Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Hydrology 282 (2003) 56–75
Journal of Hydrology
www.elsevier.com/locate/jhydrol

Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States

L.E. Hay[a,*], M.P. Clark[b]

[a]*Water Resources Division, US Geological Survey, Denver, CO, USA*
[b]*Center for Science, Technology and Policy Research, University of Colorado, Boulder, CO, USA*

Abstract

This paper examines the hydrologic model performance in three snowmelt-dominated basins in the western United States to dynamically- and statistically downscaled output from the National Centers for Environmental Prediction/National Center for Atmospheric Research Reanalysis (NCEP). Runoff produced using a distributed hydrologic model is compared using daily precipitation and maximum and minimum temperature timeseries derived from the following sources: (1) NCEP output (horizontal grid spacing of approximately 210 km); (2) dynamically downscaled (DDS) NCEP output using a Regional Climate Model (RegCM2, horizontal grid spacing of approximately 52 km); (3) statistically downscaled (SDS) NCEP output; (4) spatially averaged measured data used to calibrate the hydrologic model (Best-Sta) and (5) spatially averaged measured data derived from stations located within the area of the RegCM2 model output used for each basin, but excluding Best-Sta set (All-Sta).

In all three basins the SDS-based simulations of daily runoff were as good as runoff produced using the Best-Sta timeseries. The NCEP, DDS, and All-Sta timeseries were able to capture the gross aspects of the seasonal cycles of precipitation and temperature. However, in all three basins, the NCEP-, DDS-, and All-Sta-based simulations of runoff showed little skill on a daily basis. When the precipitation and temperature biases were corrected in the NCEP, DDS, and All-Sta timeseries, the accuracy of the daily runoff simulations improved dramatically, but, with the exception of the bias-corrected All-Sta data set, these simulations were never as accurate as the SDS-based simulations. This need for a bias correction may be somewhat troubling, but in the case of the large station-timeseries (All-Sta), the bias correction did indeed 'correct' for the change in scale. It is unknown if bias corrections to model output will be valid in a future climate. Future work is warranted to identify the causes for (and removal of) systematic biases in DDS simulations, and improve DDS simulations of daily variability in local climate. Until then, SDS based simulations of runoff appear to be the safer downscaling choice.
Published by Elsevier B.V.

Keywords: Statistical downscaling; Dynamical downscaling; NCEP/NCAR reanalysis; Hydrologic modelling

1. Introduction

The climate of most of the western United States is characterized by semiarid to arid conditions with

* Corresponding author.
E-mail address: lhay@usgs.gov (L.E. Hay).

0022-1694/$ - see front matter Published by Elsevier B.V.
doi:10.1016/S0022-1694(03)00252-X

the highest precipitation totals occurring in the mountainous regions. Between 50 and 70% of the annual precipitation in these areas falls as snow (Serreze et al., 1999). Melt of the snowpack in the spring provides most of the surface water for the western United States. Water in the western United States is highly allocated between diverse user groups and is subject to mounting pressure by the growing population and changes in institutional practices (Pulwarty, 1995; Diaz and Anderson, 1995). These user pressures, coupled with the effects of climatic variability and potential climatic change, have stimulated research efforts to develop better water management tools.

One potential management tool is the use of atmospheric model output from a global-scale forecast model such as the National Centers for Environmental Prediction/National Center for Atmospheric Research Reanalysis (NCEP) in hydrologic models. Given the large systematic biases and the poor skill present in NCEP precipitation and temperature estimates in some regions, it is necessary to explore methods that may improve upon these global-scale models. Techniques in widespread use are regional climate modeling (Dynamical DownScaling—DDS) and statistical post-processing (Statistical DownScaling—SDS) of NCEP output (e.g. Wilks, 1995; Wilby et al., 2000; Antolik, 2000).

DDS techniques use Regional Climate Models nested within a global-scale model. Regional Climate Models are run at finer horizontal resolution than the global-scale models, and thus provide a more accurate depiction of important model components such as terrain height and cloud physics. However, Regional Climate Models suffer from similar bias problems as the global-scale models (see Takle et al. (1999) and Hay et al. (2002)), and are overly demanding on current computer resources.

SDS techniques develop empirical relations between features reliably simulated in global-scale models at grid-box scales (e.g. 500 hPa geopotential height) and surface predictands at sub-grid scales (e.g. precipitation occurrence and amounts). An advanced SDS system entered in the 1996–1997 National Collegiate Weather Forecasting Contest finished better than approximately 97% of the human forecasters who entered the contest (Vislocky and Fritsch, 1997). The disadvantage of SDS is that the SDS equations must be developed using an archive of forecasts from the same model that is used in the operational setting. SDS is ultimately limited by the assumption of temporal stationarity in the empirical relations (i.e. skillful SDS results for the present climate do not necessarily translate to skillful forecasts of future climate).

The non-stationarity in empirical climate relations is well documented (e.g. Ramage, 1983). DDS does not suffer from the non-stationarity shortcomings present in SDS techniques. Though some parameterization in a Regional Climate Model may have an empirical basis, DDS simulations of local climate are more physically based than SDS and thus are more acceptably transferable from current to future climates. However, DDS simulations of current climate have not been extensively tested (Takle et al., 1999). There is a strong need for a systematic assessment of current Regional Climate Model output in order to evaluate the skill of (and confidence in) Regional Climate Model simulations, especially as drivers for impact assessment models, and to identify areas for model improvement.

Wilby et al. (2000) examined the hydrological response in the Animas River basin of Colorado to DDS and SDS output from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis. They found that in terms of modeling hydrology, both statistical and dynamical downscaling provided greater skill than the coarse-resolution data used to drive the downscaling. The output from the Regional Climate Model used in the dynamical downscaling was simulated by RegCM2 (Giorgi et al., 1988), using the continental US domain and a grid spacing of 52 km. Despite the higher level of sophistication and physical realism associated with DDS, hydrographs simulated using DDS precipitation and temperature were not generally as accurate as those simulated using SDS precipitation and temperature.

This paper will compare the hydrologic model results using precipitation and temperature timeseries derived from: (1) NCEP output (horizontal grid spacing of approximately 210 km); (2) dynamically downscaled (DDS) NCEP output using a Regional

Table 1
Study basins

		Study basin		
		Animas River at Durango	East Fork of the Carson River near Gardnerville	Cle Elum river near Roslyn
State		Colorado	California/Nevada	Washington
Gauging station ID		09361500	10309000	12479000
Drainage area (km^2)		1792	922	526
Elevation range (m)		2000–3700	1600–3000	680–1800
Number of HRUs		121	96	124
Number of stations (excluding Best-Sta) within RegCM2-buffered area shown in Fig. 2 (All-Sta)	Precipitation	38	37	27
	Temperature	30	21	14
Number of RegCM2 gridpoints within RegCm2-buffered area shown in Fig. 2		8	7	5
Number of NCEP gridpoints within NCEP-buffered area shown in Fig. 2		25	24	26
Best 3-station sets (Best-Sta)	Precipitation	Durango, Cascade, Lizard, head Pass	Twin lakes hagan's meadow lobdell	Fish lakes stampede stevens pass
	Temperature	Durango vallecito dam rico	Tahoe valley twin lakes blue lakes	Baring cle elum stampede

Climate Model (RegCM2, horizontal grid spacing of approximately 52 km); (3) statistically downscaled (SDS) NCEP output; and (4) spatially averaged measured data used to calibrate the hydrologic model (Best-Sta). A final data set was introduced in order to provide a fair means of comparing the relative performance of downscaled and station-based runoff simulations. This set consisted of spatially averaged measured data derived from stations located within the area of the RegCM2 model output used for each basin, but excluding Best-Sta set (All-Sta). The All-Sta timeseries are comparable in scale to the DDS model resolution and provide an appropriate test to determine if output at this scale can be used for simulation of basin-scale hydrology.

Three mountainous basins were chosen for this analysis: (1) Animas River at Durango, Colorado (Animas); (2) East Fork of the Carson River near Gardnerville, Nevada (Carson); and (3) Cle Elum River near Roslyn, Washington (Cle Elum). The surface hydrology of these basins is dominated by snowmelt. The Carson and Cle Elum basins are also characterized by frequent rain-on-snow events in the winter months. Tables 1 and 2 list some of the defining features of each basin, and Fig. 1 shows the location of each. In this study, for each of the three basins, daily precipitation and temperature are derived from statistically and dynamically downscaled NCEP output and used as inputs to a hydrologic model. Since the hydrological response of a basin is an integration of the regional weather conditions (in time and space), the results presented here will provide comparison of the overall realism of statistically and dynamically downscaled precipitation and temperature timeseries for three mountainous basins in the United States.

2. Data

For each basin, the following daily data/output were compiled for the purpose of hydrologic modeling: (a) measured-station data; (b) NCEP/NCAR reanalysis (NCEP); and (c) statistically (SDS) and (d) dynamically (DDS) downscaled NCEP.

2.1. Station data

In order to assess the performance of the NCEP-, SDS-, and DDS-based simulations of runoff an appropriate baseline was developed using measured station data. Our hydrologic modeling strategy

Table 2
Elevation ranges

		Elevations (m) ranges for each study basin								
		Animas River at Durango			East Fork of the Carson River near Gardnerville			Cle Elum River near Roslyn		
		Minimum	Mean	Maximum	Minimum	Mean	Maximum	Minimum	Mean	Maximum
HRUs		2011	3060	3728	1645	2300	2959	680	1337	1799
Best 3-station (Best-Sta)	Precipitation	2010	2609	3109	2438	2560	2804	1027	1162	1241
	Temperature	2010	2341	2682	1906	2260	2438	235	678	1219
All Stations (All-Sta)	Precipitation	1720	2686	3536	718	1909	2804	52	813	1829
	Temperature	1720	2328	3536	718	1104	2804	52	328	1640
RegCM2 gridpoints		1895	2579	2987	1586	1926	2166	279	802	1401
NCEP gridpoints		1168	1928	2480	25	1194	2024	28	880	1632

consists of selecting the station set that provides the best simulation of runoff (Best-Sta), and then tune a small select group of model parameters to provide the best possible simulation of runoff given the chosen station-set timeseries (see Hay et al., 2003). The SDS output is developed based on the Best-Sta timeseries for each basin. No such calibration is performed for the NCEP or DDS model inputs. Thus, use of the calibrated Best-Sta mean timeseries to assess NCEP and DDS based runoff simulations will lead to conclusions that are favorable to the station-based simulations and unfavorable to the NCEP and DDS based simulations. Tables 1 and 2 summarize some of the characteristics of the Best-Sta timeseries for each basin.

To provide a fair means of comparing the relative performance of downscaled and station-based runoff simulations, input timeseries for each basin consisting of regionally averaged station measurements were developed. These timeseries (hereafter referred to as 'All-Sta') include data on precipitation and temperature for all stations that fell within the DDS model output domain (RegCM2) for each basin (see Section 2.4 and Fig. 2), but excludes the station set used for the hydrologic model calibration (Best-Sta). The Best-Sta set is used in this study to: (1) provide the best possible set of hydrologic model parameters (used in every hydrologic simulation); (2) train the equations used for SDS; and (3) correct for bias in the NCEP, DDS, and All-Sta timeseries. The NCEP, DDS, and All-Sta timeseries are corrected for systematic bias to distinguish errors in hydrologic simulations associated with model biases from errors in hydrologic simulations associated with model problems in capturing daily meteorological variations. Tables 1 and 2 summarizes the All-Sta timeseries for each basin. All input timeseries are described in Table 3.

Daily maximum and minimum temperatures and precipitation data from stations in and around each basin were compiled from the National Weather Service (NWS) and Snow Telemetry (SNOTEL) data bases. The NWS data were retrieved from the Utah Climate Center's Weather Data Online (http://climate.usu.edu/Free/). SNOTEL data were retrieved from the Natural Resources Conservation Service (ftp://162.79.124.23/data/snow/snotel/snothist/).

2.2. NCEP/NCAR reanalysis

The NCEP/NCAR Reanalysis produced a retroactive 51-year (1948–1998) record of global atmospheric fields derived from a Numerical Weather Prediction model kept unchanged over the analysis period and constrained by observations. Data assimilated into the model consist primarily of free atmospheric variables such as upper-air temperature, pressure heights, and humidity from rawinsondes, pibals, dropsondes, and satellite retrievals. Output fields represent analyses (e.g. 500 hPa heights), which are strongly influenced by the data assimilated, as well as forecast surface variables such as precipitation, soil

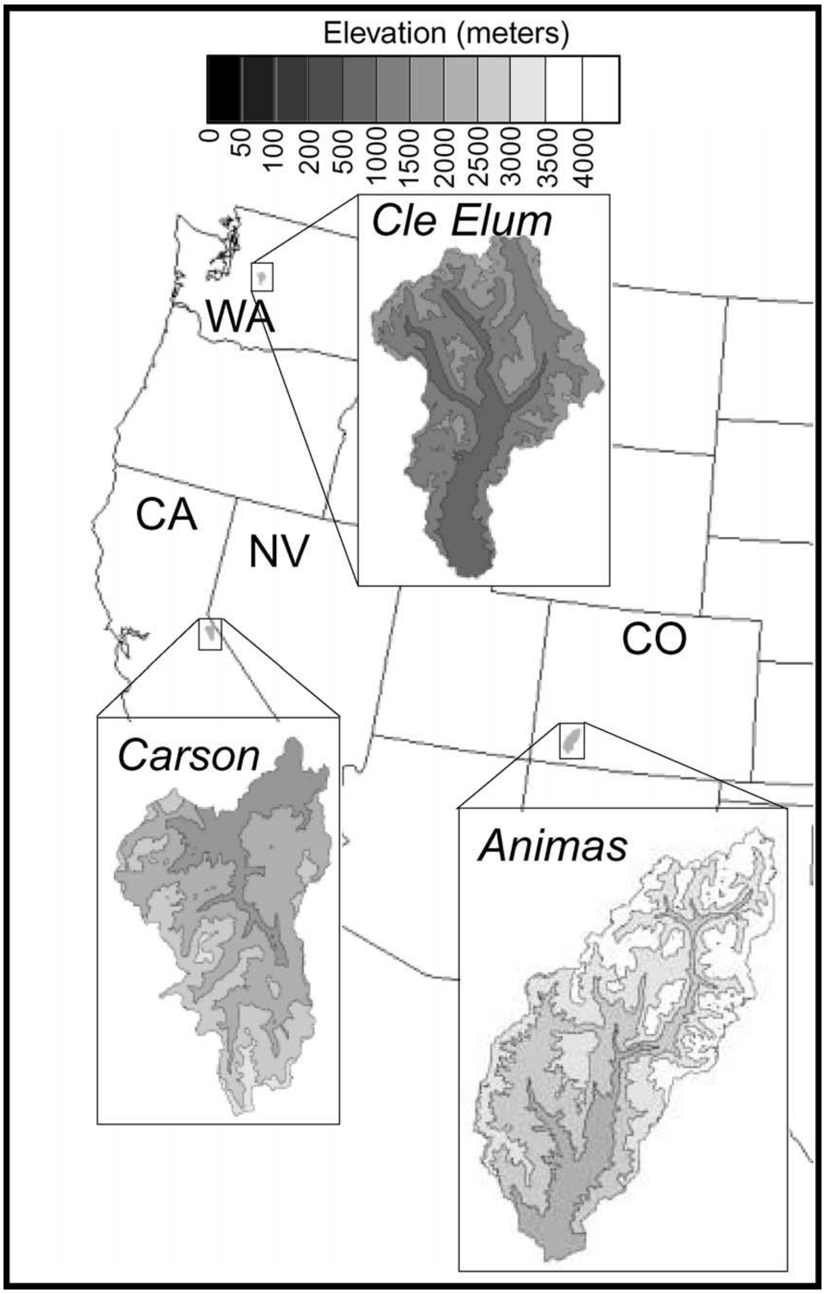

Fig. 1. Location of study basins.

moisture, evaporation, and radiative fluxes. Use of a 'frozen' model in the Reanalysis eliminates pseudo-climate jumps in archived timeseries associated with frequent upgrades in the operational modeling system used at NCEP, and allows an assessment and correction of systematic problems in the model. The model used for the reanalysis is identical to the Medium Range Forecast model implemented operationally at NCEP in January 1995 (Basist and Chelliah, 1997), except that the horizontal resolution

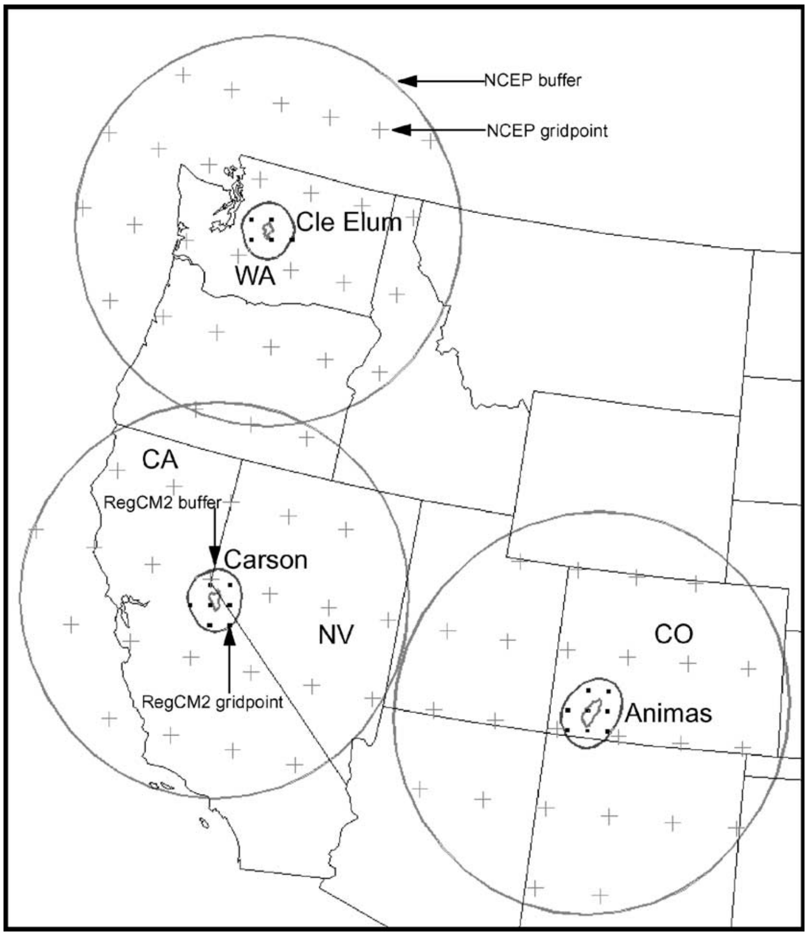

Fig. 2. NCEP and RegCM2 gridpoints used in each basin.

is twice as coarse in the reanalysis version. The model employs a horizontal grid spacing of approximately 210 km, with 28 vertical sigma levels.

Daily NCEP precipitation and temperature time-series were calculated by extracting NCEP output from within a 500 km buffered area for each basin (see Fig. 2) and interpolating to the mean of the Best-Sta location (average of three stations) using Cressman (inverse-distance) interpolation. The search radius of 500 km is somewhat arbitrary. Due to the coarse horizontal resolution of the NCEP output (2.5°), a search radius of at least 200 km is generally required to ensure a basin is completely surrounded by NCEP gridpoints. We use the larger 500 km radius to compensate for problems with horizontal moisture diffusion in the NCEP Reanalysis model.

Table 3
Meteorological inputs (precipitation and maximum and minimum temperature) to the hydrologic model

#	Abbreviation	Description
1	Best-Sta	Best 3-station set
2	All-Sta	All stations within RegCM2-buffered area shown in Fig. 2 but excluding ‘Best-Sta’ stations
3	NCEP	All gridpoints within NCEP-buffered area shown in Fig. 2
4	DDS	Dynamically downscaled NCEP output (using RegCM2). All RegCM2 gridpoints within RegCM2-buffered area shown in Fig. 2
5	SDS	Statistically downscaled NCEP output
6	Bias-All	‘All-Sta’ stations with a bias correction applied
7	Bias-NCEP	‘NCEP’ output with a bias correction applied
8	Bias-DDS	‘DDS’ with a bias correction applied

As discussed by Kistler et al. (2001), the formulation of horizontal moisture diffusion in the reanalysis model causes moisture convergence, leading to isolated 'bulls eyes' of high precipitation in some gridpoints. The 500 km search radius provides a spatial smoothing of the NCEP precipitation field, and more realistic interpolated timeseries in each of the basins examined in this study. Table 2 summarizes some of the characteristics of the NCEP gridpoints chosen for each basin. Note for each basin, maximum and minimum temperature were computed from average temperature based on the monthly diurnal temperature range calculated from the Best-Sta timeseries for each basin.

2.3. Statistical downscaling

In the SDS technique, atmospheric variables included in the NCEP forecast archive were used as predictors in a multiple linear regression approach to forecast precipitation occurrence, precipitation amounts, maximum temperature, and minimum temperature for the three basins in this study. To provide a fairly complete description of forecasted atmospheric conditions, a large pool of potential predictor variables (approximately 350 variables) was examined. Predictor variables include geopotential height, temperature, wind, and humidity at five pressure levels (300, 500, 700, 850, and 1000 hPa), various surface flux variables (e.g. downwelling shortwave radiation flux, 24 h accumulated precipitation), and computed variables such as vorticity advection, zonal and meridional moisture fluxes, and stability indices. All predictor variables are taken from within a 500 km buffered area for each basin (see Fig. 2) and interpolated to the mean of the Best-Sta location (average of three stations) using Cressman (inverse-distance) interpolation. This buffered area was identical to that used to extract the NCEP precipitation and temperature timeseries described above.

The SDS equations were developed using multiple linear regression with forward selection (Antolik, 2000). The forward selection procedure first identifies the predictor variable (e.g. 500 hPa height) which explains the most variance of the predictand (e.g. maximum temperature at a point location). It then searches through the remaining variables, and selects the variable that most reduces the remaining unexplained variance in combination with the variable already chosen. If the improvement in explained variance exceeds a given threshold (taken here as 1%), the variable is included in the multiple linear regression equation. The remaining variables are examined in the same way until no further improvement is obtained based on the correlation threshold.

To assess skill, the downscaling operation is conducted via cross validation (Michaelsen, 1987). For analysis of each month, the data for a year was withheld. All calculations were performed without it, and the downscaling was evaluated with the reserved data. This process was repeated for each year in the observed record. For example, the NCEP output and observed values for 1985 are held out. A downscaled NCEP value for 1985 is found based on the downscaling relations found from all other years. We then hold out the NCEP output and observed values for 1986, and compute 1986 values for downscaling relationships for all other years. By this approach, the downscaled NCEP value for any given year is independent of the observed value for that year.

A final step in the downscaling procedure is stochastic modeling of the residuals in the multiple linear regression equations to provide an assessment of model uncertainty and permit the generation of probabilistic forecasts. For maximum and minimum temperature, this is achieved by extracting a random number from a normal gaussian distribution (mean of zero and standard deviation of one), multiplying the random number by the standard deviation of the regression residuals, and adding this product to the forecast of temperature. For precipitation, we first determine precipitation occurrence. A random number is drawn from a uniformly distributed distribution ranging from zero to one. If the random number is lower than the forecasted probability of precipitation occurrence, the day is classified as a precipitation day. Precipitation amounts are only forecasted for precipitation days. After forecasting precipitation amounts, residuals are modeled stochastically using methods identical to those used for maximum and minimum temperature, and then the forecasted (normally distributed) precipitation amounts are transformed back to the original gamma-type distribution of observed precipitation using the non-parametric probability transform techniques described above. The stochastic modeling of the regression residuals inflates the variance of precipitation and temperature forecasts, reducing problems of variance underestimation that are typical of regression-based models.

2.4. Dynamical downscaling

The Regional Climate Model selected to dynamically downscale the NCEP output was RegCM2 (Giorgi et al., 1988). We used the 10-year run performed by the Atmospheric Sciences Department, Iowa State University. This version of RegCM2 used the continental US domain of the PIRCS experiments (see Fig. 1 in Takle et al. (1999)). A 10 years run (1979–1988) was conducted using 6 h output from NCEP to define initial and boundary conditions. These were supplemented by observations of water-surface temperature in the Gulf of California and the Great Lakes, which are poorly resolved in the reanalysis. Precipitation was simulated using the Grell (1993) convection scheme and the simple warm-cloud explicit moisture scheme of Hsie et al. (1984). The simulations also used the NCAR Community Climate Model version 2 radiation package (Briegleb, 1992), the BATS version 1e surface package (Dickinson et al., 1992), and the non-local boundary layer turbulence scheme of Holtslag et al. (1990).

The RegCM2 grid spacing is 52 km on a Lambert conformal projection of the middle latitudes. Fig. 2 shows the RegCM2-gridpoints chosen for analysis in each of the three study basins. A buffer equal to that of the RegCM2 grid spacing was generated around each basin boundary and all RegCM2-gridpoints that fell within this buffered area were chosen for this analysis (see Fig. 2). Note the RegCM2 buffered area is much smaller than an NCEP grid box (Fig. 2). Table 2 summarizes some of the characteristics of the RegCM2 gridpoints chosen for each basin.

3. Hydrologic model

The hydrologic model chosen for this study is the US Geological Survey's Precipitation Runoff Modeling System (PRMS) (Leavesley et al., 1983; Leavesley and Stannard, 1995). PRMS is a distributed-parameter, physically based watershed model. Distributed parameter capabilities are provided by partitioning a watershed into Hydrologic Response Units (HRUs). Basin and HRU delineation, characterization, and parameterization were done for each basin using a geographic information system (GIS) interface. HRUs were delineated identically for each basin by (1) subdividing the basin into two flow planes for each channel; (2) subdividing the basin using three equal area elevation bands; and (3) intersecting the flow-plane map with the elevation-band map. The number of HRUs resulting from this process for each basin are listed in Table 1. The elevation ranges of the HRUs are listed in Table 2.

PRMS uses daily inputs of the climate variables precipitation (PRCP), maximum temperature (TMAX), minimum temperature (TMIN), and solar radiation. TMAX, TMIN, and PRCP are available at most climate stations across the United States. Solar radiation is generally not measured at the climate stations used in this study, so shortwave and longwave radiation were computed empirically using algorithms in PRMS (see Leavesley et al. (1983) for more information). TMAX, TMIN, and PRCP are estimated for each HRU by using the xyz methodology described in Hay et al. (2003). The xyz methodology uses measured TMAX, TMIN, and PRCP from a group of stations (or gridpoints) and spatially distributes from one point (a single daily mean value) to each HRU in a basin. The method allows for station data and gridpoints (i.e. DDS and NCEP timeseries) to be distributed similarly, both starting as a single daily mean value (note SDS output is at the average of the Best-Sta station location in each basin).

PRMS is conceptualized as a series of reservoirs (impervious zone, soil zone, subsurface, and groundwater) whose outputs combine to produce streamflow. For each HRU, a water balance is computed daily and an energy balance is computed twice each day. The sum of the water balances of each HRU, weighted by unit area, produces the daily watershed response.

Hydrologic model parameters describing topographic, vegetation, and soils characteristics were generated for each HRU from four digital databases: (1) USGS 3-arc second digital elevation models; (2) State soils geographic (STATSGO) 1-km gridded soils data (US Department of Agriculture, 1994); (3) US Forest Service 1-km gridded vegetation type and density data (US Department of Agriculture, 1992); and (4) USGS 1-km gridded Land Use/ Land Cover data (Anderson et al., 1976). An objective parameter estimation and calibration procedure was used to prevent biasing parameter estimates to any particular meteorological timeseries (Leavesley et al., 2002). Using this procedure, no changes are made to GIS generated spatial parameters. Calibration focused on the water balance parameters affecting potential evapotranspiration and precipitation distribution, and on subsurface and ground-water parameters affecting hydrograph shape and timing (Leavesley et al., 2002). Other model parameters were based on parameter sets from model applications to comparable basins in the same region (Leavesley et al., 1992).

4. Hydrologic model input data

The hydrologic model PRMS was forced with a daily mean TMAX, TMIN, and PRCP value derived from the following sources: (1) measured-station data; (2) NCEP output; (3) DDS; and (4) SDS output. DDS output for the United States was available from 1979–1988. To remove the bias from the state variables in each basin, PRMS was initialized with station data from October 1, 1977 to December 31, 1978. Then, 10 years (1979–1988) of TMAX, TMIN, and PRCP were distributed to the HRUs in each basin using daily mean TMAX, TMIN, and PRCP values from climate stations, NCEP output, and downscaled NCEP output. Due to the stochastic nature of SDS, the SDS-based simulations of runoff included 100 ensembles. Figs. 3–5 show the daily basin TMAX, TMIN, and PRCP mean by month for the meteorological input timeseries listed in Table 3 (excluding the bias-corrected timeseries) for the Animas, Carson, and Cle Elum basins, respectively. The SDS values are represented by a range computed from the 100 ensembles.

4.1. Temperature

Figs. 3a,b–5a,b show the daily basin TMAX and TMIN mean by month computed using the Best-Sta, All-Sta, NCEP, DDS, and SDS timeseries for the Animas, Carson, and Cle Elum basins, respectively. Examination of the TMAX and TMIN indicate problems with the All-Sta, NCEP, and DDS temperature data when compared with the Best-Sta. The All-Sta TMAX values are similar to Best-Sta values, but the All-Sta TMIN values are higher than Best-Sta values in the Animas and Carson River basins. NCEP TMAX values are consistently higher than Best-Sta values in the Carson River basin. NCEP TMAX values in the Animas and Cle Elum River basins are generally lower than Best-Sta values in the winter and higher in the summer months. DDS TMAX values are lower than Best-Sta values. DDS TMIN values are consistently higher than Best-Sta values in the Animas River basin. SDS temperature values are nearly identical to the Best-Sta values in all three basins.

A simple bias correction was performed on the raw All-Sta, NCEP, and DDS TMAX and TMIN timeseries to produce the Bias-All, Bias-NCEP, and Bias-DDS TMAX and TMIN timeseries, respectively. Biases were removed in the NCEP (and DDS and All-Sta) timeseries by (1) computing a monthly climatology of the NCEP TMAX and TMIN for each day; (2) subtracting the daily NCEP value of TMAX and TMIN from that climatology (to produce a daily anomaly value); and (3) adding the daily TMAX and TMIN anomaly from the NCEP model to the corresponding Best-Sta monthly station climatology of TMAX and TMIN. Because there were only 10 years of DDS output available for this study, an independent timeseries was not used to produce the TMAX and TMIN bias corrections. Due to the nature of the TMAX and TMIN bias correction, the monthly climatologies of Bias-NCEP, Bias-DDS, Bias-All, and Best-Sta are the same as the Best-Sta data (not shown).

TMAX and TMIN bias adjustments ‘correct’ the monthly mean values of All-Sta, DDS, and NCEP, but daily values of Bias-DDS generally do not contain the day-to-day variability present in the Best-Sta, All-Sta, NCEP, or SDS daily values for any of the basins. Fig. 6 shows for each basin the R-square values for TMAX and TMIN calculated between daily Best-Sta values and: (1) All-Sta; (2) Bias-All; (3) DDS; (4) Bias-DDS; (5) NCEP; (6) Bias-NCEP; and (7) SDS. R-Square values using SDS are as good or better than when using Bias-All. R-square values using Bias-DDS

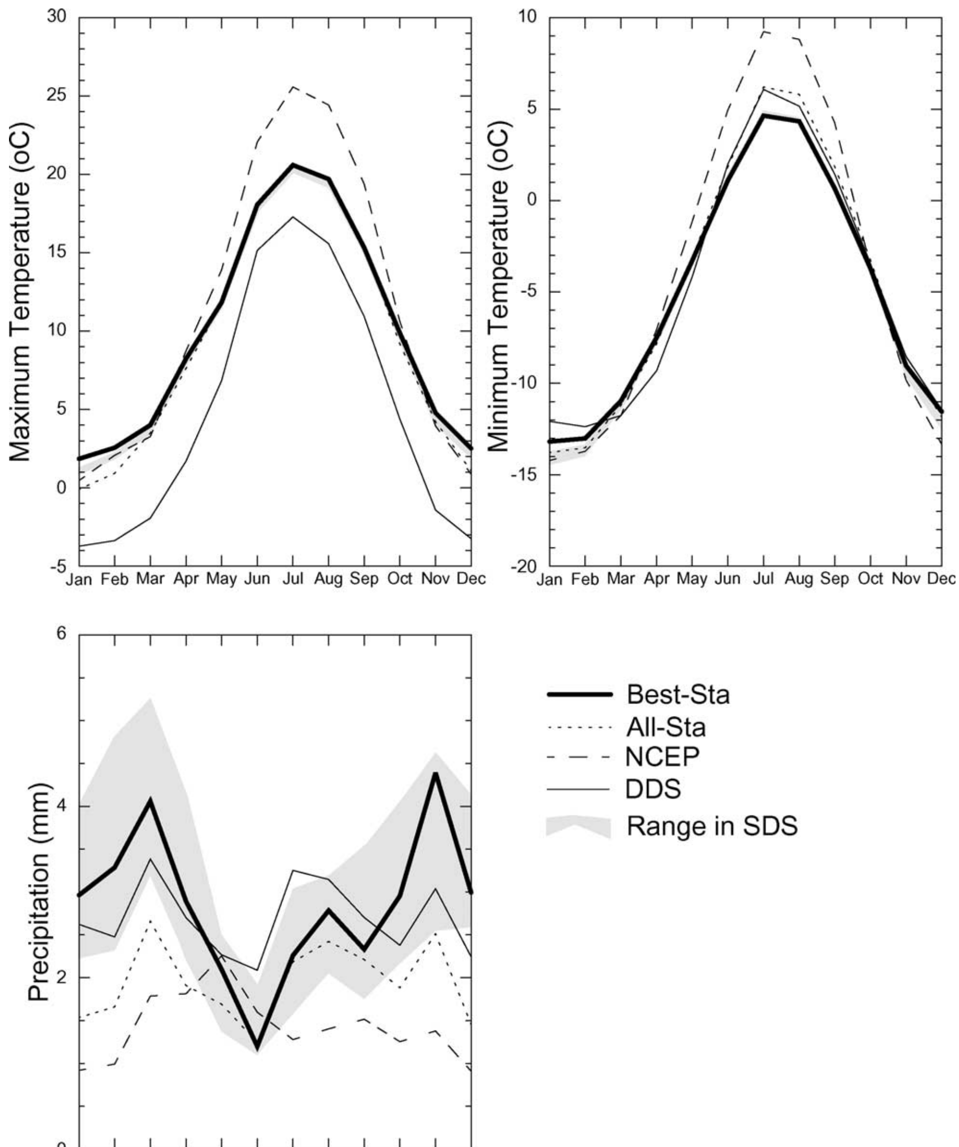

Fig. 3. Daily basin mean TMAX, TMIN, and PRCP by month for the Animas River basin.

output are the lowest of the bias-corrected input timeseries (solid horizontal line in each plot). In all cases Bias-NCEP R-square values are better than Bias-DDS (dotted horizontal line in each plot). This implies that TMAX and TMIN station data compiled at the scale of the DDS and NCEP output still contain the day-to-day variability present in the Best-Sta timeseries. This is interesting considering the huge domain used to extract the NCEP output (see Fig. 2). The DDS output may have identical monthly means values, but does not contain the day-to-day variability in TMAX and TMIN present in the measured station data.

4.2. Precipitation

Figs. 3c, 4c and 5c show the daily basin PRCP mean by month computed using the Best-Sta, All-Sta,

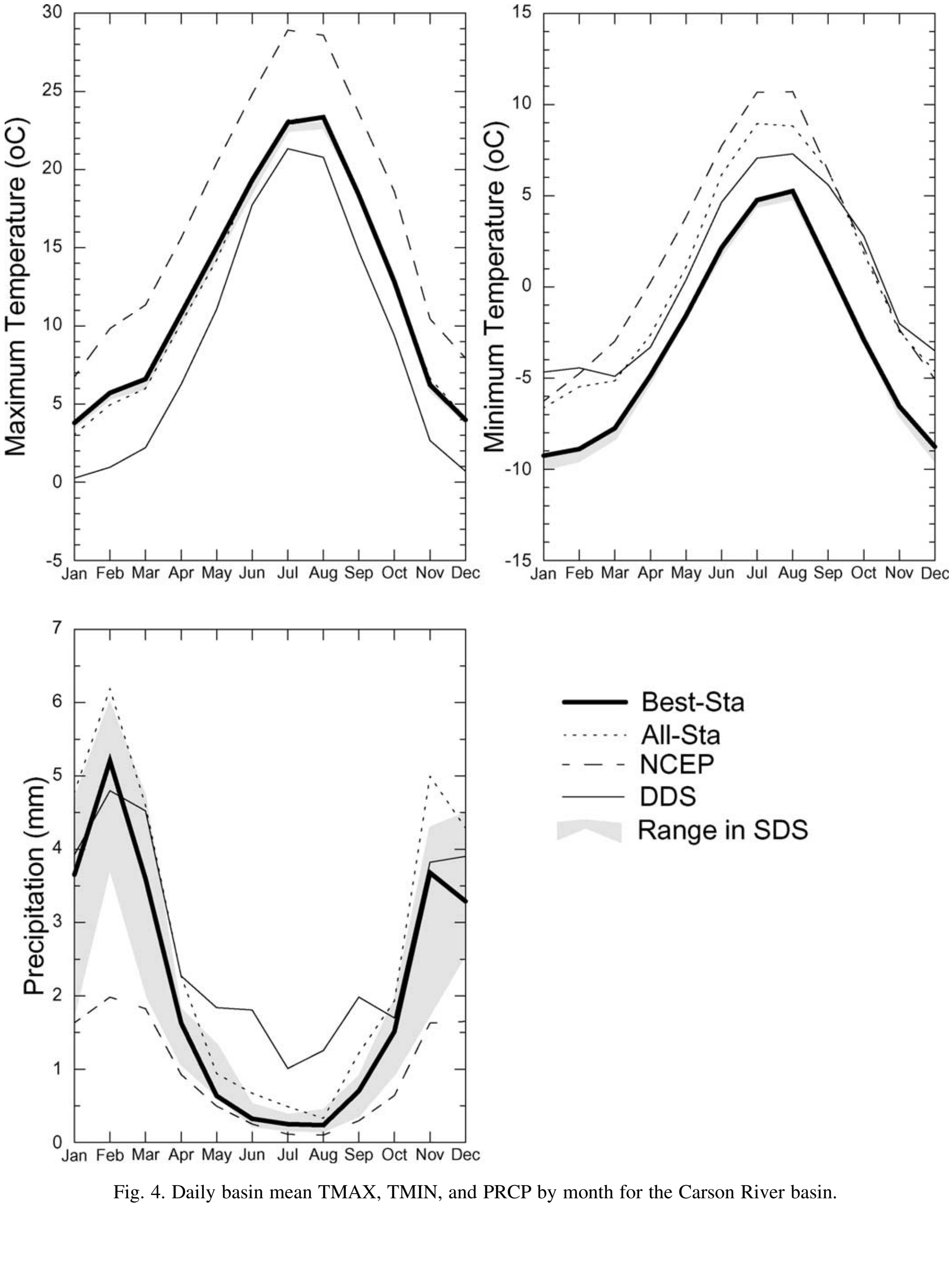

Fig. 4. Daily basin mean TMAX, TMIN, and PRCP by month for the Carson River basin.

NCEP, DDS, and SDS timeseries for the Animas, Carson, and Cle Elum basins, respectively. Comparison of the All-Sta, NCEP, DDS, and SDS with the Best-Sta PRCP timeseries shows that they all capture the gross aspects of the seasonal cycle of PRCP in all three basins, although there are some large discrepancies when using NCEP, All-Sta and DDS.

Based on these results, the raw NCEP, DDS, and All-Sta PRCP timeseries were 'corrected' for biases. The bias corrections were made on a monthly basis

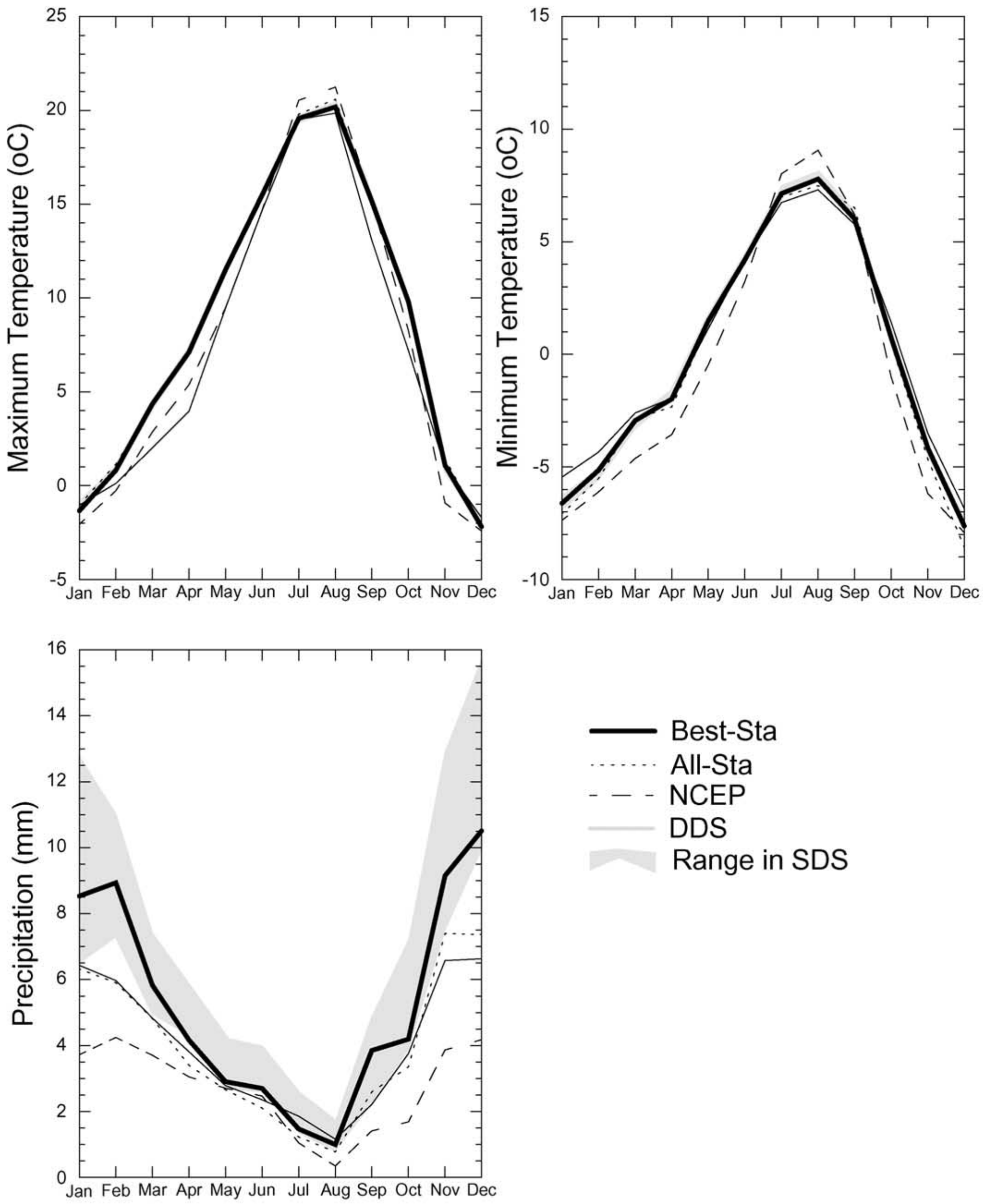

Fig. 5. Daily basin mean TMAX, TMIN, and PRCP by month for the Cle Elum River basin.

using a gamma transform which preserved the PRCP distribution. This procedure is similar to the transform method suggested by Panofsky and Brier (1968). The NCEP (and All-Sta and DDS) PRCP biases were corrected using the following steps: (1) force the NCEP PRCP values to have the same number of PRCP days as the Best-Sta timeseries (Table 1). This was accomplished by (a) ranking the NCEP PRCP output and (b) setting all values to zero with ranks equal to or lower than the number of dry days in

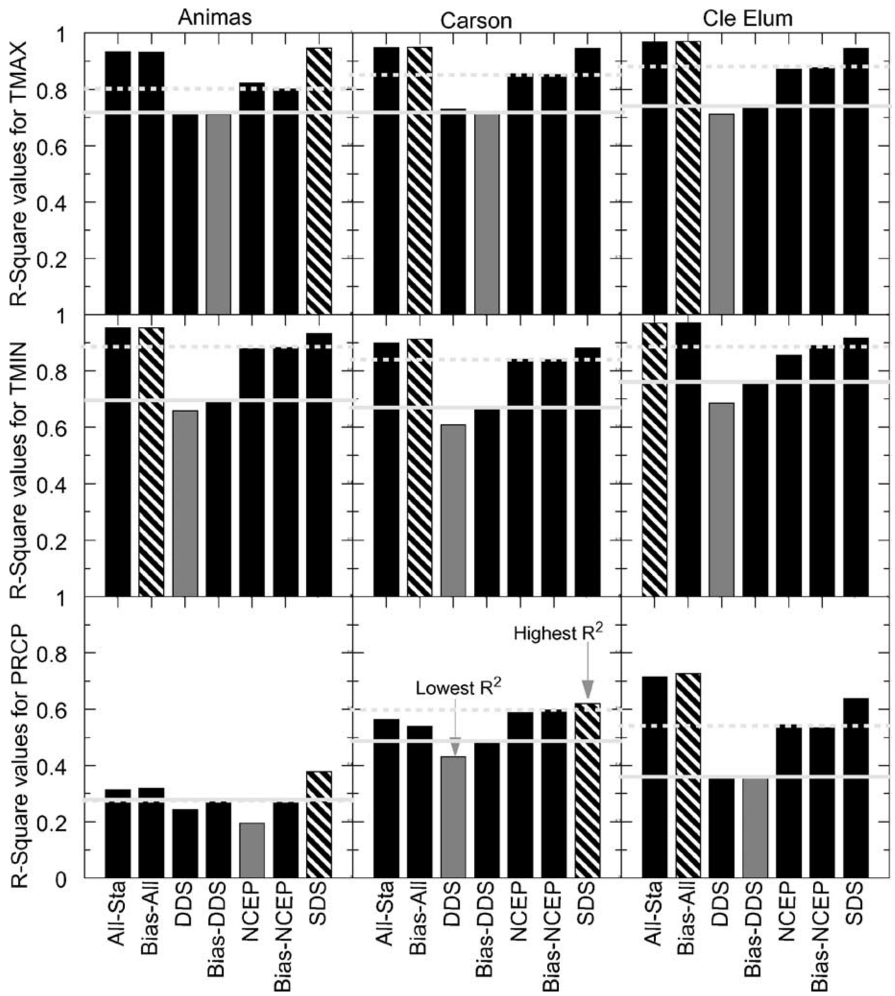

Fig. 6. TMAX, TMIN, and PRCP R-square values for each basin calculated using Best-Sta timeseries and: (1) All-Sta; (2) Bias-All; (3) DDS; (4) Bias-DDS; (5) NCEP; (6) Bias-NCEP; and (7) SDS daily timeseries.

the Best-Sta timeseries; (2) fit a gamma distribution to the resultant Best-Sta and NCEP timeseries (restricted to PRCP days); (3) for each NCEP PRCP day (i.e. all NCEP values above the thresholds identified in step (1b)), compute the cumulative probability in the gamma distribution fitted to the NCEP output, and then replace the raw NCEP value with the PRCP amount associated with the matched cumulative probability in the gamma distribution fitted to the Best-Sta data. Because there were only 10 years of DDS output available for this study, an independent timeseries was not used to produce the DDS PRCP bias corrections. Monthly values of Bias-NCEP, Bias-DDS, and Bias-All are similar to Best-Sta data (not shown).

The bias adjustments to the All-Sta, NCEP, and DDS PRCP may ‘correct’ the monthly mean values (Figs. 3c–5c), but daily values of the bias-corrected timeseries do not contain the day-to-day variability present in the Best-Sta values for any of the basins. Fig. 6 shows for each basin the R-square values for PRCP calculated between daily Best-Sta values and: (1) All-Sta; (2) Bias-All; (3) DDS; (4) Bias-DDS; (5) NCEP; (6) Bias-NCEP; and (7) SDS. R-square values are not greater than 0.75 for any of the basins indicating that none of the input timeseries contain the day-to-day variability seen in the Best-Sta timeseries. Note that the SDS R-square values are the highest in two out of three basins and are always greater than the DDS values.

5. Hydrologic model output

Figs. 7–9 show scatter plots of measured versus simulated daily runoff and the corresponding Nash-Sutcliffe (NS) goodness-of-fit statistic (Nash and Sutcliffe, 1970) for the three basins, respectively,

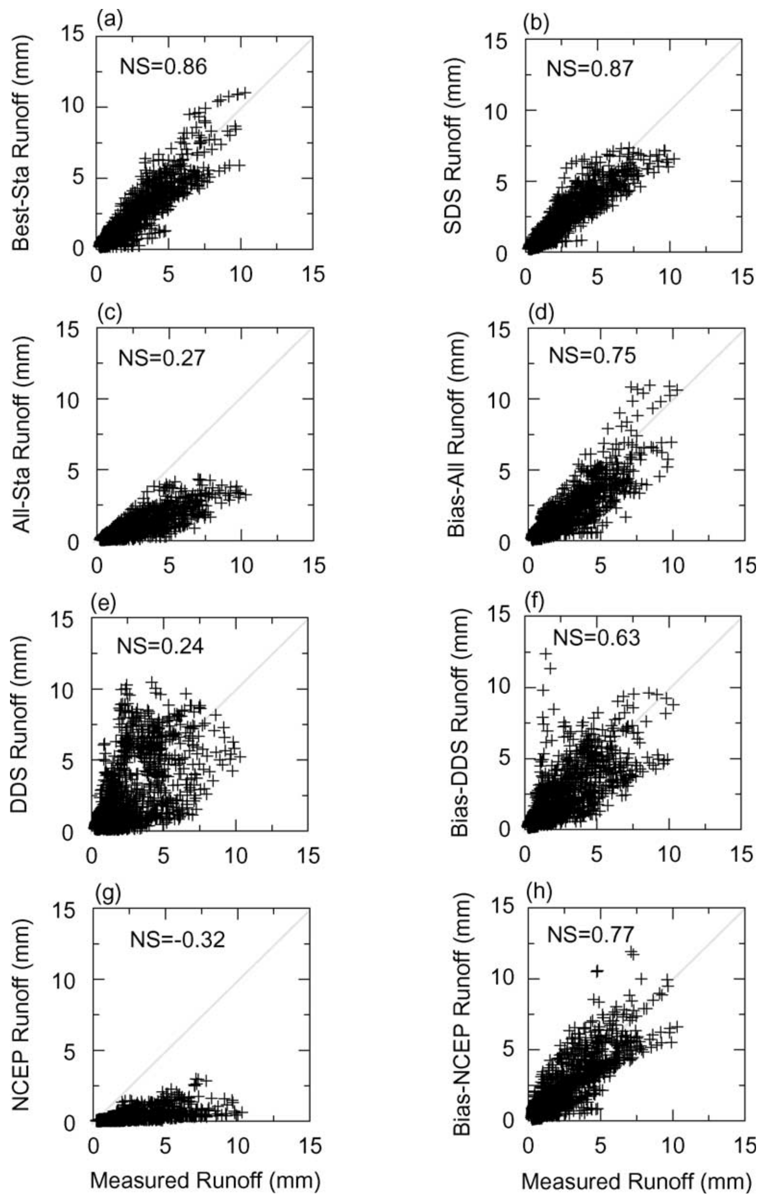

Fig. 7. Measured versus simulated runoff and corresponding Nash-Sutcliffe Goodness-of-Fit statistic (NS) for the Animas River basin. NS computed between measured runoff and PRMS simulated runoff using: (a) Best-Sta; (b) SDS; (c) All-Sta; (d) Bias-All; (e) DDS; (f) Bias-DDS; (g) NCEP; and (h) Bias-NCEP input timeseries.

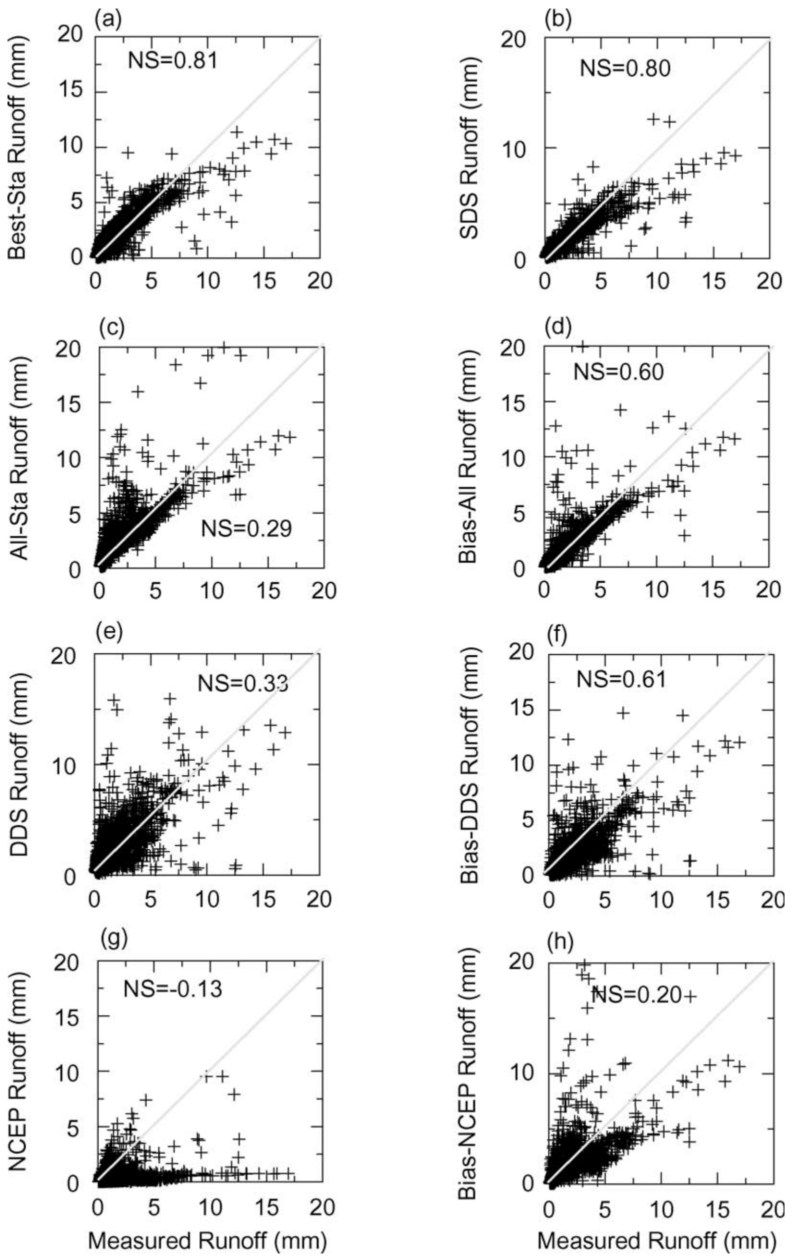

Fig. 8. Measured versus simulated runoff and corresponding Nash-Sutcliffe Goodness-of-Fit statistic (NS) for the Carson River basin. NS computed between measured runoff and PRMS simulated runoff using: (a) Best-Sta; (b) SDS; (c) All-Sta; (d) Bias-All; (e) DDS; (f) Bias-DDS; (g) NCEP; and (h) Bias-NCEP input timeseries.

using the input timeseries listed in Table 3. Fig. 10 summarizes the NS values listed in Figs. 7–9. For PRMS runoff simulated using the Best-Sta timeseries (Fig. 7–9a), the NS values are all above 0.75, indicating a good fit even with minimal calibration of the PRMS model parameters. As expected, PRMS outputs simulated using All-Sta, DDS, and NCEP (Figs. 7c,e,g–9c,e,g), result in much lower NS

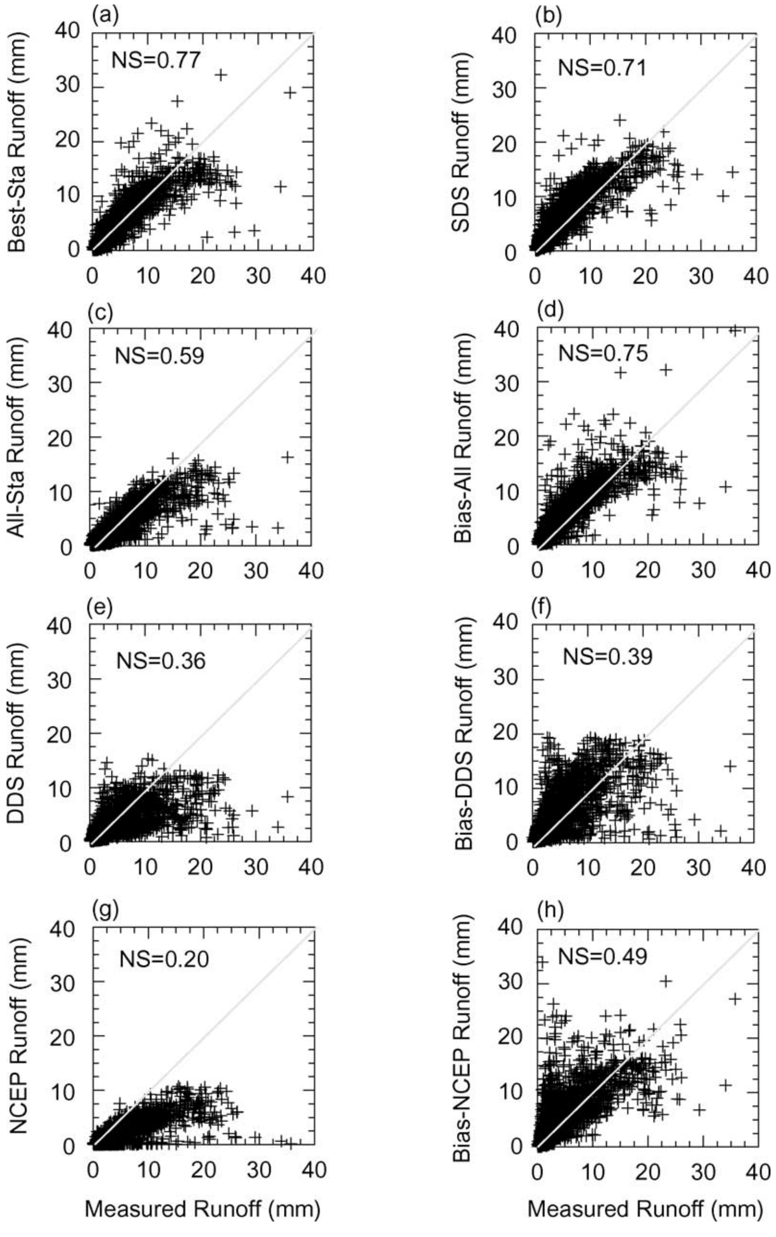

Fig. 9. Measured versus simulated runoff and corresponding Nash-Sutcliffe Goodness-of-Fit statistic (NS) for the Cle Elum River basin. NS computed between measured runoff and PRMS simulated runoff using: (a) Best-Sta; (b) SDS; (c) All-Sta; (d) Bias-All; (e) DDS; (f) Bias-DDS; (g) NCEP; and (h) Bias-NCEP input timeseries.

values (from 0.3 to 0.6 for All-Sta, less than 0.4 for DDS and below 0.2 for NCEP). For PRMS outputs simulated using bias-corrected input, model skill improves, but in all basins the NS values are significantly lower than those simulated using the Best-Sta or SDS input timeseries. PRMS model runoff simulated with SDS is as good as that simulated using Best-Sta data with a slight exception in the Cle Elum basin. Note that runoff simulated using Bias-DDS has the lowest NS values

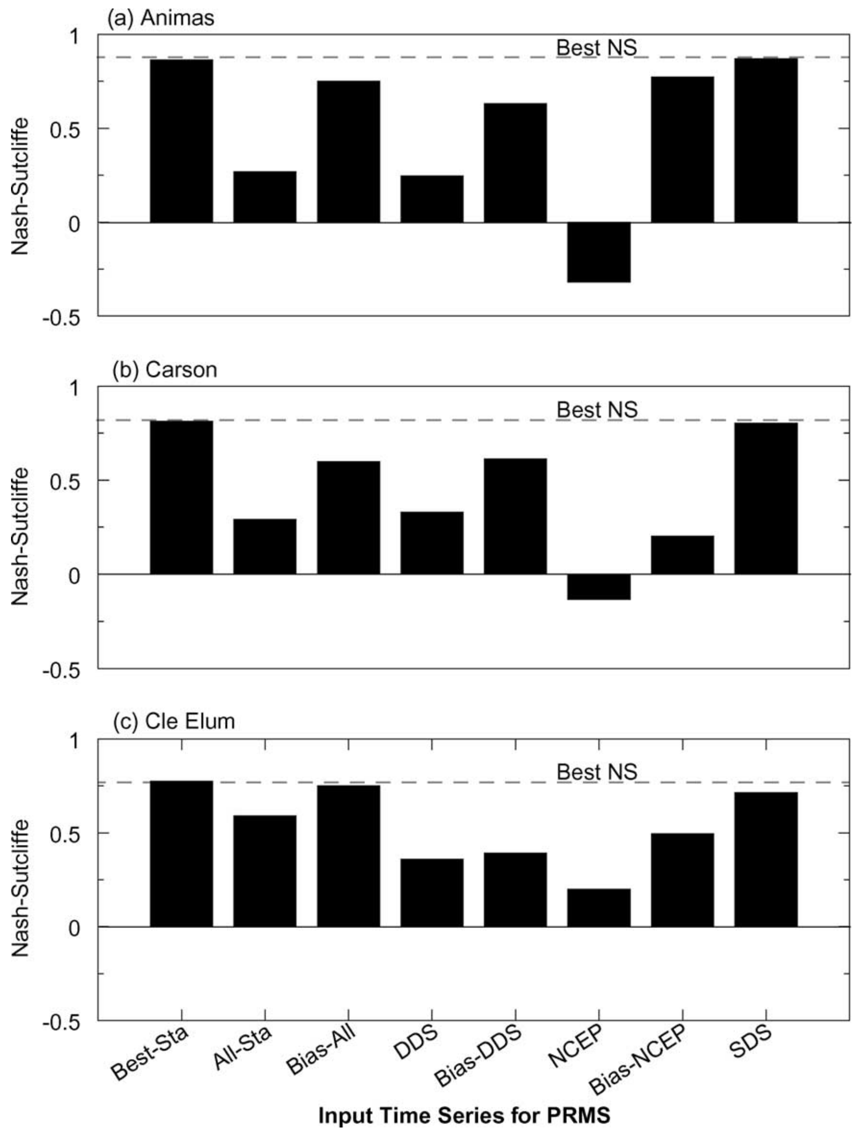

Fig. 10. Nash-Sutcliffe Goodness-of-Fit statistic (NS) for the: (a) Animas; (b) Carson; and (c) Cle Elum River basins.

of the bias-corrected results with the exception of the Carson River basin.

6. Summary and discussion

This study was initiated to examine possibilities for using statistically and dynamically downscaled output in hydrologic applications. Eight types of input timeseries (TMAX, TMIN, and PRCP) were tested using the hydrologic model PRMS (listed in Table 3). PRMS runoff simulated using Best-Sta and SDS timeseries produced realistic daily runoff in all three study basins (Figs. 7a,b–9a,b and 10). Analysis of the daily basin mean TMAX, TMIN, and PRCP by month computed using All-Sta, NCEP, and DDS timeseries indicated that these timeseries needed a bias correction (Figs. 3–5). PRMS runoff simulated using bias-corrected All-Sta timeseries (Bias-All) produced realistic daily runoff that was as good or better than that produced using the Bias-NCEP or Bias-DDS timeseries. The All-Sta timeseries were tested to determine if an area as large as that covered by gridpoints used in the DDS timeseries (see Fig. 2) could produce realistic TMAX, TMIN, and PRCP for basin-scale modeling. Results indicate that, after bias correction, the large-scale Bias-All timeseries performed almost as well as the Best-Sta timeseries, indicating that large-scale data can be made appropriate for hydrologic modeling in the three basins examined for this study. The NCEP timeseries were extracted from a significantly larger area than that used for the DDS or All-Sta (see Fig. 2). Interestingly enough, runoff simulated using Bias-NCEP was better than that using Bias-DDS in two out of three basins (Fig. 10). The most significant results were from runoff produced using the SDS input. In all three basins, SDS-based runoff simulations were almost as good or better than that produced using Best-Sta, Bias-All, Bias-DDS, or Bias-NCEP (Fig. 10).

Fig. 6 indicated that the day-to-day variability, present in the Best-Sta PRCP timeseries, was not replicated in any of the other timeseries (though SDS had the highest R-square values in two out of three basins). Snowmelt-dominated basins are much more strongly controlled by TMAX. The day-to-day variability present in the Best-Sta TMAX timeseries was best represented by the All-Sta and SDS timeseries. DDS-based timeseries had the lowest R-square values for TMAX. In snowmelt dominated basins, daily variations in PRCP tend to be of less importance. A correct volume of PRCP over the accumulation season (e.g. as represented in the April 1 snowpack) is what the hydrologic model requires. Because variability of runoff in snowmelt-dominated basins are much more strongly controlled by daily variations in temperature rather than precipitation, runoff simulated using SDS timeseries were as realistic as those produced using the Best-Sta timeseries. These results are consistent with Wilby and Dettinger's (2000) study of snowmelt-dominated basins in the Sierra Nevada. In their study they concluded that much of the hydrologic 'skill' arises from the fact that the snowpack integrates individual precipitation events throughput the winter season. The errors in daily precipitation, tend to cancel each other out. In contrast to rainfall-dominated basins, where the skill in modeling runoff is dependent on the skill of capturing discrete precipitation events, the skill of hydrologic simulations in snowmelt-dominated basins is due to simulations of the total snowpack and melt processes (which are largely influenced by TMAX).

7. Conclusions

The hydrologic response in three snowmelt-dominated basins in the western United States to dynamically- and statistically downscaled output from the National Centers for Environmental Prediction/National Center for Atmospheric Research Reanalysis (NCEP) was examined. Runoff produced using a distributed hydrologic model was compared using daily precipitation and maximum and minimum temperature timeseries derived from the following sources: (1) NCEP output; (2) dynamically downscaled (DDS) NCEP output using a Regional Climate Model (RegCM2); 3) statistically downscaled (SDS) NCEP output; (4) spatially averaged measured data used to calibrate the hydrologic model (Best-Sta) and (5) spatially averaged measured data derived from stations located within the area of the RegCM2 model output used for each basin, but excluding the Best-Sta set (All-Sta). The All-Sta timeseries are comparable in scale to the DDS model resolution and provide an appropriate test to determine if output at this scale can be used for simulation of basin-scale hydrology.

The All-Sta, NCEP, DDS, and SDS timeseries capture the gross aspects of the seasonal cycles of TMAX, TMIN, and PRCP. However, in all three basins large systematic biases in the All-Sta, NCEP, and DDS simulations of TMAX, TMIN, and PRCP were evident, which translated into unrealistic simulations of runoff. The All-Sta, NCEP, and DDS timeseries were corrected for biases (Bias-All, Bias-NCEP, and Bias-DDS, respectively).

Simulated runoff based on Best-Sta, SDS, Bias-All, Bias-NCEP, and Bias-DDS output were evaluated on a daily basis. SDS-based simulations of runoff were as good or better than Best-Sta. Bias-All based simulations of runoff were as good or of poorer quality than SDS. Bias-NCEP based simulations were as good or of poorer quality than Bias-All. Bias-DDS based simulations of runoff were always of poorer quality than Best-Sta and SDS. In two out of the three basins, Bias-NCEP based simulations were better than Bias-DDS. These results indicate that in snowmelt-dominated basins, TMAX, TMIN, and PRCP averaged over a large area can have the daily variations necessary for basin-scale modeling (as evidenced by the Bias-All results). The snowmelt-dominated basins are strongly controlled by accurate estimates of TMAX, therefore capturing daily variations in PRCP was found to be less important, and only the volume of PRCP over the accumulation season needs to be correct.

In conclusion, climate data of similar resolution to that of the DDS model can be made appropriate for basin-scale modeling when a bias correction is applied. This need for statistical correction (essentially a magnitude correction) may be somewhat troubling, but in the case of the large station-timeseries (All-Sta), the magnitude correction did indeed 'correct' for the change in scale. This was not shown to be true for the bias-corrected DDS output in two of the three basins. The DDS output could be 'corrected' for magnitude but did not contain the day-to-day variability needed for basin-scale modeling in snowmelt-dominated basins, present in the All-Sta and SDS timeseries. The major advantage of using Regional Climate Model output to simulate runoff is their physical realism. It is unknown if statistical corrections to model output will be valid in a future climate. Future work is warranted to identify the causes for (and removal of) systematic biases in DDS simulations, and improve DDS simulations of daily variability in local climate. Until then, SDS-based simulations of runoff appear to be the safer downscaling choice.

Acknowledgements

This work was funded by NOAA grant NA16GP1587 and by the NOAA Regional Integrated Sciences and Assessment Program (NOAA Cooperative Agreement NA17RJ1229). Regional Climate Model output was kindly provided by William J. Gutowski (Departments of Agronomy, Geological and Atmospheric Sciences, Iowa State University, Ames, Iowa). The authors would also like to thank Robert L. WIlby (Department of Geography, King's College London, Strand, London).

References

Anderson, J.R., Hardy, E.E., Roach, J.T., Witmer, R.E., 1976. A land use land cover classification system for use with remote sensor data, US Geological Survey Professional Paper 964, 28 pp.

Antolik, M.S., 2000. An overview of the National Weather Service's centralized statistical quantitative precipitation forecasts. J. Hydrol., 306–337.

Basist, A.N., Chelliah, M., 1997. Comparison of tropospheric temperatures derived from the NCEP/NCAR reanalysis NCEP operational analysis and the microwave sounding unit. Bull. Am. Meteorol. Soc. 78 (7), 1431–1447.

Briegleb, B.P., 1992. Delta-Eddington approximation for solar radiation in the NCAR community climate model. J. Geophys. Res. 97, 7603–7612.

Diaz, H.F., Anderson, C.A., 1995. Precipitation trends and water consumption related to population in the southwestern United States. Water Resour. Res. 31, 713–730.

Dickinson, R.E., Henderson-Sellers, A., Kennedy, P.J., 1992. Biosphere-atmosphere transfer scheme (BATS) version 1e as coupled to NCAR community climate model, NCAR Tech. Note 387 + STR, 72 pp.

Giorgi, F., Mearns, L.O., Shields, C., Mayer, L., 1988. A regional model study of the importance of local versus remote controls of the drought and the 1993 flood over the central United States. J. Climate 9, 1150–1161.

Grell, G.A., 1993. Prognostic evaluation of assumptions used by cumulus parameterizations. Mon. Weath. Rev. 121, 764–787.

Hay, L.E., Clark, M.P., Wilby, R.L., Gutowski, W.J., Leavesley, G.G., Pan, Z., Arritt, R.W., Takle, E.S., 2002. Use of regional climate model output for hydrologic simulations. J. Hydrometeorol. 3, 571–590.

Holtslag, A.A.M., de Bruijn, E.I.F., Pan, H.L., 1990. A high resolution air mass transformation model for short-range weather forecast. Mon. Weath. Rev. 118, 1561–1575.

Hsie, E.-Y., Anthes, R.A., Keyser, D., 1984. Simulations of frontogenesis in a moist atmosphere using alternative parameterizations of condensation and precipitation. J. Atmos. Sci. 41, 2701–2716.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Wollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., Fiorino, M., 2001. The NCEP-NCAR 50-year reanalysis: Monthly Means CD-ROM and Documentation. Bull. Amer. Meteorol. Soc. 82, 247–267.

Leavesley, G.H., Stannard, L.G., 1995. The precipitation-runoff modeling system-PRMS. In: Singh, V.P, (Ed.), Computer Models of Watershed Hydrology, Water Resources Publications, Highlands Ranch, CO, pp. 281–310, Chapter 9.

Leavesley, G.H., Lichty, R.W., Troutman, B.M., Saindon, L.G., 1983. Precipitation-runoff modeling system: User's manual US Geological Survey Water Investigation Report. 83–4238.

Leavesley, G.H., Markstrom, S.L., Restrepo, P.J., Viger, R.J., 2002. A modular approach to addressing model design, scale, and parameter estimation issues in distributed hydrological modeling. Hydrol. Process. 16, 173–187.

Michaelsen, J., 1987. Cross-validation in statistical climate forecast models. J. Climatol. Appl. Meteorol. 26, 1589–1600.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—a discussion of principles. J. Hydrol. 10, 282–290.

Pulwarty, R.S., 1995. Adaptive management of the Colorado River: The role of science and scientists. In: Melis, T., Wegner, D. (Eds.), Department of Interior Glen Canyon Environmental Studies: Aquatic, Geomorphic and Climatic Integration Group Report, Flagstaff, AZ, US Government Print, Washington, DC, pp. 36–49.

Ramage, C.S., 1983. Teleconnections and the siege of time. J. Climatol. 3, 223–231.

Serreze, M.C., Clark, M.P., Armstrong, R.L., McGinnis, D.A., Pulwarty, R.S., 1999. Characteristics of the western US snowpack from SNOTEL data. Water Resour. Res. 35, 2145–2160.

Takle, E.S., Gutowski, W.J. Jr., Arritt, R.W., Pan, Z., Anderson, C.J., Silva, R., Caya, D., Chen, S.-C., Christensen, J.H., Hong, S.-Y., Juang, H.-M.H., Katzfey, J.J., Lapenta, W.M., Laprise, R., Lopez, P., McGregor, J., Roads, J.O., 1944. Project to intercompare regional climate simulations (PIRCS): description and initial results. J. Geophys. Res. 104, 19443–19461.

Vislocky, R., Fritsch, J.M., 1997. An automated, observations-based system for short-term prediction of ceiling and visibility. Weath. Forecasting 12, 31–43.

Wilby, R.L., Dettinger, M.D, 2000. Stream.ow changes in the Sierra Nevada, CA simulated using statistically downscaled General Circulation Model output. In: McLaren, S, Kniveton, D. (Eds.), Linking Climate Change to Land Surface Change, Kluwer, Netherlands.

Wilby, R.L., Hay, L.E., Gutowski, W.J., Arritt, R.W., Takle, E.S., Pan, Z., Leavesley, G.H., Clark, M.P., 2000. Hydrological responses to dynamically and statistically downscaled climate model output. GRL 27, 1199–1202.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences: An Introduction, Academic Press, San Diego, pp. 467.