J. Clim.*, **19**, 2122–2143, doi:10.1175/JCLI3761.1.

626 Dankers, R., O. B. Christensen, L. Feyen, M. Kalas, and A. de Roo, 2007: Evaluation of very
627 high-resolution climate model data for simulating flood hazards in the Upper Danube Basin.
628 *J. Hydrol.*, **347**, 319–331, doi:10.1016/j.jhydrol.2007.09.055.

629 Ek, M. B., 2003: Implementation of Noah land surface model advances in the National Centers
630 for Environmental Prediction operational mesoscale Eta model. *J. Geophys. Res.*, **108**, 1–
631 16, doi:10.1029/2002JD003296.

632 Elsner, M. M., S. Gangopadhyay, T. Pruitt, L. D. Brekke, N. Mizukami, and M. P. Clark, 2014:
633 How Does the Choice of Distributed Meteorological Data Affect Hydrologic Model
634 Calibration and Streamflow Simulations? *J. Hydrometeorol.*, **15**, 1384–1403,
635 doi:10.1175/JHM-D-13-083.1.

636 Finnerty, B. D., M. B. Smith, D.-J. Seo, V. Koren, and G. E. Moglen, 1997: Space-time scale
637 sensitivity of the Sacramento model to radar-gage precipitation inputs. *J. Hydrol.*, **203**, 21–
638 38, doi:10.1016/S0022-1694(97)00083-8.

639 Fowler, H. J., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to impacts
640 studies: recent advances in downscaling techniques for hydrological modelling. *Int. J.*
641 *Climatol.*, **27**, 1547–1578, doi:10.1002/joc.1556. http://doi.wiley.com/10.1002/joc.1556.

642 Fry, J., and Coauthors, 2011: Completion of the 2006 National Land Cover Database for the
643 Conterminous United States. *Photogramm. Eng. Remote Sens.*, **77**, 858–864.

644 Gao, Y., J. a. Vano, C. Zhu, and D. P. Lettenmaier, 2011: Evaluating climate change over the
645 Colorado River basin using regional climate models. *J. Geophys. Res.*, **116**, D13104,
646 doi:10.1029/2010JD015278.

647 Graham, L. P., J. Andréasson, and B. Carlsson, 2007: Assessing climate change impacts on

hydrology from an ensemble of regional climate models, model scales and linking methods – a case study on the Lule River basin. *Clim. Change*, **81**, 293–307, doi:10.1007/s10584-006-9215-2.

Gutmann, E. D., R. M. Rasmussen, C. Liu, K. Ikeda, D. J. Gochis, M. P. Clark, J. Dudhia, and G. Thompson, 2012: A Comparison of Statistical and Dynamical Downscaling of Winter Precipitation over Complex Terrain. *J. Clim.*, **25**, 262–281, doi:10.1175/2011JCLI4109.1.

Hansen, M. C., R. S. DeFries, J. R. Townshend, and R. Sohlberg, 2000: Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int. J. Remote Sens.*, **21**, 1331–1364.

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, doi:10.3178/hrl.2.61.

Hartmann, A., T. Wagener, A. Rimmer, J. Lange, H. Brielmann, and M. Weiler, 2013: Testing the realism of model structures to identify karst system processes using water quality and quantity signatures. *Water Resour. Res.*, **49**, 3345–3358, doi:10.1002/wrcr.20229.

Herbst, M., H. V. Gupta, and M. C. Casper, 2009: Mapping model behaviour using Self-Organizing Maps. *Hydrol. Earth Syst. Sci.*, **13**, 395–409.

Hoerling, M., and J. Eischeid, 2007: Past peak water in the Southwest. *Southwest Hydrol.*, **6**, 18–19.

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. *Mon. Weather Rev.*, **134**, 2318–2341, doi:10.1175/MWR3199.1.

Hrachowitz, M., and Coauthors, 2014: Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water Resour. Res.*, **50**, 7445–7469, doi:10.1002/2014WR015484.

Ikeda, K., and Coauthors, 2010: Simulation of seasonal snowfall over Colorado. *Atmos. Res.*, **97**, 462–477, doi:10.1016/j.atmosres.2010.04.010.

Janjić, Z., 1994: The Step-Mountain Eta Coordinate Model: Further Developments of the Convection, Viscous Sublayer, and Turbulence Closure Schemes. *Mon. Weather Rev.*, **122**, 927–945.

Jensen, M., D. Rob, and C. Franzoy, 1969: Scheduling irrigations using climate-crop-soil data. *National Conference on Water Resources Engineering of the American Society of Civil Engineers*, New Orleans, 20.

Jiang, T., Y. D. Chen, C. Xu, X. Chen, X. Chen, and V. P. Singh, 2007: Comparison of

682 hydrological impacts of climate change simulated by six hydrological models in the
683 Dongjiang Basin, South China. *J. Hydrol.*, **336**, 316–333,
684 doi:10.1016/j.jhydrol.2007.01.010.

685 Jones, R. N., F. H. S. Chiew, W. C. Boughton, and L. Zhang, 2006: Estimating the sensitivity of
686 mean annual runoff to climate change using selected hydrological models. *Adv. Water*
687 *Resour.*, **29**, 1419–1429, doi:10.1016/j.advwatres.2005.11.001.

688 Jordan, R., 1991: A One-Dimensional Temperature Model for a Snow Cover. *Cold Reg. Res.*
689 *Eng. Lab, U.S. Army Corps Eng. Hanover, N.H.*, Spec. Rept. 91–16.

690 Kawase, H., T. Yoshikane, M. Hara, F. Kimura, T. Yasunari, B. Ailikun, H. Ueda, and T. Inoue,
691 2009: Intermodel variability of future changes in the Baiu rainband estimated by the pseudo
692 global warming downscaling method. *J. Geophys. Res.*, **114**, D24110,
693 doi:10.1029/2009JD011803.

694 Kay, A. L., H. N. Davies, V. A. Bell, and R. G. Jones, 2009: Comparison of uncertainty sources
695 for climate change impacts: flood frequency in England. *Clim. Change*, **92**, 41–63,
696 doi:10.1007/s10584-008-9471-4.

697 Kleinn, J., C. Frei, J. Gurtz, D. Lüthi, P. L. Vidale, and C. Schär, 2005: Hydrologic simulations
698 in the Rhine basin driven by a regional climate model. *J. Geophys. Res. D Atmos.*, **110**, 1–
699 18, doi:10.1029/2004JD005143.

700 Kollat, J. B., P. M. Reed, and T. Wagener, 2012: When are multiobjective calibration trade-offs
701 in hydrologic models meaningful? *Water Resour. Res.*, **48**, 1–19,
702 doi:10.1029/2011WR011534.

703 Koren, V. I., B. D. Finnerty, J. C. Schaake, M. B. Smith, D.-J. Seo, and Q.-Y. Duan, 1999: Scale
704 dependencies of hydrologic models to spatial variability of precipitation. *J. Hydrol.*, **217**,
705 285–302, doi:10.1016/S0022-1694(98)00231-5.

706 Lauer, A., C. Zhang, O. Elison-Timm, Y. Wang, and K. Hamilton, 2013: Downscaling of climate
707 change in the hawaii region using CMIP5 results: On the choice of the forcing fields*. *J.*
708 *Clim.*, **26**, 10006–10030, doi:10.1175/JCLI-D-13-00126.1.

709 Leavesley, G. H., and L. G. Stannard, 1995: The precipitation-runoff modeling system-PRMS.
710 *Computer Models of Watershed Hydrology*, V.P. Singh, Ed., Water Resources Publications,
711 Highlands Ranch, CO, 281–310.

712 Leavesley, G. H., R. W. Lichty, B. M. Troutman, and L. G. Saindon, 1983: *Precipitation-Runoff*
713 *Modeling System: User's Manual*.

714 Ley, R., M. C. Casper, H. Hellebrand, and R. Merz, 2011: Catchment classification by runoff
715 behaviour with self-organizing maps (SOM). *Hydrol. Earth Syst. Sci.*, **15**, 2947–2962,

716 doi:10.5194/hess-15-2947-2011.

717 Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based
718 model of land surface water and energy fluxes for general circulation models. *J. Geophys.*
719 *Res.*, **99**, 14,415.14.428.

720 ——, E. F. Wood, and D. P. Lettenmaier, 1996: Surface soil moisture parameterization of the
721 VIC-2L model: Evaluation and modification. *Glob. Planet. Change*, **13**, 195–206,
722 doi:10.1016/0921-8181(95)00046-1.

723 ——, J. Guo, and L. R. Leung, 2004: Assessment of the effects of spatial resolutions on daily
724 water flux simulations. *J. Hydrol.*, **298**, 287–310, doi:10.1016/j.jhydrol.2003.07.007.

725 Livneh, B., E. A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K. M. Andreadis, E. P. Maurer, and
726 D. P. Lettenmaier, 2013: A long-term hydrologically based dataset of land surface fluxes
727 and states for the conterminous United States: Update and extensions. *J. Clim.*, **26**, 9384–
728 9392, doi:10.1175/JCLI-D-12-00508.1.

729 Lobligeois, F., V. Andréassian, C. Perrin, P. Tabary, and C. Loumagne, 2014: When does higher
730 spatial resolution rainfall information improve streamflow simulation? An evaluation using
731 3620 flood events. *Hydrol. Earth Syst. Sci.*, **18**, 575–594, doi:10.5194/hess-18-575-2014.

732 Ludwig, R., and Coauthors, 2009: The role of hydrological model complexity and uncertainty in
733 climate change impact assessment. *Adv. Geosci.*, **21**, 63–71.

734 Majone, B., C. I. Bovolo, A. Bellin, S. Blenkinsop, and H. J. Fowler, 2012: Modeling the
735 impacts of future climate change on water resources for the Gállego river basin (Spain).
736 *Water Resour. Res.*, **48**, W01512, doi:10.1029/2011WR010985.

737 Maurer, E., A. W. Wood, and D. P. Lettenmaier, 2002: A Long-Term Hydrologically Based
738 Dataset of Land Surface Fluxes and States for the Conterminous United States. *J. Clim.*, **15**,
739 3237–3251.

740 McMillan, H., M. Gueguen, E. Grimon, R. Woods, M. Clark, and D. E. Rupp, 2014: Spatial
741 variability of hydrological processes and model structure diagnostics in a 50 km 2
742 catchment. *Hydrol. Process.*, **28**, 4896–4913, doi:10.1002/hyp.9988.

743 Meehl, G. A., C. Covey, K. E. Taylor, T. Delworth, R. J. Stouffer, M. Latif, B. McAvaney, and
744 J. F. B. Mitchell, 2007: THE WCRP CMIP3 Multimodel Dataset: A New Era in Climate
745 Change Research. *Bull. Am. Meteorol. Soc.*, **88**, 1383–1394, doi:10.1175/BAMS-88-9-1383.

746 Mendoza, P. A., and Coauthors, 2015: Effects of hydrologic model choice and calibration on the
747 portrayal of climate change impacts. *J. Hydrometeorol.*, **16**, 762–780, doi:10.1175/JHM-D-
748 14-0104.1.

749 Mendoza, P. A., M. P. Clark, N. Mizukami, E. D. Gutmann, J. R. Arnold, L. D. Brekke, and B.

750 Rajagopalan, 2016: How do hydrologic modeling decisions affect the portrayal of climate
751 change impacts? *Hydrol. Process.*, **30**, 1071–1095, doi:10.1002/hyp.10684.
752 http://doi.wiley.com/10.1002/hyp.10684.

753 Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Am. Meteorol.*
754 *Soc.*, **87**, 343–360, doi:10.1175/BAMS-87-3-343.

755 Miller, W., R. A. Butler, T. Piechota, J. Prairie, K. Grantz, and G. DeRosa, 2012: Water
756 Management Decisions using Multiple Hydrologic Models Within the San Juan River Basin
757 Under Changing Climate Conditions. *J. Water Resour. Plan. Manag.*, **138**, 412–420,
758 doi:10.1061/(ASCE)WR.1943-5452.0000237.

759 Miller, W. P., T. C. Piechota, S. Gangopadhyay, and T. Pruitt, 2011: Development of streamflow
760 projections under changing climate conditions over Colorado River basin headwaters.
761 *Hydrol. Earth Syst. Sci.*, **15**, 2145–2164, doi:10.5194/hess-15-2145-2011.

762 Milly, P. C. D., K. A. Dunne, and A. V Vecchia, 2005: Global pattern of trends in streamflow
763 and water availability in a changing climate. *Nature*, **438**, 347–350,
764 doi:10.1038/nature04312.

765 Mitchell, K. E., and Coauthors, 2004: The multi-institution North American Land Data
766 Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a
767 continental distributed hydrological modeling system. *J. Geophys. Res.*, **109**, D07S90,
768 doi:10.1029/2003JD003823.

769 Mizukami, N., and Coauthors, 2014: Hydrologic Implications of Different Large-Scale
770 Meteorological Model Forcing Datasets in Mountainous Regions. *J. Hydrometeorol.*, **15**,
771 474–488, doi:10.1175/JHM-D-13-036.1.

772 ——, and Coauthors, 2016: Implications of the Methodological Choices for Hydrologic
773 Portrayals of Climate Change over the Contiguous United States: Statistically Downscaled
774 Forcing Data and Hydrologic Models. *J. Hydrometeorol.*, **17**, 73–98, doi:10.1175/JHM-D-
775 14-0187.1. http://journals.ametsoc.org/doi/abs/10.1175/JHM-D-14-0187.1.

776 Najafi, M. R., H. Moradkhani, and I. W. Jung, 2011: Assessing the uncertainties of hydrologic
777 model selection in climate change impact studies. *Hydrol. Process.*, **25**, 2814–2826,
778 doi:10.1002/hyp.8043.

779 Nakicenovic, N., and Coauthors, 2000: *IPCC Special Report on Emissions Scenarios*. Cambridge
780 Univ. Press, Cambridge, UK, 612 pp.

781 Niu, G., and Z. Yang, 2006: Effects of frozen soil on snowmelt runoff and soil water storage at a
782 continental scale. *J. Hydrometeorol.*, 937–953.

783 Niu, G.-Y., Z.-L. Yang, R. E. Dickinson, and L. E. Gulden, 2005: A simple TOPMODEL-based

784 runoff parameterization (SIMTOP) for use in global climate models. *J. Geophys. Res.*, **110**,
785 D21106, doi:10.1029/2005JD006111.

786 ——, and Coauthors, 2011: The community Noah land surface model with
787 multiparameterization options (Noah-MP): 1. Model description and evaluation with local-
788 scale measurements. *J. Geophys. Res.*, **116**, D12109, doi:10.1029/2010JD015139.

789 Oleson, K. W., and Coauthors, 2010: *Technical Description of version 4.0 of the Community*
790 *Land Model (CLM)*. Boulder, Colorado, USA, 257 pp.

791 Olsson, J., P. Berg, and A. Kawamura, 2015: Impact of RCM Spatial Resolution on the
792 Reproduction of Local, Subdaily Precipitation. *J. Hydrometeorol.*, **16**, 534–547,
793 doi:10.1175/JHM-D-14-0007.1.

794 Oudin, L., A. Kay, V. Andréassian, and C. Perrin, 2010: Are seemingly physically similar
795 catchments truly hydrologically similar? *Water Resour. Res.*, **46**, W11558,
796 doi:10.1029/2009WR008887.

797 Pfannerstill, M., B. Guse, and N. Fohrer, 2014: Smart low flow signature metrics for an
798 improved overall performance evaluation of hydrological models. *J. Hydrol.*, **510**, 447–458,
799 doi:10.1016/j.jhydrol.2013.12.044.

800 Pokhrel, P., and H. V. Gupta, 2009: Regularized calibration of a distributed hydrological model
801 using available information about watershed properties and signature measures. *New*
802 *Approaches to Hydrological Prediction in Data-sparse Regions (Proc. of Symposium HS.2*
803 *at the Joint IAHS and IAH Convention, Hyderabad, India, September 2009)*, 20–25.

804 ——, K. K. Yilmaz, and H. V. Gupta, 2012: Multiple-criteria calibration of a distributed
805 watershed model using spatial regularization and response signatures. *J. Hydrol.*, **418-419**,
806 49–60, doi:10.1016/j.jhydrol.2008.12.004.

807 Poulin, A., F. Brissette, R. Leconte, R. Arsenault, and J.-S. Malo, 2011: Uncertainty of
808 hydrological modelling in climate change impact studies in a Canadian, snow-dominated
809 river basin. *J. Hydrol.*, **409**, 626–636, doi:10.1016/j.jhydrol.2011.08.057.

810 Prein, A. F., G. J. Holland, R. M. Rasmussen, J. Done, K. Ikeda, M. P. Clark, and C. H. Liu,
811 2013: Importance of Regional Climate Model Grid Spacing for the Simulation of Heavy
812 Precipitation in the Colorado Headwaters. *J. Clim.*, **26**, 4848–4857, doi:10.1175/JCLI-D-12-
813 00727.1.

814 Prudhomme, C., and H. Davies, 2009: Assessing uncertainties in climate change impact analyses
815 on the river flow regimes in the UK. Part 2: future climate. *Clim. Change*, **93**, 197–222,
816 doi:10.1007/s10584-008-9461-6.

817 Rasmussen, R., and Coauthors, 2011: High-Resolution Coupled Climate Runoff Simulations of

818 Seasonal Snowfall over Colorado: A Process Study of Current and Warmer Climate. *J.*
819 *Clim.*, **24**, 3015–3048, doi:10.1175/2010JCLI3985.1.

820 ——, and Coauthors, 2014: Climate Change Impacts on the Water Balance of the Colorado
821 Headwaters: High-Resolution Regional Climate Model Simulations. *J. Hydrometeorol.*, **15**,
822 1091–1116, doi:10.1175/JHM-D-13-0118.1.

823 Rasmussen, S. H., J. H. Christensen, M. Drews, D. J. Gochis, and J. C. Refsgaard, 2012: Spatial-
824 Scale Characteristics of Precipitation Simulated by Regional Climate Models and the
825 Implications for Hydrological Modeling. *J. Hydrometeorol.*, **13**, 1817–1835,
826 doi:10.1175/JHM-D-12-07.1. http://journals.ametsoc.org/doi/abs/10.1175/JHM-D-12-07.1.

827 Ray, A., J. Barsugli, K. Averyt, K. Wolter, M. Hoerling, N. Doesken, B. Udall, and R. Webb,
828 2008: *Climate change in Colorado: a synthesis to support water resources management and*
829 *adaptation*. 52 pp.

830 van Roosmalen, L., J. H. Christensen, M. B. Butts, K. H. Jensen, and J. C. Refsgaard, 2010: An
831 intercomparison of regional climate model data for hydrological impact studies in Denmark.
832 *J. Hydrol.*, **380**, 406–419, doi:10.1016/j.jhydrol.2009.11.014.

833 Sawicz, K., T. Wagener, M. Sivapalan, P. A. Troch, and G. Carrillo, 2011: Catchment
834 classification: empirical analysis of hydrologic similarity based on catchment function in
835 the eastern USA. *Hydrol. Earth Syst. Sci.*, **15**, 2895–2911, doi:10.5194/hess-15-2895-2011.

836 Schär, C., C. Frei, D. Lüthi, and H. C. Davies, 1996: Surrogate climate-change scenarios for
837 regional climate models. *Geophys. Res. Lett.*, **23**, 669–672, doi:10.1029/96GL00265.

838 Shrestha, R., Y. Tachikawa, and K. Takara, 2006: Input data resolution analysis for distributed
839 hydrological modeling. *J. Hydrol.*, **319**, 36–50, doi:10.1016/j.jhydrol.2005.04.025.

840 ——, ——, and ——, 2007: Selection of scale for distributed hydrological modelling in
841 ungauged basins. *Predictions in Ungauged Basins: PUB Kick-off (Proceedings of the PUB*
842 *Kick-off meeting held in Brasilia, 20-22 November 2002)*, 290–297.

843 Skamarock, W. C., and Coauthors, 2008: *A Description of the Advanced Research WRF Version*
844 *3*. Boulder, Colorado, USA,.

845 Steele-Dunne, S., P. Lynch, R. McGrath, T. Semmler, S. Wang, J. Hanafin, and P. Nolan, 2008:
846 The impacts of climate change on hydrology in Ireland. *J. Hydrol.*, **356**, 28–45,
847 doi:10.1016/j.jhydrol.2008.03.025.

848 Stewart, I. T., D. R. Cayan, and M. D. Dettinger, 2005: Changes toward earlier streamflow
849 timing across western North America. *J. Clim.*, **18**, 1136–1155, doi:10.1175/JCLI3321.1.

850 Suklitsch, M., A. Gobiet, A. Leuprecht, and C. Frei, 2008: High Resolution Sensitivity Studies
851 with the Regional Climate Model CCLM in the Alpine Region. *Meteorol. Zeitschrift*, **17**,

852 467–476, doi:10.1127/0941-2948/2008/0308.

853 Surfleet, C. G., and D. Tullos, 2013: Uncertainty in hydrologic modelling for estimating
854 hydrologic response due to climate change (Santiam River, Oregon). *Hydrol. Process.*, **27**,
855 3560–3576, doi:10.1002/hyp.9485.

856 ——, D. Tullos, H. Chang, and I.-W. Jung, 2012: Selection of hydrologic modeling approaches
857 for climate change assessment: A comparison of model scale and structures. *J. Hydrol.*,
858 **464-465**, 233–248, doi:10.1016/j.jhydrol.2012.07.012.

859 Teutschbein, C., and J. Seibert, 2010: Regional climate models for hydrological impact studies at
860 the catchment scale: A review of recent modeling strategies. *Geogr. Compass*, **4**, 834–860,
861 doi:10.1111/j.1749-8198.2010.00357.x.

862 Thiessen, A. H., 1911: Precipitation Averages for Large Areas. *Mon. Weather Rev.*, **39**, 1082–
863 1089, doi:10.1175/1520-0493(1911)39<1082b:PAFLA>2.0.CO;2.

864 Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit Forecasts of Winter
865 Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a
866 New Snow Parameterization. *Mon. Weather Rev.*, **136**, 5095–5115,
867 doi:10.1175/2008MWR2387.1.

868 Tramblay, Y., C. Bouvier, P.-A. Ayral, and A. Marchandise, 2011: Impact of rainfall spatial
869 distribution on rainfall-runoff modelling efficiency and initial soil moisture conditions
870 estimation. *Nat. Hazards Earth Syst. Sci.*, **11**, 157–170, doi:10.5194/nhess-11-157-2011.

871 USDA (U.S. Department of Agriculture), 1992: *Forest Land Distribution Data for the United*
872 *States*.

873 ——, 1994: *State Soil Geographic (STATSGO) Database – Data Use Information*. Misc. Pub.
874 107 pp.

875 Vano, J. A., and D. P. Lettenmaier, 2014: A sensitivity-based approach to evaluating future
876 changes in Colorado River discharge. *Clim. Change*, **122**, 621–634, doi:10.1007/s10584-
877 013-1023-x.

878 ——, T. Das, and D. P. Lettenmaier, 2012: Hydrologic Sensitivities of Colorado River Runoff to
879 Changes in Precipitation and Temperature. *J. Hydrometeorol.*, **13**, 932–949,
880 doi:10.1175/JHM-D-11-069.1.

881 Velázquez, J. A., and Coauthors, 2013: An ensemble approach to assess hydrological models'
882 contribution to uncertainties in the analysis of climate change impact on water resources.
883 *Hydrol. Earth Syst. Sci.*, **17**, 565–578, doi:10.5194/hess-17-565-2013.

884 Verseghy, D., 1991: CLASS—A Canadian land surface scheme for GCMs. I. Soil model. *Int. J.*
885 *Climatol.*, **11**, 111–133.

886 Vicuña, S., R. D. Garreaud, and J. McPhee, 2011: Climate change impacts on the hydrology of a
887 snowmelt driven basin in semiarid Chile. *Clim. Change*, **105**, 469–488,
888 doi:10.1007/s10584-010-9888-4.

889 Wagener, T., K. van Werkhoven, P. Reed, and Y. Tang, 2009: Multiobjective sensitivity analysis
890 to understand the information content in streamflow observations for distributed watershed
891 modeling. *Water Resour. Res.*, **45**, W02501, doi:10.1029/2008WR007347.

892 Wayand, N. E., A. F. Hamlet, M. Hughes, S. I. Feld, and J. D. Lundquist, 2013: Intercomparison
893 of Meteorological Forcing Data from Empirical and Mesoscale Model Sources in the North
894 Fork American River Basin in Northern Sierra Nevada, California. *J. Hydrometeorol.*, **14**,
895 677–699, doi:10.1175/JHM-D-12-0102.1.

896 van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang, 2008: Characterization of watershed
897 model behavior across a hydroclimatic gradient. *Water Resour. Res.*, **44**, 1–16,
898 doi:10.1029/2007WR006271.

899 ——, ——, ——, and ——, 2009: Sensitivity-guided reduction of parametric dimensionality for
900 multi-objective calibration of watershed models. *Adv. Water Resour.*, **32**, 1154–1169,
901 doi:10.1016/j.advwatres.2009.03.002.

902 Wi, S., F. Dominguez, M. Durcik, J. Valdes, H. F. Diaz, and C. L. Castro, 2012: Climate change
903 projection of snowfall in the Colorado River Basin using dynamical downscaling. *Water*
904 *Resour. Res.*, **48**, W05504, doi:10.1029/2011WR010674.

905 Wilby, R. L., 2005: Uncertainty in water resource model parameters used for climate change
906 impact assessment. *Hydrol. Process.*, **19**, 3201–3219, doi:10.1002/hyp.5819.

907 Wood, A., L. Leung, V. Sridhar, and D. Lettenmaier, 2004: Hydrologic implications of
908 dynamical and statistical approaches to downscaling climate model outputs. *Clim. Change*,
909 **62**, 189–216.

910 Wood, E. F., D. P. Lettenmaier, and V. G. Zartarian, 1992: A Land-Surface Hydrology
911 Parameterization With Subgrid Variability for General Circulation Models. *J. Geophys.*
912 *Res.*, **97**, 2717–2728, doi:10.1029/91JD01786.

913 Xu, C., 1999: From GCMs to river flow: a review of downscaling methods and hydrologic
914 modelling approaches. *Prog. Phys. Geogr.*, **23**, 229–249,
915 doi:10.1177/030913339902300204.

916 Yang, Z.-L., and Coauthors, 2011: The community Noah land surface model with
917 multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *J.*
918 *Geophys. Res.*, **116**, 1–16, doi:10.1029/2010JD015140.

919 Yilmaz, K. K., H. V. Gupta, and T. Wagener, 2008: A process-based diagnostic approach to

920 model evaluation: Application to the NWS distributed hydrologic model. *Water Resour.*
921 *Res.*, **44**, W09417, doi:10.1029/2007WR006716.

922 Zhao, G., G. Hörmann, N. Fohrer, and J. Gao, 2009: Impacts of spatial data resolution on
923 simulated discharge, a case study of Xitiaoxi catchment in South China. *Adv. Geosci.*, **21**,
924 131–137.

925

List of Figures

Table 1: Three study watersheds' characteristics. Hydrologic variables correspond to the period Oct/2000-Sep/2008. P, R, PE, RE and DI denote basin-averaged mean annual values of precipitation, runoff, potential evapotranspiration, runoff efficiency (R/P) and dryness index (PE/P), respectively.

Location	USGS ID	Area (km^2)	Mean basin elevation (m.a.s.l.)	Mean annual runoff (mm/yr)	Mean Precipitation from WRF (mm/yr)	Mean annual PE* (mm/yr)	Mean annual RE (R/P)	Mean annual DI (PE/P)
Yampa at Steamboat Springs	09239500	1468	2674	228	717	953	0.32	1.33
East at Almont	09112500	748	3127	327	782	757	0.42	0.97
Animas at Durango	09361500	1819	3098	365	883	885	0.41	1.00

*PE obtained from PRMS by using a Jensen-Haise formulation (Jensen et al. 1969)

971 Table 2: Summary of data sources and simulation setup used in this study

Model	Vegetation data	Soil data	Forcing variables*	Spatial/temporal discretization
PRMS	USGS 1-km gridded vegetation type and density data (USDA 1992)	State soils geographic (STATSGO) 1-km gridded soils data (USDA 1994)	Daily precipitation; maximum and minimum daily temperature.	4 km and Δt = 24 h
VIC	UMD 1-km Global Land Cover Classification (Hansen et al. 2000)	State soils geographic (STATSGO) 1-km gridded soils data (USDA 1994)	Precipitation, temperature, shortwave and longwave radiation, wind speed, relative humidity and air pressure.	4 km and Δt = 1 h
Noah-LSM and Noah-MP	National Land Cover Data Base, 2006 (Fry et al. 2011).	State soils geographic (STATSGO) 1-km gridded soils data (USDA 1994)	Precipitation, temperature, shortwave and longwave radiation, wind speed, relative humidity and air pressure.	4 km and Δt = 1 h

972 *Air temperature at 2 m and wind speed at 10 m are used for hydrologic simulations.

973

974 Table 3: Overview of hydrologic model components used in this study

Model	Snow accumulation and melt	Canopy storage	Moisture in the soil column/surface runoff	Baseflow
PRMS	2-layer energy/mass balance model. Snowpack energy balance is computed every 12 hours.	Precipitation can be intercepted by and evaporated from the plant canopy. Precipitation that is not intercepted by the canopy layer (throughfall) is distributed to the watershed land surface. Interception of precipitation by the plant canopy is computed during a time step as a function of plant-cover density and the storage available on the predominant plant-cover type in each HRU.	Surface runoff and infiltration are computed using a non-linear variable-source-area method allowing for cascade flow.	The groundwater zone is conceptualized as a linear reservoir (ie. baseflow is computed as a linear function of groundwater storage).
VIC	2-layer energy/mass balance model.	Water enters 1-layer canopy reservoir, and can leave as canopy evaporation, transpiration or throughfall. Canopy throughfall occurs when additional precipitation exceeds the storage capacity of the canopy. Different vegetation classes are allowed within a unique grid cell via a 'mosaic' approach, where energy and water balance terms are computed independently for each coverage class (vegetation and bare soil).	An infiltration capacity function is defined. Vertical movement of moisture through soil follows 1-D Richards equation.	Defined as a function of the soil moisture in the third layer (Arno formulation). The function is linear below a soil moisture threshold, and becomes nonlinear above that threshold.
Noah-LSM	1-layer energy/mass balance model that simulates snow accumulation, sublimation, melting and heat exchange at snow-atmosphere and snow-soil interfaces.	One canopy layer, simple canopy resistance. Simple Jarvis type of canopy resistance function, single linearized energy balance equation representing combined ground/vegetation surface, considering seasonal LAI and green vegetation fraction.	Surface runoff is computed as the difference between throughfall and a maximum infiltration rate. Vertical movement of moisture through soil layers follows 1-D Richards equation.	Computed as the product of a scaling factor between 0 and 1 and the hydraulic conductivity of the bottom layer.
Noah-MP	3-layer energy/mass balance model that represents percolation, retention and refreezing of meltwater within the snowpack.	Snow interception includes loading/unloading, melt/refreeze capabilities, and sublimation of canopy-intercepted snow, along with detailed representation of transmission and attenuation of radiation through the canopy, within- and below-canopy turbulence, and different options to represent the biophysical controls on transpiration.	Surface runoff is an exponential function of depth to water table. Vertical movement of moisture through soil layers follows 1-D Richards equation.	Baseflow is parameterized as an exponential decaying function of the water table level (SIMTOP).

975 Table 4: Signature measures used to evaluate model performance and projected changes in
976 catchment behavior

Notation	Short description	Equation	Hydrologic process
RR	Runoff Ratio	$RR = R / P$	Overall water balance (ET processes).
CTR	Center Time of Runoff	$CTR = \frac{\sum_{i=1}^{N} t_i Q_i}{\sum_{i=1}^{N} Q_i}$	Seasonality of runoff.
FMS	FDC Mid-segment Slope	$FMS = \frac{\log(Q_{m1}) - \log(Q_{m2})}{m_1 - m_2}$	Variability, or flashiness, of the flow magnitudes.
FLV	FDC Low-segment Volume	$FLV = \sum_{l=1}^{L} [\log(Q_l) - \log(Q_L)]$	Measure of the long-term baseflow processes.

977 R: basin-averaged mean annual runoff.
978 P: basin-averaged mean annual precipitation.
979 Q_{m1}: flow with exceedance probability of $m_1 = 0.2$.
980 Q_{m2}: flow with exceedance probability of $m_2 = 0.7$.
981 $l = 1,2,...,L$ is the index into the array of flow values located within the low-flow segment (0.7-
982 1.0 exceedance probabilities), being L the index for minimum flow.
983 N: total number of days in a water year.

984

985

986 Table 5: Sample standard deviation in the error of simulated signature measures (CTRL - Obs)
987 computed from hydrologic multi-model ensemble outputs obtained with different forcing
988 datasets. RR, CTR, FMS and FLV denote runoff ratio, center of time of runoff, flow duration
989 curve mid-segment slope, and flow duration curve low-flow volumes, respectively.

Signature	Yampa		East		Animas	
	Experiment 1	Experiment 2	Experiment 1	Experiment 2	Experiment 1	Experiment 2
RR						
4 km	0.02		0.03		0.04	
12 km	0.06	0.03	0.03	0.04	0.06	0.02
36 km	0.06	0.04	0.07	0.01	0.08	0.01
CTR (days since Oct. 1)						
4 km	2.32		5.96		3.60	
12 km	9.37	4.25	7.99	7.88	6.96	4.35
36 km	18.54	6.29	8.31	10.46	7.50	7.22
FMS ($\log(m^3/s)$)						
4 km	0.65		0.31		0.07	
12 km	1.45	0.92	0.58	0.41	0.14	0.05
36 km	1.81	1.50	0.43	0.55	0.24	0.08
FLV ($\log(m^3/s)$)						
4 km	172.54		73.81		47.44	
12 km	296.98	267.63	122.76	78.59	100.06	47.87
36 km	2238.45	285.34	145.06	87.10	122.86	46.62

990

991

992 Table 6: Sample standard deviation (mm/year) in the partitioning of mean annual precipitation
993 into mean annual runoff and mean annual ET, computed from hydrologic multi-model ensemble
994 outputs obtained with different forcing datasets. Larger numbers indicate that a large spread in
995 precipitation partitioning is introduced by the choice of hydrologic model.

Forcing data	Yampa		East		Animas	
	Experiment 1	Experiment 2	Experiment 1	Experiment 2	Experiment 1	Experiment 2
Current climate (CTRL)						
4 km	26.4		40.7		55.4	
12 km	47.8	33.0	37.4	45.5	65.2	36.2
36 km	41.3	51.7	59.1	20.1	77.2	23.4
Future climate (PGW)						
4 km	48.5		31.5		53.0	
12 km	63.6	57.7	47.2	32.4	75.6	37.1
36 km	50.0	76.4	85.3	44.5	85.1	21.0

996

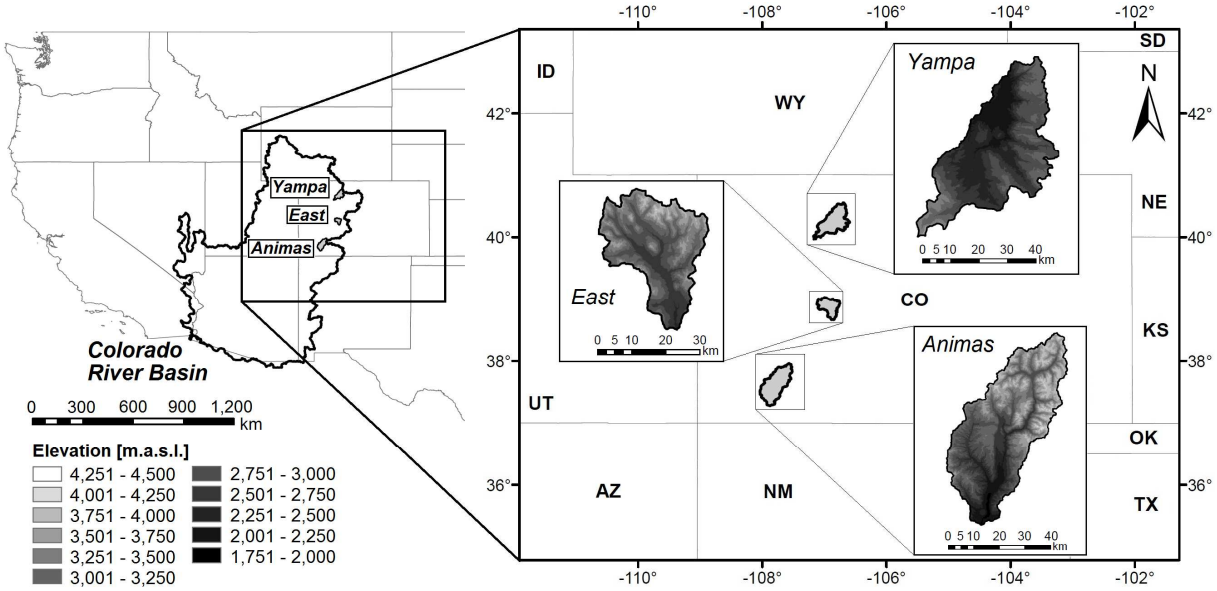
997 Table 7: Sample standard deviation (mm/year) in projected changes in mean annual runoff and
998 mean annual ET, computed from (a) multiple hydrologic model structures and a single forcing
999 dataset, and (b) multiple forcing datasets and a single hydrologic model structure. Larger
1000 numbers indicate that a large spread in projected changes is introduced by the choice of
1001 hydrologic model or forcing dataset.

Experiment	Yampa		East		Animas	
	Experiment 1	Experiment 2	Experiment 1	Experiment 2	Experiment 1	Experiment 2
(a) Fixed forcing dataset						
4 km	23.9		23.3		8.8	
12 km	20.5	26.5	16.7	25.1	15.2	12.6
36 km	12.9	30.5	31.5	27.5	12.8	17.8
(b) Fixed hydrologic model						
PRMS	12.4	1.2	28.5	7.4	27.4	10.5
VIC	10.0	4.8	23.1	8.2	30.2	5.1
Noah-LSM	15.9	6.1	14.9	2.3	20.2	5.1
Noah-MP	20.3	1.9	22.6	5.0	26.6	9.1

1002

Table 8: Sample standard deviation in projected changes (PGW - CTRL) in signature measures of catchment behavior, computed from hydrologic multi-model ensemble outputs obtained with different forcing datasets. RR, CTR, FMS and FLV denote runoff ratio, center of time of runoff, flow duration curve mid-segment slope, and flow duration curve low-flow volumes, respectively.

Signature	Yampa		East		Animas	
	Experiment 1	Experiment 2	Experiment 1	Experiment 2	Experiment 1	Experiment 2
RR						
4 km	0.02		0.02		0.01	
12 km	0.03	0.03	0.02	0.02	0.02	0.01
36 km	0.02	0.03	0.03	0.02	0.01	0.01
CTR (days since Oct. 1)						
4 km	4.10		2.90		3.25	
12 km	2.29	4.54	1.73	2.71	3.29	3.11
36 km	1.65	4.20	3.38	2.95	2.01	2.28
FMS ($\log(m^3/s)$)						
4 km	0.24		0.19		0.11	
12 km	0.17	0.22	0.17	0.19	0.23	0.10
36 km	0.12	0.44	0.17	0.19	0.16	0.08
FLV ($\log(m^3/s)$)						
4 km	268.65		21.40		50.33	
12 km	150.66	170.56	53.93	31.58	45.50	54.55
36 km	401.64	85.76	101.33	45.96	63.04	47.15



1008 Figure 1: Location of the basins of interest: Yampa at Steamboat Springs (1468 km^2), East at
1009 Almont (748 km^2), and Animas at Durango (1819 km^2).

1010

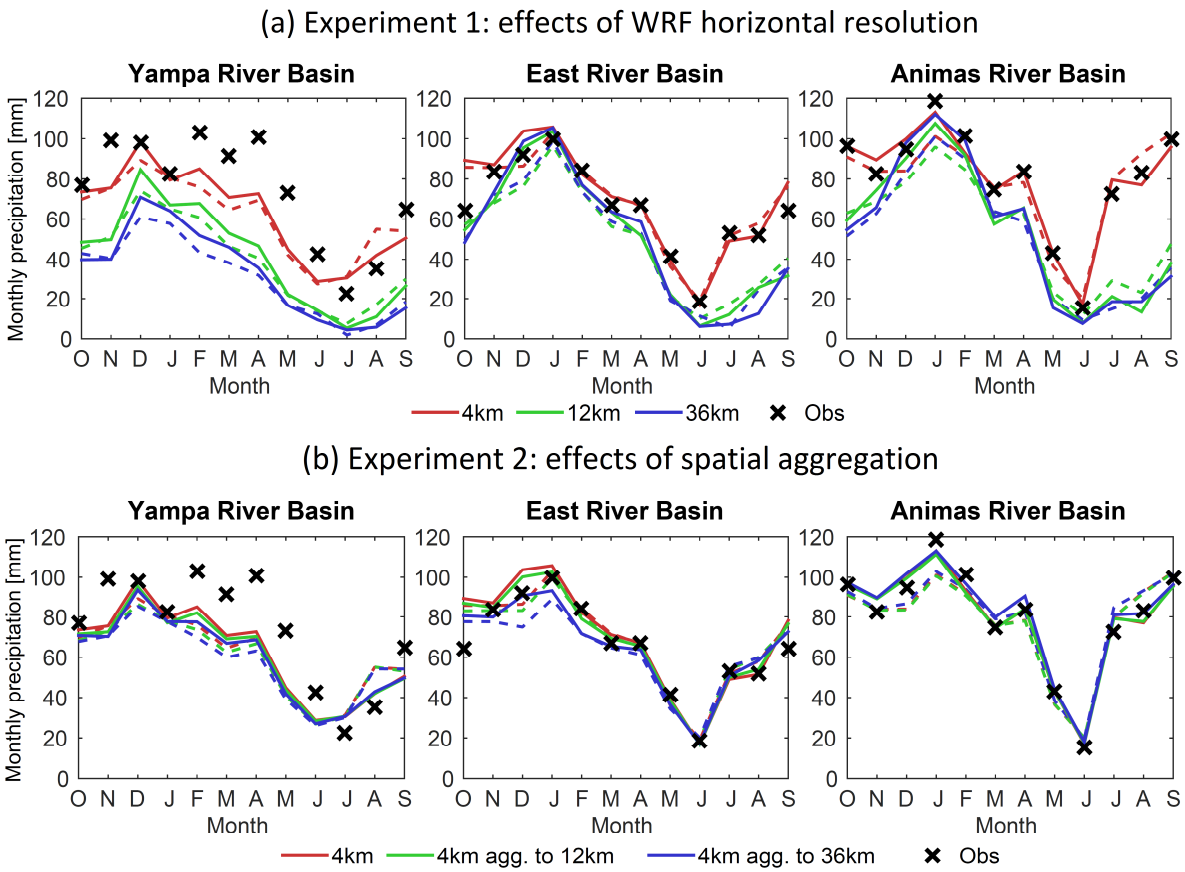


Figure 2: Basin-averaged observed (x symbols) and simulated (colored lines) monthly precipitation values for current (CTRL, dashed lines) and future (PGW, solid lines) WRF outputs used in (a) experiment 1 (effects of WRF horizontal resolution) and (b) experiment 2 (effects of spatial aggregation), for period Oct/2002 - Sep/2008. Basin-averaged observed precipitation was estimated using neighboring SNOTEL stations, whose weights were assigned using the Thiessen interpolation method (Thiessen 1911).

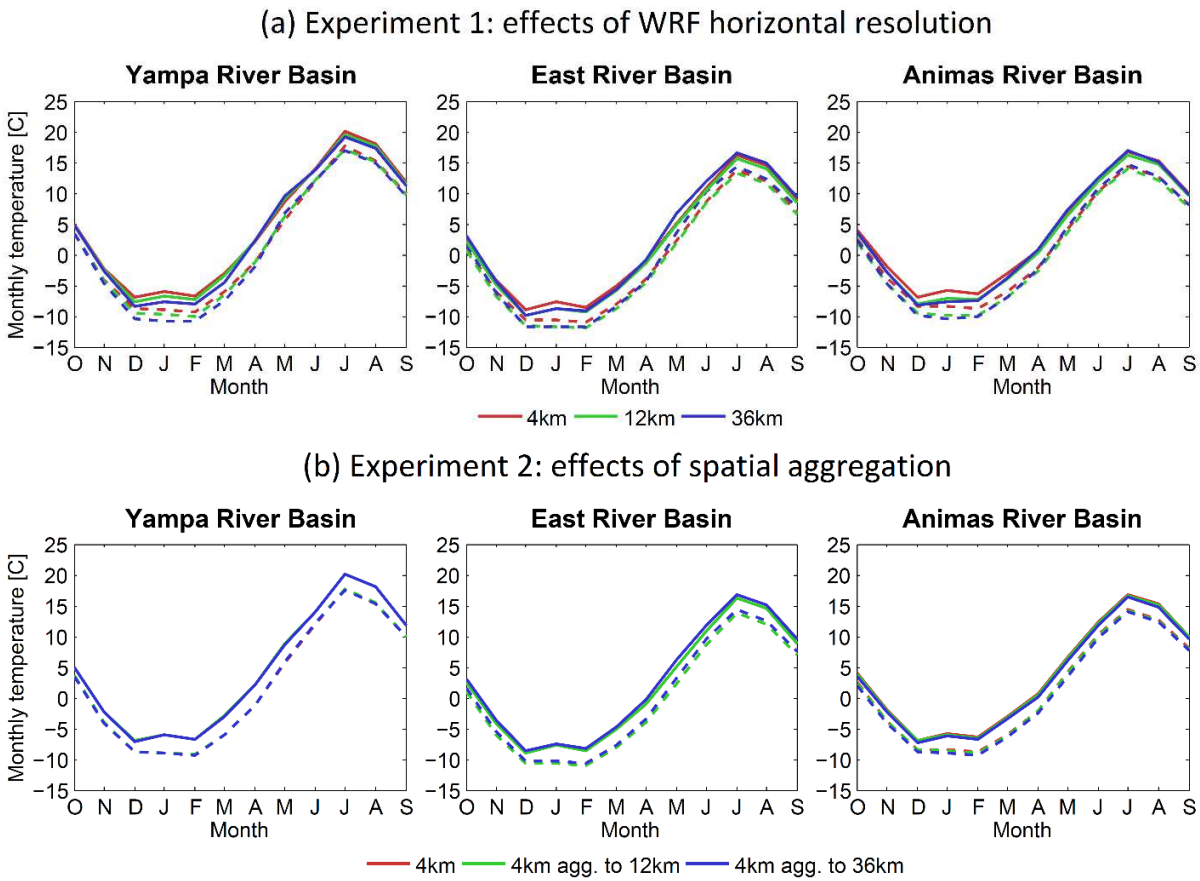


Figure 3: Same as in Figure 2, but for basin-averaged monthly temperature.

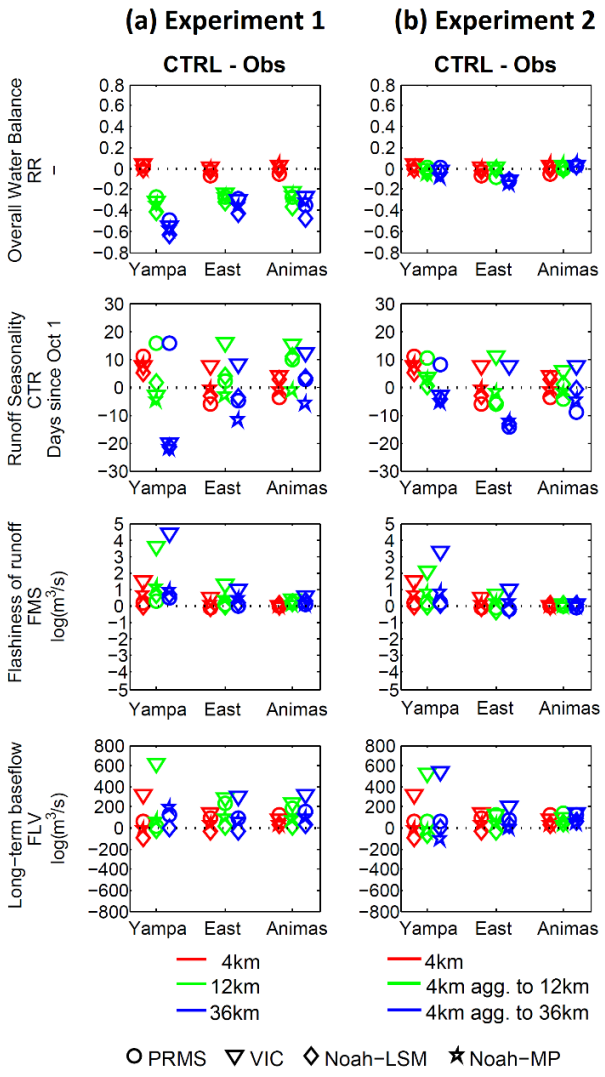
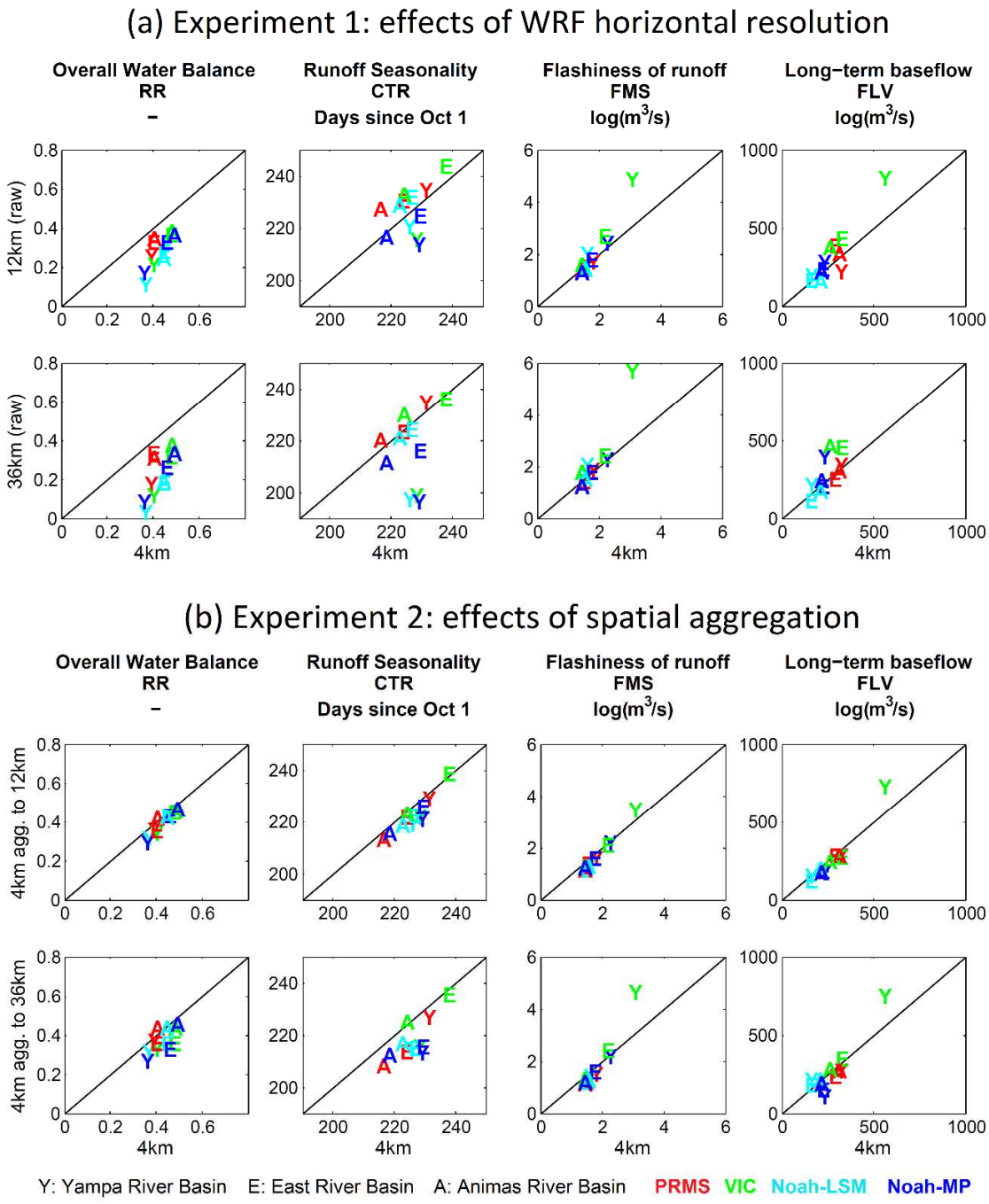


Figure 4: Difference between simulated (CTRL) and observed (Obs) signature measures of hydrologic behavior (period Oct/2002 - Sep/2008) obtained from various hydrologic model structures (i.e. different symbols) and forcing datasets (i.e. different colors). Results are displayed for (a) experiment 1 (effects of WRF horizontal resolution) and (b) experiment 2 (effects of spatial aggregation).



1028

1029 Figure 5: Impact of (a) WRF horizontal resolution (experiment 1) and (b) spatial aggregation of
1030 WRF 4-km resolution datasets on simulated hydrologic signature measures. Each column
1031 contains results for a specific metric, while different rows contain outputs from 12-km and 36-
1032 km (y axis) versus model outputs using WRF datasets with 4 km horizontal grid space (x axis).
1033 In each panel, different letters represent basins and different colors depict results from various
1034 hydrologic models (see legend for details).

1035

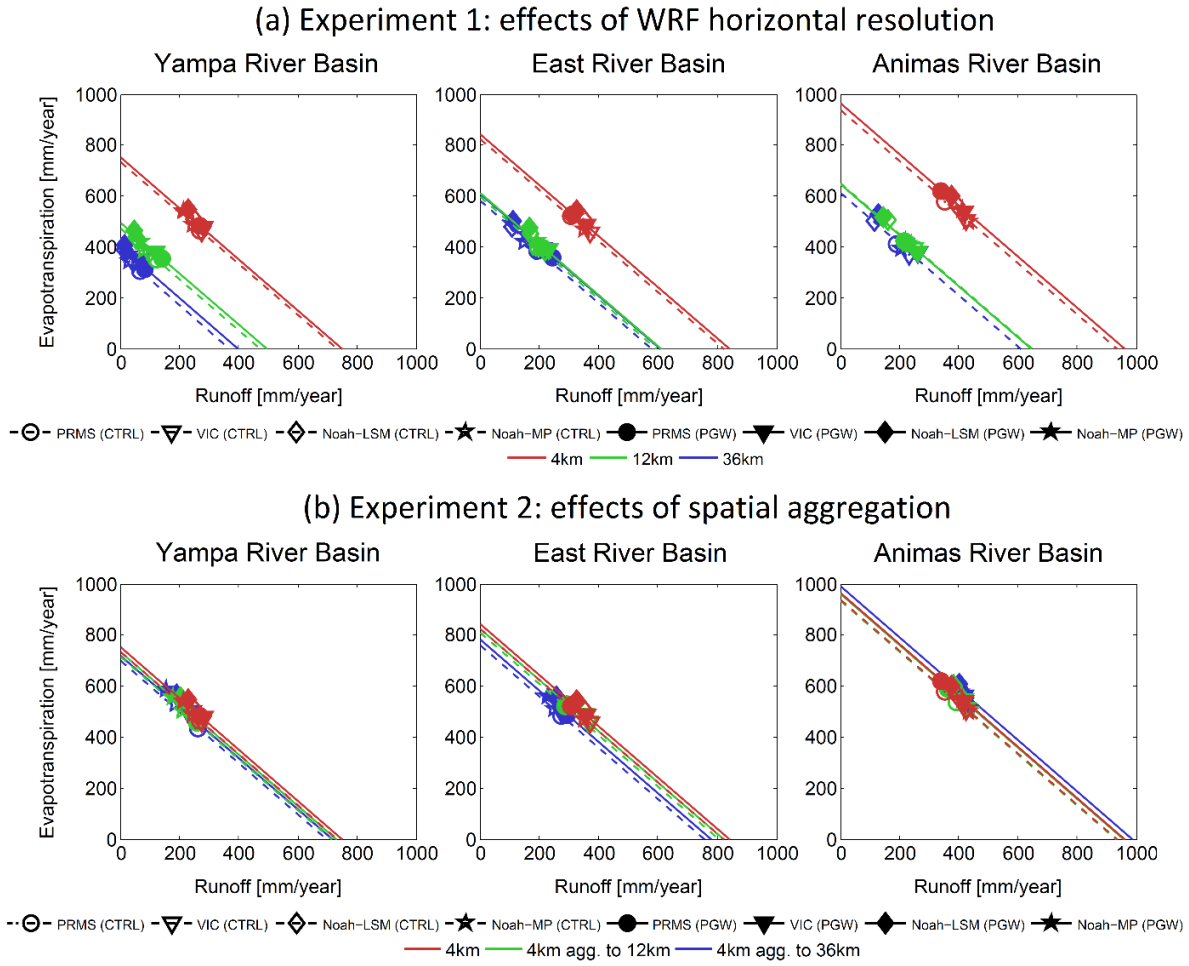
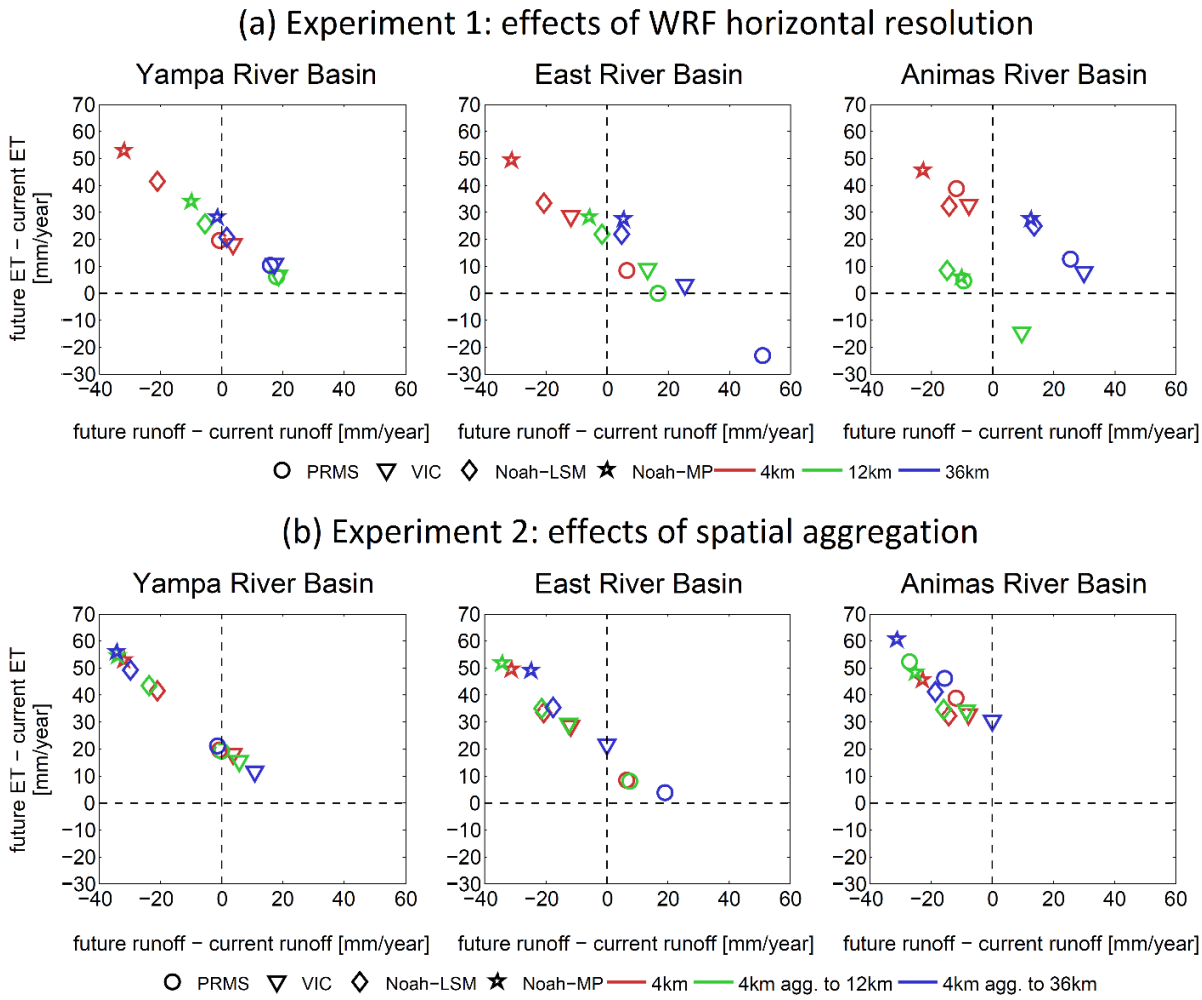


Figure 6: Partitioning of current (CTRL) and future (PGW) basin-averaged mean annual precipitation (diagonal, mm/year) into basin-averaged mean annual runoff (x axis, mm/year) and evapotranspiration (y axis, mm/year) obtained from various model structures (i.e. different symbols) and forcing datasets (i.e. different colors) for the period Oct/2002 - Sep/2008. Results are displayed for (a) experiment 1 (effects of WRF horizontal resolution) and (b) experiment 2 (effects of spatial forcing aggregation).



1045

1046 Figure 7: Projected changes in basin-averaged mean annual runoff (x axis, mm/year) and
1047 evapotranspiration (y axis, mm/year) obtained from various model structures (i.e. different
1048 symbols) and forcing datasets (i.e. different colors) for the period Oct/2002 - Sep/2008. Results
1049 are displayed for (a) experiment 1 (effects of WRF horizontal resolution) and (b) experiment 2
1050 (effects of spatial forcing aggregation).

1051

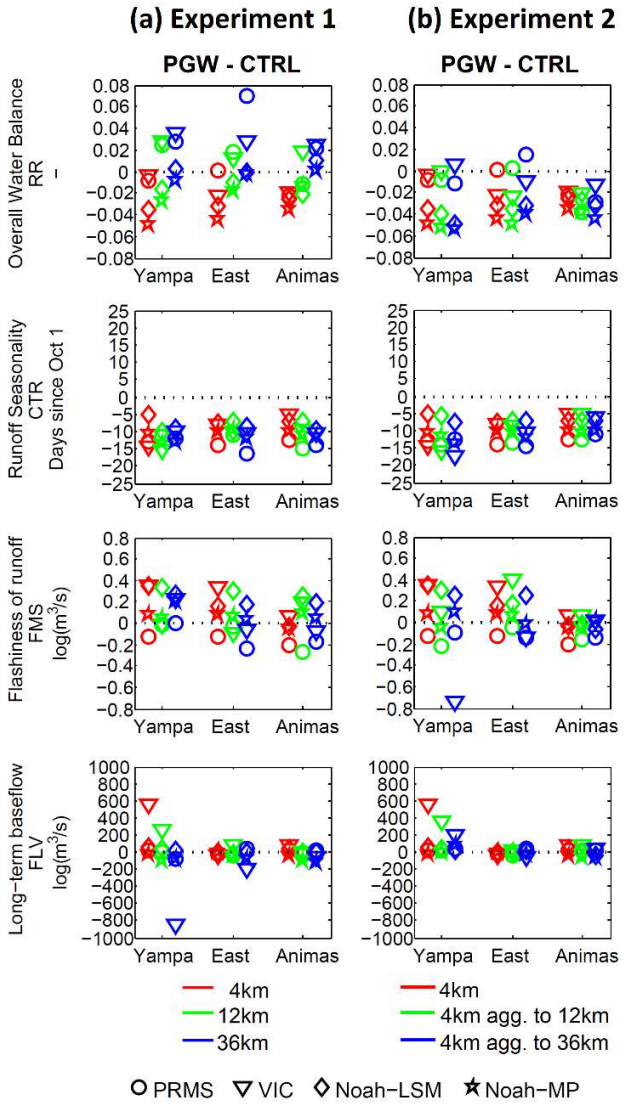


Figure 8: Difference between future (PGW) and current (CTRL) simulated signature measures of hydrologic behavior, obtained from various hydrologic model structures (i.e. different symbols) and forcing datasets (i.e. different colors) over a six-average water year (Oct/2002 - Sep/2008). Results are displayed for (a) experiment 1 (effects of WRF horizontal resolution) and (b) experiment 2 (effects of spatial forcing aggregation).