Hydrological responses to dynamically and statistically downscaled climate model output

Robert L. Wilby,[1,2] Lauren E. Hay,[3] William J. Gutowski Jr.,[4] Raymond W. Arritt,[4] Eugene S. Takle,[4] Zaitao Pan,[4] George H. Leavesley,[3] and Martyn P. Clark[5]

Abstract. Daily rainfall and surface temperature series were simulated for the Animas River basin, Colorado using dynamically and statistically downscaled output from the National Center for Environmental Prediction/ National Center for Atmospheric Research (NCEP/NCAR) re-analysis. A distributed hydrological model was then applied to the downscaled data. Relative to raw NCEP output, downscaled climate variables provided more realistic simulations of basin scale hydrology. However, the results highlight the sensitivity of modeled processes to the choice of downscaling technique, and point to the need for caution when interpreting future hydrological scenarios.

Introduction

The spatial resolution of General Circulation Models (GCMs) is too coarse to represent regional climate variations at the scales required for environmental impact assessments. Two techniques have been developed that attempt to counter this deficiency: semi-empirical (statistical) downscaling (SDS) of GCM outputs, and regional climate models (RCMs) nested within a GCM (Giorgi and Mearns, 1991). To date, very few studies have directly compared SDS and RCM output (e.g. Kidson and Thompson, 1998; Mearns et al., 1999; Murphy, 1999).

Statistical downscaling is analagous to the "model output statistics" (MOS) and "perfect prog" approaches used for short-range numerical weather prediction (Klein and Glahn, 1974). Both applications use correlations with climate variables at the synoptic scale (such as geopotential height fields) to simulate weather at the local scale (such as single site precipitation). Common SDS procedures involve weather-type classification, linear and non-linear regression, or modifications to stochastic weather generators (see Wilby and Wigley, 1997). A key strength of SDS is the low computational demand which facilitates the generation of ensembles of climate realizations. However, realistic SDS scenarios are contingent on strong/stationary empirical relationships, and on the choice of predictor variable(s) and transfer function(s) used for the downscaling (see Winkler et al., 1997).

[1] Division of Geography, University of Derby, Derbyshire, UK
[2] National Center for Atmospheric Research, Boulder, CO
[3] Water Resources Division, US Geological Survey, Denver, CO
[4] Departments of Agronomy, Geological and Atmospheric Sciences, Iowa State University, Ames, IA
[5] Cooperative Institute for Research in Environmetal Sciences, University of Colorado, Boulder, CO

Paper number 1999GL006078.
0094-8276/00/1999GL006078$05.00

Regional climate models simulate sub-GCM grid scale climate features dynamically at resolutions of 20-50 kilometres given time-varying atmospheric conditions supplied by the GCM bounding a specified domain (see reviews by McGregor, 1997; Giorgi and Mearns, 1999). The main advantage of RCMs is their ability to respond in physically consistent ways to different external forcings (such as land-surface or atmospheric chemistry changes). RCMs can also resolve important atmospheric processes such as orographic precipitation better than the driving GCM (Jones et al., 1995). However, RCMs are computationally demanding and require orders of magnitude more computer time than SDS to compute equivalent scenarios.

Ultimately the realism of both SDS and RCM scenarios depends on the quality of the climate data providing the boundary conditions. When directly compared, the two approaches can yield different regional climate scenarios because each utilizes different aspects of GCM output. For example, most SDS methods assume that the local variable is primarily a function of synoptic forcing, whereas all vertical levels of the atmosphere, including the surface, are considered by RCMs (Mearns et al., 1999). However, very little is known about the significance of such differences once assimilated by non-linear impact models. Here, differences in daily precipitation and temperature for the Animas River basin, southwest Colorado are examined, using raw NCEP data (as an analogue for GCM-scale output), SDS, and RCM simulations for current climate conditions. The simulated surface climate variables were used to drive a distributed hydrological model. Since the hydrological response of the basin is an integration of the regional climate (in time and space), the results provide insights into the overall "value-added" (or lost) to hydrological model skill due to the choice of downscaling technique.

Data and Methods

The Animas River basin, southwest Colorado, has a drainage area of 1820 km^2 and an elevation range of approximately 2000 to 4000 meters. Area-average daily precipitation (P), and maximum and minimum temperatures (Tmax, Tmin) were computed for the water years (WYs) 1980 to 1995 using two Snow Telemetry and one National Weather Service station. Area-average data for WYs 1987-95 were used to calibrate the SDS method; and WYs 1980-86 to evaluate all models.

Both the SDS and RCM were driven by gridded (approx. 200 km grid spacing) variables obtained from the NCEP/NCAR re-analysis (Kalnay et al., 1996). The SDS method (see Wilby et al., 1999) uses step-wise multiple linear regression to identify parsimonious sets of NCEP atmospheric variables – at the grid-point nearest the Animas basin – to predict local P, Tmax and Tmin. Separate regression equations were produced for each climatological season (i.e., DJF, MAM, JJA, SON) and surface variable (i.e., P, Tmax and Tmin).

Daily time series of precipitation occurrence and log-transformed, wet-day amounts were each regressed against three daily NCEP variables: [1] total column precipitable water (kg/m^2); [2] 500 hPa geopotential height (m); and [3] the meridional component of the wind computed from the 500 hPa geopotential height field (m/s). Daily time series of Tmax and Tmin were regressed against four daily NCEP variables: [1] mean sea level pressure (hPa); [2] the zonal component of the geostrophic wind at sea level (m/s); [3] total column precipitable water (kg/m^2); and [4] 500 hPa geopotential height (m). The rainfall and temperature models also include lag-1 autocorrelation functions, and stochastic representations of regression residuals. Conventional Monte Carlo methods are then used to generate ensembles of climate realizations using multiple passes of the NCEP predictors (as in Wilby et al., 1999).

RCM output was produced by RegCM2 (Giorgi et al., 1996), employing the continental U.S. domain of the Project to Intercompare Regional Climate Simulations (see Fig. 1 in Takle et al., 1999). Initial and boundary conditions from the NCEP/NCAR re-analysis were supplemented by observations of water-surface temperature in the Gulf of California and the Great Lakes, which are under-resolved in the re-analysis. Model grid spacing equates to 52 km on a Lambert conformal projection of the middle latitudes. This spacing gives minimal resolution of the Animas basin (just three RegCM2 grid boxes) and presents a severe test of the RCM's ability to resolve the local climate at a distance of >1500 km from the forcing frame. (In this respect, the SDS has some advantage over the RCM because it uses NCEP output closer to the Animas).

The Precipitation-Runoff Modeling System (PRMS) (Leavesley and Stannard, 1995) was used to simulate daily runoff (Q) given time series of surface climate variables (P, Tmax, and Tmin). PRMS partitioned the watershed into 34 hydrologic response units (HRUs) using distributed basin characteristics such as slope, aspect, elevation, vegetation and soil type. PRMS parameters were estimated from available topographic, soils and vegetation data but were <u>not</u> fit using an optimization algorithm. Downscaled climate variables were distributed to each HRU using observed monthly mean statistical relations between each surface variable and the elevation (z) at 37 stations in and around the basin. Additionally, the mean elevation of the RegCM2 and NCEP grid cells were adjusted on a monthly basis by solving for elevation using observed monthly z-climate relations. The fictitious elevations were then used to distribute P, Tmax, and Tmin as in the case of observed and SDS data (for more details see Hay et al., 2000). This correction is necessary because of the coarseness of the RegCM2 and NCEP grids, which cause the model terrain heights to depart from the actual elevation of the basin.

Results

Daily P, Tmax, Tmin and Q scenarios were generated for the verification period (WYs 1980-86) using: [1] area-average station data (Station), [2] an ensemble of twenty (SDS) realizations, [3] raw RCM output (RegCM2), [4] raw re-analysis output (NCEP), [5] elevation corrected RCM output ($RegCM2_{adj}$) and [6] elevation corrected re-analysis output ($NCEP_{adj}$). Table 1 reports the percentage of variance explained (E) by each model as well as the mean bias (D) of simulated daily Q (%), P (%), Tmax and Tmin (°C), with respect to observed data.

Fig. 1a shows that the NCEP output captures the timing of the June minimum and March maximum wet-day amounts but underestimates total rainfall by 47% in the uncorrected case and by 36% in $NCEP_{adj}$ (Table 1). In comparison, RegCM2 output for both the corrected and uncorrected cases is closer to the observed seasonal cycle, with D-values of –6 and –5% respectively. The SDS ensemble spans the observed rainfall regime in all months except for June (too high) and November (too low), but underestimates the rainfall total by 6%. It is noteworthy that relative to temperature and runoff (see below), the E-statistics for daily P are low for all models, ranging between 14% (NCEP) and 26% ($RegCM2_{adj}$).

In contrast, Tmin (Fig. 1b) is well represented by all methods. NCEP output shows a warm bias in summer and cold bias in winter that is reduced by elevation correction in $NCEP_{adj}$ (although overall D increased slightly from –0.4 to –0.7°C). The comparable monthly biases are smaller in RegCM2 and the corrected output of $RegCM2_{adj}$ has D = –0.1°C. However, NCEP and RegCM2 show large biases in Tmax (Fig. 1c) with D = –2.0 and –4.6°C respectively. NCEP has a significant cold bias in winter and spring that is partially offset by correction (D = –1.0°C), whereas RegCM2 has a cold bias throughout the year that is removed by $RegCM2_{adj}$ except in November to March (D = –0.9°C). In comparison, SDS has D = –0.5°C for Tmax.

The integrated results of the rainfall and temperature biases are reflected in simulated Q. Table 1 shows that all methods underestimate Q (with the exception of RegCM2). However, even Station data explained less than 85% of the variance in Q, suggesting that a component of the bias may be attributed

Table 1. Percentage of explained variance (E) and mean bias (D) in observed daily P, Tmax, Tmin and Q due to different downscaling methods.

	P		Tmin		Tmax		Q	
	E (%)	D (%)	E (%)	D (°C)	E (%)	D (°C)	E (%)	D (%)
Station	-	-	-	-	-	-	84	-4
SDS*	18	-6	88	+0.1	90	-0.5	78	-22
NCEP	14	-47	75	-0.4	81	-2.0	75	-65
$NCEP_{adj}$	15	-36	74	-0.7	79	-1.0	72	-54
RegCM2	26	-6	66	+0.9	72	-4.6	48	+5
$RegCM2_{adj}$	26	-5	67	-0.1	72	-0.9	69	-11

* mean results from an ensemble of twenty members

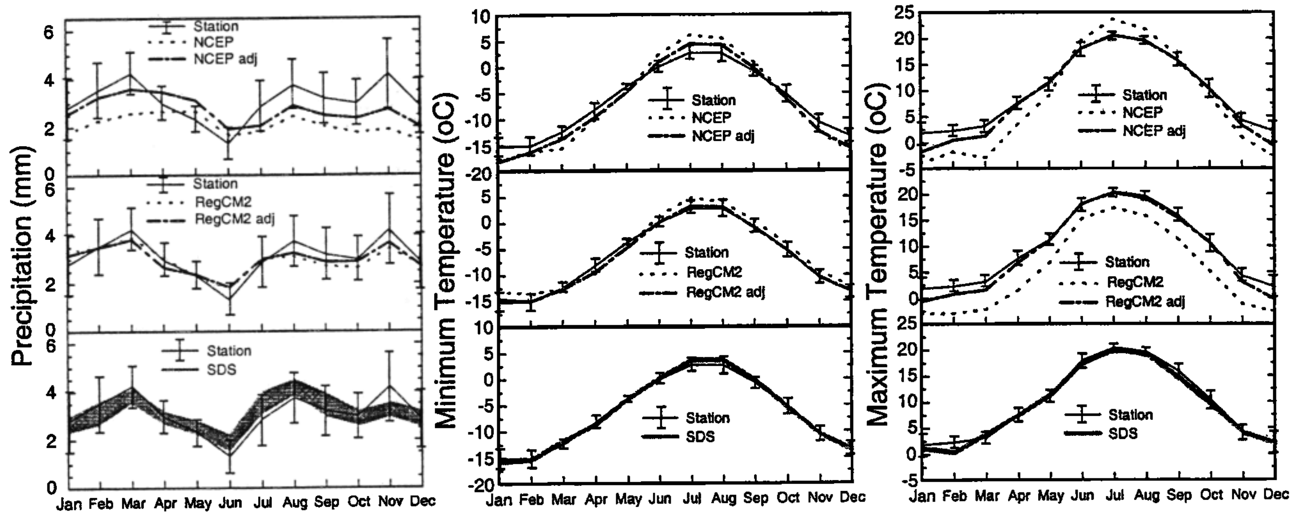

Figure 1. Downscaled monthly mean daily (a) P, (b) Tmin and (c) Tmax for WYs 1980-86, compared with area-averaged station data for the Animas River basin. The error bars for Station data correspond to ± 2 SE. The dotted lines for NCEP and RegCM2 represent the uncorrected model output. The gray shading for the SDS results shows the range of values produced by an ensemble of twenty members.

to the hydrological model (and calibration data in the case of SDS). Overall, SDS had the highest E (78%) despite a larger underestimate of total runoff than RegCM2. NCEP underestimated Q by 65% compared with +5% for RegCM2 and −22% for SDS. Elevation correction of the RegCM2 output increased E from 48% to 69% but did not improve the bias in Q (from +5 to −11%).

Fig. 2 compares the relative skill of the downscaling methods at simulating daily runoff using annual values of the coefficient of efficiency (CE) (Nash and Sutcliffe,1970). The station data provided the best simulation results with the majority of the years having CE scores of 0.8 or higher. The magnitude of these values indicates that, even though parameters were not optimized, the performance of the hydrological model is still quite good.

Overall the CE scores for Station data fall within the ensemble range of the SDS method. The CE scores for $RegCM2_{adj}$ and $NCEP_{adj}$ also lie within the bounds of the SDS ensemble for all years except WY1984 and WY1981 respectively, when the dynamical models had lower skill. In comparison, the skill of NCEP was lower than that of SDS in

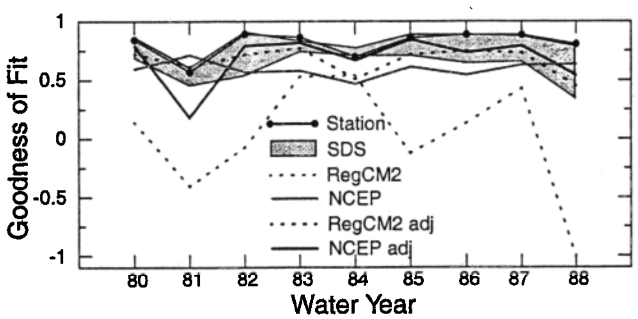

Figure 2. Nash-Sutcliffe scores of simulated versus observed daily Q computed on a water year basis for different downscaling methods. The goodness of fit scores for the Station output provide a measure of errors due to the hydrological model and/or choice of stations since observed P, Tmax and Tmin were used in this case.

all years apart from WY1981 when the modeled runoff from NCEP was closest to observed. The least skillful model output was obtained from RegCM2: the negative CE scores in WYs 1981, 1985 and 1988 indicate that the observed mean of Q is a better predictor of daily Q than the model (Wilcox et al., 1990).

Discussion

The preceding results provide insights into the relative strengths and weaknesses of each technique for basin-scale hydrological modeling. The most straightforward procedure is to use coarse resolution re-analysis output (in this case gridded P, Tmax and Tmin) and apply this information directly to the hydrological model. These data provide a datum with which the "value-added" of downscaling may be assessed. However, underestimation of total P (due to shortfalls in all seasons except summer), combined with biases in Tmax (of the order −2°C) result in lower E (Table 1) and CE scores for NCEP than Station and SDS simulations (Fig. 2).

The next level of sophistication is to correct NCEP output for systematic biases. Elevation corrections were shown to reduce errors in monthly Tmax and Tmin but were less successful for P (Fig. 1a). However, the net result of the $NCEP_{adj}$ corrections was an improvement in CE for all but two of the WYs, and a slight reduction in the bias of Q.

The SDS technique is of intermediate complexity and has the advantage of efficiently producing ensembles of surface climate variables given a very limited set of gridded predictor variables. Overall, SDS had much greater skill for Tmax and Tmin than for P, and returned the highest value of E for daily Q. The relatively low E-statistic for daily P reflects the large stochasticity of this variable in the SDS model. Conversely, the high skill for Q was attributed to well-timed snowpack melt (as regulated by Tmax and Tmin) and reasonable estimates of *gross* snowpack accumulation (rather than the sequence of individual precipitation events).

The RegCM2 and $RegCM2_{adj}$ scenarios were the most time consuming and computationally demanding to produce, so the

SDS ensemble was compared with a single realization of the RCM. Despite the higher level of sophistication and physical realism of the model, the uncorrected RegCM2 monthly P, Tmax and Tmin regimes were not generally as skillful as those of SDS. The cold bias in Tmax (also noted by Giorgi et al., 1993) leads to more persistent snow-pack and delayed spring melt. Hence, the E-scores for the RegCM2 were lower than those for SDS even though the total runoff was better estimated by the RCM. As with NCEP, the elevation corrections did yield gains in model verification performance such that the annual CE scores for $RegCM2_{adj}$ were within the range generated by SDS. This gain in skill implies that the quality of the uncorrected RegCM2 climate simulations is largely constrained by the bias in Tmax (which was greater in RegCM2 than in NCEP output).

From our single-basin study it is concluded that the SDS and RCM methods have greater skill (in terms of modeling hydrology) than the coarse resolution data used to drive the downscaling. The SDS has the advantage of requiring very few parameters – an attribute that makes this procedure attractive for many hydrological applications. The RCM output, once elevation corrected, provides better estimates of the water balance than the raw and corrected NCEP output. However, since the methods provide varying results, care must be taken in interpreting scenarios of basin-scale hydrology under both present and future climate forcing.

Acknowledgements. This research was supported by ACACIA, EPRI, NOAA (NA86GP0572), the NASA RESAC Program (NAG13-99005) and the National Science Foundation (EAR-9634329). We gratefully acknowledge the comments of Greg McCabe and two anonymous referees. NCAR is sponsored by the National Science Foundation.

References

Giorgi, F. and Mearns, L.O. 1991. Approaches to the simulation of regional climate change. A review. *Rev. Geophys.*, **29**, 191-216.

Giorgi, F. and Mearns, L.O. 1999. Introduction to special section: Regional climate modeling revisited. *J. Geophys. Res.*, **104**, 6335-6352.

Giorgi, F., Bates, G.T. and Nieman, S.J. 1993: The multiyear surface climatology of a regional atmospheric model over the Western United States. *J. Climate*, **6**, 75-95.

Giorgi, F. et al. 1996. A regional model study of the importance of local versus remote controls of the 1988 drought and the 1993 flood over the central United States. *J. Climate*, **9**, 1150-1161.

Hay, L.E., et al. 2000. A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States. *J. Am. Wat. Resour. Assoc.*, in press.

Jones, R.G., Murphy, J.M. and Noguer, M. 1995. Simulation of climate change over Europe using a nested regional-climate model. I: Assessment of control climate, including sensitivity to location of lateral boundaries. *J. Roy. Met. Soc.*, **121**, 1413-1449.

Kalnay, E. et al. 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437-471.

Kidson, J. W. and Thompson, C.S. 1998. A comparison of statistical and model-based downscaling techniques for estimating local climate variations. *J. Climate*, **11**, 735-753.

Klein, W.H. and Glahn, H.R. 1974. Forecasting local weather by means of model output statistics. *Bull. Amer. Meteor. Soc.*, **55**, 1217-1227.

Leavesley, G.H., and 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, Colorado, p. 281-310.

McGregor, J.J. 1997. Regional climate modelling. *Meteorol. Atmos. Phys.*, **63**, 105-117.

Mearns, L.O. et al. 1999. Comparison of climate change scenarios generated from regional climate model experiments and statistical downscaling. *J. Geophys. Res.*, **104**, 6603-21.

Murphy, J. 1999. An evaluation of statistical and dynamical techniques for downscaling local climate. *J. Climate*, **12**, 2256-2284.

Nash, J.E. and Sutcliffe, J.V. 1970. River flow forecasting through conceptual models, I, A discussion of principle. *J. Hydrol.*, **10**, 282-290.

Takle, E. S. et al. 1999. Project to Intercompare Regional Climate Simulations (PIRCS): Description and initial results. *J. Geophys. Res.*, **104**, 19443-19461.

Wilby, R.L. and Wigley, T.M.L. 1997. Downscaling General Circulation Model output: a review of methods and limitations. *Prog. Phys. Geog.*, **21**, 530-548.

Wilby, R.L., Hay, L.E. and Leavesley, G.H. 1999. A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River basin, Colorado. *J. Hydrol.*, **225**, 67-91.

Wilcox, B.P., et al. 1990. Predicting runoff from rangeland catchments: A comparison of two models. *Wat. Resour Res.*, **26**, 2401-2410.

Winkler, J.A., et al. 1997. The simulation of daily temperature series from GCM output. Part II: Sensitivity analysis of an empirical transfer function methodology. *J. Climate*, **10**, 2514-2532.

R. L. Wilby, University of Derby, Kedleston Road, Derby, DE22 1GB, UK. (e-mail: r.l.wilby@derby.ac.uk)

L. E. Hay and G. H. Leavesley, US Geological Survey, P.O. Box 25046, Lakewood, CO 80225. (e-mail: lhay@usgs.gov; george@usgs.gov)

R. W. Arritt, W. J. Gutowski, Z. Pan and E. S. Takle, Iowa State University, Ames, IA 50011. (e-mail: rwarritt, gutowski, panz; or gstakle@iastate.edu)

M. P. Clark, University of Colorado, P.O. Box 449, Boulder, CO 80309. (e-mail: mpclark@kryos.colorado.edu)

(Received July 13, 1999; revised November 22, 1999; accepted February 14, 2000)