HYDROLOGICAL PROCESSES
Hydrol. Process. (2013)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.9734

Evaluation of dynamically downscaled reanalysis precipitation data for hydrological application

Satish Bastola[1]* and Vasubandhu Misra[1,2,3]

[1] *Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, FL USA*
[2] *Department of Earth, Ocean and Atmospheric Sciences, Florida State University, Tallahassee, FL USA*
[3] *Florida Climate Institute, Florida State University, Tallahassee, FL USA*

Abstract:

Skilful and reliable precipitation data are essential for seasonal hydrologic forecasting and generation of hydrological data. Although output from dynamic downscaling methods is used for hydrological application, the existence of systematic errors in dynamically downscaled data adversely affects the skill of hydrologic forecasting. This study evaluates the precipitation data derived by dynamically downscaling the global atmospheric reanalysis data by propagating them through three hydrological models. Hydrological models are calibrated for 28 watersheds located across the southeastern United States that is minimally affected by human intervention. Calibrated hydrological models are forced with five different types of datasets: global atmospheric reanalysis (National Centers for Environmental Prediction/Department of Energy Global Reanalysis and European Centre for Medium-Range Weather Forecasts 40-year Reanalysis) at their native resolution; dynamically downscaled global atmospheric reanalysis at 10-km grid resolution; stochastically generated data from weather generator; bias-corrected dynamically downscaled; and bias-corrected global reanalysis. The reanalysis products are considered as surrogates for large-scale observations. Our study indicates that over the 28 watersheds in the southeastern United States, the simulated hydrological response to the bias-corrected dynamically downscaled data is superior to the other four meteorological datasets. In comparison with synthetically generated meteorological forcing (from weather generator), the dynamically downscaled data from global atmospheric reanalysis result in more realistic hydrological simulations. Therefore, we conclude that dynamical downscaling of global reanalysis, which offers data for sufficient number of years (in this case 22 years), although resource intensive, is relatively more useful than other sources of meteorological data with comparable period in simulating realistic hydrological response at watershed scales. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS reanalysis; bias correction; rainfall–runoff model; dynamic downscaling

Received 8 July 2012; Accepted 9 January 2013

INTRODUCTION

Global climate models (GCMs) are the most widely used tools for modelling future climate. The output GCMs or global reanalyses do well in representing large-scale atmospheric circulation features, but they lack in reproducing surface variables (e.g. precipitation) at a scale that is often required for hydrological application (Xu 1999; Wilby *et al*., 2000; Fowler *et al*., 2007). This is also the reason why output from global climate models cannot be directly utilized for evaluating future hydrologic impact of climate change (Wood *et al*., 2004). Therefore, dynamic downscaling (e.g. Mesinger *et al*., 2006; Kanamitsu and Kanamaru, 2007) has now become an important component of watershed-scale hydrologic modelling.

The regions that exhibit significant interannual variations in precipitation, e.g. southeastern United States (SEUS; Wang *et al*., 2010; Chan and Misra, 2010; Seager *et al*., 2008; Ropelewski and Halpern, 1987), present a challenge to produce reliable hydrological forecasts. However, for sustainable development of water resources in such regions, reliable hydrological forecasts are necessary for making optimal decisions on water resource management.

Yao and Georgakakos (2001) showed that inclusion of future climate information into hydrological models lead to more accurate hydrologic forecasts, which can result in economic benefits (e.g. increases in hydropower revenues). One of the primary obstacles to producing good hydrological forecasts is the absence or lack of skilful and reliable meteorological forcing. Another limitation is the inherent uncertainties of the hydrological models. Furthermore, the domain of this study, i.e. the SEUS, is subject to extreme weather and climate events (e.g. hurricanes, seasonal droughts and floods) that are usually difficult for climate models to simulate or forecast. Together, these factors pose a challenge in developing hydrological models for watersheds in the region (HEPEX, 2006). More importantly, over the SEUS region, seasonal climate prediction models do not show very promising skill in forecasting precipitation and temperature in summer (Stefanova *et al*., 2012b). High-resolution rainfall data that capture spatial and temporal distributions of rainfall in great detail is essential for hydrological application.

*Correspondence to: Satish Bastola, Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, FL, USA. E-mail: satish_bastola@yahoo.com

Two approaches have come to the fore to bridge the gap that exists between the availability of meteorological forcing from relatively coarse resolution global models and the requirement of finer resolution forcing for hydrological models: development of high-resolution regional reanalysis (e.g. Mesinger *et al*., 2006) or creation of high-resolution atmospheric data from downscaling global reanalysis (e.g. Kanamitsu and Kanamaru, 2007; Stefanova *et al*., 2012a). The conventional regional reanalysis use global reanalysis as a lateral boundary forcing along with regional data assimilation system to produce a relatively high-resolution meteorological dataset. Many studies claim that dynamical downscaling from global reanalysis amounts to regional reanalysis (Von Storch *et al*., 2000; Kanamitsu and Kanamaru, 2007). Kanamitsu and Kanamaru (2007) argue that reanalysis products obtained from downscaling global reanalysis can offer computational advantage and greater regional detail compared with the current advanced regional reanalysis. In the dynamic downscaling of global reanalysis, regional climate models (RCMs) are forced at their lateral boundaries by the global reanalysis to generate the regional climate. An RCM with a relatively high fidelity is a useful tool in describing regional scale climate conditions and in producing high-resolution meteorological data. RCMs have the potential to improve the spatial detail of simulated climate because of improved representation of subgrid scale features. Although the ability of RCMs to reproduce the observations has improved significantly over the years, the use of output from the RCMs is likely to be limited by the systematic errors in global reanalysis products (e.g. Mooney *et al*., 2011; Frei *et al*., 2003; Hagemann *et al*., 2004). Hongwei *et al*. (2012) observed significant differences in the simulations of the weather research and forecasting model forced with three global reanalysis datasets. Similarly, Stefanova *et al*. (2012a) observed significant biases in the generated regional climate using the regional spectral model (Kanamitsu and Kanamaru 2007) forced with global reanalysis of the National Centers for Environmental Prediction/Department of Energy Global Reanalysis (NCEP-R2; Kanamitsu *et al*., 2002) and the European Centre for Medium-Range Weather Forecasts (ECMWF) 40-year Reanalysis (ERA-40; Uppala *et al*., 2006).

In many hydrological applications, the hydrological models are forced with downscaled meteorological data without a proper assessment of the model forcing. The biases that exist in the forcing data will subsequently propagate through the hydrological models, resulting in simulation biases that have direct consequences on forecast quality and operational usefulness. Therefore, to use the high-resolution climate model data for hydrological modelling and for operational purposes, the output from the RCM needs to be suitably corrected for bias (Wilby *et al*., 2000). Bias-correction methods are usually applied to correct systematic errors in the output of atmospheric models.

Bias-correction methods include correction to the mean (e.g. Bastola and François, 2012), standardization (Wilby *et al*., 2004) or quartile-based mapping (Li *et al*., 2010). Standardization of a variable (i.e. by subtracting a mean from it and dividing the difference by the standard deviation) is widely used in statistical downscaling methods to reduce bias. Similarly, the quantile mapping method applies correction not only to the mean but also to the distribution. Therefore, this method corrects for errors in variability as well.

Wood *et al*. (2004) used the quantile-based approach to correct monthly simulated variables to force a hydrological model. This approach is relatively simple and extends the correction of the means to the shape of the distribution. In other words, this method is capable of correcting errors in variability as well. However, it maps only modelled values to observed values. Therefore, extreme values (e.g. in a projected future climate) that may not have been observed in the past cannot be properly corrected. This quantile-based approach has been successfully implemented in hydrological applications. This method was however found to produce bias, as it does not preserve the relationship between precipitation and temperature (Zhang and Georgakakos, 2011). More recently, Piani and Haerter (2012) presented a statistical bias-correction methodology that allows climate model output of both temperature and precipitation to be used as forcing for impact models.

The objective of this study is to evaluate regional and global reanalysis datasets for the SEUS region by propagating them through a suite of conceptual hydrological models and comparing the subsequent output to observations. For this purpose, a number of watersheds within the region are selected. With the widespread availability of terrestrial data based on geographic information system, distributed hydrologic models, which explicitly account for spatial variation in topography, meteorological inputs, and water movement, have become popular (e.g., Ajami *et al*., 2006; Georgakakos *et al*., 2004). Distributed modelling holds significant promise for better simulation of streamflow with high spatial resolution. However, uncertainty in model structure and parameter and uncertainty in rainfall estimates contribute to significant uncertainty in streamflow (Georgakakos *et al*., 2004). Moreover, for the watersheds selected in this study, the rainfall time series, a principal input to drive hydrological models, is available only at a watershed scale. Therefore, this study focuses on only spatially lumped conceptual models. Moreover, identification of streamflow records suitable for the evaluation of downscaling methods presents somewhat different problems than those that exist for the precipitation and temperature records. In the United States, almost all gauged streams are affected to some extent by human interventions such as upstream reservoirs, upstream diversion works, and interbasin transfer of water. Therefore, selecting watersheds that are free of or minimally affected by water management is essential for parameterization and subsequent application of watershed models for evaluation of model forcing. Hydrological models, identified for selected watersheds, are then forced

DOI: 10.1002/hyp

with the downscaled precipitation datasets. Simulations are also made with the raw large-scale gridded precipitation dataset. The results from the hydrological models with the different meteorological forcings are discussed in the subsequent sections.

METHODOLOGY

Watershed selection

In this study, we select watersheds that are free of or minimally affected by water management. The watersheds included in this study are from the Model Parameter Estimation Experiment (MOPEX) dataset (Schaake *et al*., 2006), which was based on Hydro Climatic Data Network (HCDN) (Slack *et al*., 1993) and hydroclimatological data set of Walilis *et al*. (1991). For the selected watersheds, the hydrometeorological data extend well beyond 1970. It is assumed that water management minimally affects the selected watersheds, as HCDN in its collection excluded watersheds that are subject to diversion, transfer and significant change in land cover. For watersheds recently impacted by human activities, HCDN included only the record for the earlier period during which the criterion for non-impairment of natural stream was met. Both networks include only those gauges believed to be unaffected by upstream regulation that have long enough data records to be suitable for climate studies. Furthermore, MOPEX used only those datasets that met the precipitation gauge density requirement.

A total of 28 watersheds located within the SEUS (Figure 1) are considered in this study. The attributes of the selected watersheds are shown in Table I. For model calibration, the aerial average rainfall and potential evapotranspiration data, both key to the forcing required for the hydrological models, are taken from MOPEX datasets, as are the streamflow data required for the estimation of hydrological model parameters.

Datasets

The two fine-scale reanalysis precipitation data obtained by dynamically downscaling two different global reanalysis following Stefanova *et al*. (2012a) were used. Stefanova *et al*. (2012a) have downscaled global atmospheric reanalysis (NCEP-R2 and ECMWF ERA-40) with the regional spectral model to generate CLARReS10_R2 and CLARReS10_ERA over the domain of the SEUS at a horizontal resolution of 10 km for the period of 1979–2001. Although both ECMWF ERA-40 and NCEP assimilate similar input data, they differ in the structure, parameterization and resolution of their data assimilation models. They also differ in their methods of processing input observations. The objective of the downscaling is to construct a high-resolution proxy dataset that can be beneficial to operational weather and climate prediction research by readily providing higher resolution data for verification. Although these products show some systematic errors, the downscaled reanalyses show good agreement with observations in terms of both the relative seasonal distribution and the diurnal structure of precipitation (Stefanova *et al*., 2012a; Misra *et al*., 2011).

Given the systematic errors associated with global and regional scale models, their output is often not directly applicable as input for rainfall–runoff models (RR). Therefore, a statistical bias-correction method is needed for correcting climate model output to produce long-term hydrological time series. Biases in RR models can result from the input data, the estimated model parameters or the simplifying assumption used in such models. These biases will then propagate through RR models, degrading the overall quality of the resulting simulation. Conse-

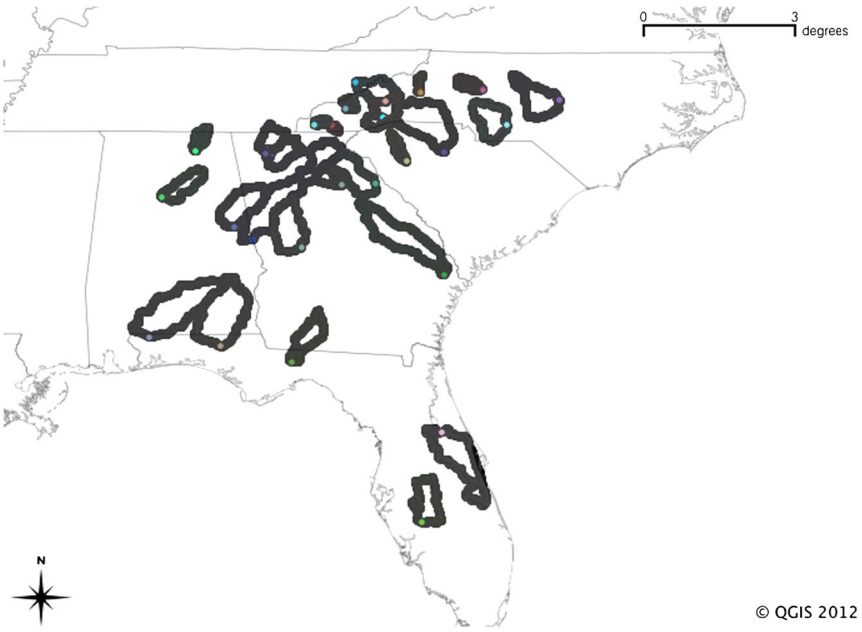

Figure 1. Location of the 28 watersheds within the southeastern United States used in this study

DOI: 10.1002/hyp

Table I. Description and characteristics of the 28 watersheds located within the region of southeastern United States

SN	Basin (USGS ID)	Lon	Lat	Area (sq mi)	Rain /pet	Annual ave runoff (cumec)	Runoff /rain	Annual rain (mm)	River system
1	2456500	−87.0	33.7	885	1.47	45.6	0.44	1425	Locust Fork at Sayre, AL
2	3574500	−86.3	34.6	320	1.53	17.3	0.45	1467	Paint Rock River near Woodville, AL
3	2414500	−85.6	33.1	1675	1.47	86.3	0.44	1425	Tallapoosa River at Wadley, AL
4	2296750	−81.9	27.2	1367	1.07	44.8	0.32	1248	Peace River at Arcadia, FL
5	2329000	−84.4	30.6	1140	1.29	49.3	0.39	1349	Ochlockonee River near Havana, FL
6	2365500	−85.8	30.8	3499	1.42	171.9	0.42	1425	Choctawhatchee River at Caryville, FL
7	2375500	−87.2	31.0	3817	1.46	201.2	0.43	1493	Escambia River near Century, FL
8	2236000	−81.4	29.0	3066	1.03	97.7	0.3	1293	St. Johns River near Deland, FL
9	2192000	−82.8	34.0	1430	1.4	65.7	0.42	1333	Broad River near Bell, GA
10	2202500	−81.4	32.2	2650	1.1	85.4	0.33	1189	Ogeechee River near Eden, GA
11	2217500	−83.4	33.9	392	1.5	19.9	0.44	1385	Middle Oconee River near Athens, GA
12	2347500	−84.2	32.7	1850	1.31	78	0.39	1317	Flint River near Culloden, GA
13	2383500	−84.8	34.6	831	1.55	48	0.46	1528	Coosawattee River near Pine Chapel, GA
14	2339500	−85.2	32.9	3550	1.48	189.3	0.44	1475	Chattahoochee River at West Point, GA
15	2387000	−84.9	34.7	687	1.55	37.2	0.46	1433	Conasauga River at Tilton, GA
16	2387500	−84.9	34.6	1602	1.53	87.6	0.45	1480	Oostanaula River at Resaca, GA
17	2102000	−79.1	35.6	1434	1.24	51	0.37	1171	Deep River at Moncure, NC
18	2118000	−80.7	35.8	306	1.47	13.9	0.44	1257	South Yadkin River near Mocksville, NC
19	2126000	−80.2	35.1	1372	1.24	48.9	0.37	1173	Rocky River near Norwood, NC
20	2138500	−81.9	35.8	67	1.77	4	0.51	1436	Linville River near Nebo, NC
21	3443000	−82.6	35.3	296	2.68	33.5	0.64	2156	French Broad River at Blantyre, NC
22	3451500	−82.6	35.6	945	2.34	70.7	0.59	1544	French Broad River at Asheville, NC
23	3504000	−83.6	35.1	52	2.08	4.6	0.57	1895	Nantahala River near Rainbow Springs, NC
24	3512000	−83.4	35.5	184	1.94	14	0.54	1720	Oconaluftee River at Birdtown, NC
25	3550000	−84.0	35.1	104	1.92	8.5	0.54	1846	Valley River at Tomotla, NC
26	2156500	−81.4	34.6	2790	1.58	139.1	0.46	1319	Broad River near Carlisle, SC
27	2165000	−82.2	34.4	236	1.38	10.6	0.41	1340	Reedy River near Ware Shoals, SC
28	3455000	−83.2	36.0	1858	2.14	114.5	0.56	1340	French Broad River near Newport, TN

quently, the biases in model outputs limit their use in water resources decision making.

In this study, quantile-based bias-correction method (Wood *et al*. 2002) is implemented to correct for systematic bias in rainfall. For each grid, the cumulative distribution function (CDF) of observed and regional reanalysis datasets is derived. The empirical CDF is applied to precipitation variables.

$$\hat{X}^{m}_{i,t} = F^{-1}{}_{\text{obs}}\left(F_{\text{mod}}\left(X^{m}_{i,t}\right)\right) \quad (1)$$

where $\hat{x}^{m}{}_{i,t}\, x^{m}{}_{i,t}$ are the tth corrected and uncorrected estimates of a variable i, t is the time step, and m is the month for the selected grid. $F_{\text{obs}}(\cdot)$ and $F_{\text{mod}}(\cdot)$ are the empirical CDFs of the observed and modelled datasets for the same grid.

The method is implemented independently on each grid, for each month, and for each variable. The method maps modelled values to observed values; therefore, extreme values outside the observed values cannot be obtained. As this analysis is carried out for a historical period, we did not account for extreme values outside the observed values, which otherwise could have been important for correcting biases for future climate scenarios.

For the evaluation of potential rainfall datasets for hydrological application, the following nine datasets (Table II), constructed from five types of dataset discussed earlier, were considered: (1) NCEP-R2 (OriR2; Kanamitsu *et al*. 2002), (2) ECMWF ERA-40 (OriERA; Uppala *et al*. 2006), (3) CLARReS10_R2 (DSR2; Stefanova *et al*. 2012b), (4) CLARReS10_ERA (DSERA; Stefanova *et al*. 2012a), (5) bias-corrected CLARReS10_R2 (BC_DSR2), (6) bias- corrected CLARReS10_ERA (BC_DSERA, (7) bias-corrected R2 (BC_OriR2), (8) bias- corrected ERA40 (BC_OriERA) and (9) synthetically generated data (using a weather generator, WGEN; Richardson and Wright 1984). To generate model forcing for hydrological model from gridded simulated rainfall, spatially averaged rainfall over each watershed is computed.

The weather generator model WGEN (Richardson and Wright, 1984) used in this study is a stochastic model that has been previously applied across the United States (e.g. Richardson and Wright, 1984). It is a statistical model used widely to generate daily sequence of weather variables. WGEN has two sets of parameters: one related to occurrence of precipitation and the other related to estimation of amounts of precipitation. It uses a first-order Markov process that requires two parameters, namely probabilities of wet day following wet day and wet day following dry day, to model precipitation occurrence. WGEN uses a two-parameter gamma distribution to model the distribution of rainfall amounts. The two parameters, shape and scale, for each location are determined from the observed records. In this study, the parameters of WGEN are estimated on the basis of aerial average rainfall taken from MOPEX project.

DOI: 10.1002/hyp

Table II. Rainfall data set used in this study

SN	Data set	Resolution (°)	Name	Remarks
1	NCEP-R2	2.5*2.5	OriR2	Raw
2	ECMWF ERA-40	2.5*2.5	OriERA	Raw
3	CLARReSS10_R2	0.1*0.1	DSR2	Downscaled (Stefanova *et al.* 2012b)
4	CLARESSeS10_ERA	0.1*0.1	DSERA	Downscaled (Stefanova *et al.* 2012b)
5	BC_NCEP-R2	0.25*0.25	OriR2_BC	Bias-corrected raw data
6	BC_ECMWF ERA-40	0.25*0.25	OriERA_BC	Bias-corrected raw data
7	BC_CLARReSS10_R2	0.25*0.25	DSR2_BC	Downscaled and bias corrected
8	BC_CLARESSeS10_ERA	0.25*0.25	DSERA_BC	Downscaled and bias corrected
9	Synthetically generated	Watershed scale	WGEN	

Hydrological models

In hydrological modelling, uncertainty stems from a variety of sources. Despite their known limitations, conceptual rainfall–runoff models continue to be widely used for assessing the impacts of climate change on water resources and for projecting potential ranges of future climate change impacts (e.g. Bastola *et al.*, 2011b). In this study, the uncertainty in simulation is accounted for by combining simulations obtained from three conceptual model structures and their behavioural model parameters. Three RR models are used in this study: the hydrologic model (HYMOD), the Nedbør-Afstrømnings model (NAM), and the TANK model.

HYMOD (Wagener *et al.*, 2001) proposed by Boyle (2001) is a conceptual and lumped model. HYMOD uses a nonlinear tank connected with two series of linear tanks in parallel to model rainfall excess mechanism. The NAM (Madsen 2000), developed at the Institute of Hydrodynamics and Hydraulic Engineering at the Technical University of Denmark, is another widely used conceptual lumped rainfall–runoff model. The model accounts for the water content in a number of different but interrelated storages, i.e. surface zone, root-zone and groundwater. Similarly, the TANK model used in this study is based on Sugawara (1995). It uses a number of vertical tanks, with side and bottom outlets, that are arranged vertically in series to model infiltration and saturated and unsaturated flows, through flow at watershed scale.

In all the three models described earlier, the snowmelt component of hydrological cycle is modelled using simple degree approach. In this approach, temperature is used as parameter for modelling snowmelt. And for this study, it is assumed that when air temperature is greater than zero, the precipitation falls as rain, and when air temperature is below zero, it falls as snow. The snowmelt is then calculated on the basis of the degree-day coefficient CS (mm/°C/day) and temperature. The snowmelt is allowed to store in a simple storage tank.

The rainfall–runoff models, discussed earlier, require the calibration of key parameters to yield reliable predictions (Gupta *et al.*, 2003). In the present application, HYMOD, NAM and TANK require 6, 10 and 16 parameters, respectively, to be estimated through model calibration. Several studies discuss problems associated with model calibration and parameter uncertainty (e.g. Kuczera, 1997; Beven *et al.*, 1992; Duan *et al.*, 1992). Hydrological modelling literature in general agrees that large combinations of parameters can result in equally acceptable model simulation (Beven, 2005). Therefore, any investigation that intends to evaluate different types of model forcing needs to address this problem, as the modelling inferences will be conditional upon the selected parameter. We therefore implement a multimodel and multiparameter simulation to conduct a conclusive evaluation of the meteorological forcing data. The generalized likelihood uncertainty estimation (GLUE) method (Beven and Binley, 1992) is implemented to account for parametrically and structurally different hydrological models, i.e. multiparameter and multimodel. In GLUE, the ensemble simulation is constructed by weighting the model prediction on the basis of each model's likelihood measure (Beven and Binley, 1992).

The methodology for the construction of ensemble simulation is briefly outlined here.

(a) For the selected models, the prior distribution or the prior range of model parameter is chosen. For simplicity, prior distribution is represented through uniform distribution from a specified range of values for each of the model parameters (Table III).
(b) The measure to evaluate model performance and a threshold that can be used to differentiate between behavioural and non-behavioural solution is selected. The threshold acts as a minimum performance acceptable for a set of parameters. Typical values of 0.5 have been selected in a number of studies (e.g. Tolson and Shoemaker 2008). In this study, Nash–Sutcliffe efficiency criterion is selected as the goodness-of-fit measure.
(c) For each of the selected model, simulations are made with parameters sampled from its prior distribution. Each simulation is then associated with a model performance. For each model simulation and the corresponding parameter set that result in model performance, a value greater than a specified threshold value is retained.
(d) The model performance value for retained simulation for the entire model is then rescaled so that their cumulative sum equals 1. Subsequently, this rescaled likelihood value that is associated with each retained simulation is subsequently used to produce likelihood weighted ensemble output and range of prediction.

DOI: 10.1002/hyp

Table III. Description of parameters and the range of values used for model calibration

SN	Model	Parameters	Range	Definition
1		*Cmax*	1–500	Maximum storage capacity
2		*Bexp*	0.1–2	A measure of spatial variability of soil moisture storage
3	HYMOD	*Alpha*	0.1–0.99	A factor that distributes water between slow and fast flow reservoir
4		*Kq*	0.001–0.1	Time constant parameter for quick flow reservoir
5		*Ks*	0.1–0.99	Time constant parameter for slow flow reservoir
6		*CS*	2.–4.	Degree day coefficient
7		*Umax*	10–20	Maximum storage capacity of surface storage
8		*Lmax*	50–250	Maximum storage capacity of lower zone storage
9		*CQOF*	0.01–0.99	Runoff coefficient for overland flow
10		*TOF*	0.0–0.7	Threshold for overland flow
11	NAM	*TIF*	0.0–0.7	Threshold for interflow
12		*TG*	0.0–0.7	Threshold for recharge
13		*CK IF*	500–1000	Time constant for interflow
14		*CK1, 2*	3.–48.	Time constant for overland flow
15		*CK BF*	500–5000	Time constant for groundwater reservoir
16		*CS*	2. –4.	Degree day coefficient
17		*A11*	0.01–0.8	Top tank, bottom outlet coefficient
18		*A12*	0–0.8	TOP tank, bottom side outlet coefficient
19		*A13*	0–0.8	TOP tank, top side outlet coefficient
20		*A21*	0–0.8	Middle tank, bottom outlet coefficient
21		*A22*	0–0.8	Middle tank, side outlet coefficient
22		*A31*	0–0.8	Bottom tank, bottom outlet coefficient
23		*A32*	0–0.8	Bottom tank, side outlet coefficient
24		*A4*	0–0.8	Groundwater storage coefficient
25		*H11*	0.1–100	Height of side outlet of TOP tank
26	TANK	*H12*	0.1–5	Height of side outlet of TOP tank
27		*H13*	0.1–6	Height of side outlet of TOP tank
28		*H22*	0.1–7	Height of side outlet of middle tank
29		*H33*	0.1–8	Height of side outlet of bottom tank
30		*K1*	0.0–0.5	Coefficient that control flow between top and middle tank
31		*K2*	0.0–0.5	Coefficient that control flow between primary and secondary storage of top tank
32		*CS*	2. –4.	Degree day coefficient

The prediction range is constructed using Equation 2 (Beven and Freer 2001).

$$P\left(\hat{Z}_t < z\right) = \sum_{i=1}^{n} L\left[f(\theta_i)\middle|\hat{Z}_{t,i} < z\right] \tag{2}$$

where $\hat{Z}$ is the variable simulated by the model at time t, $f(\theta_i)$ is the ith set of behavioural model parameter, L is the rescaled likelihood value and n is the total number of behavioural simulation, i.e. sum of behavioural parameters for set of selected model.

The hydrological simulation is run at a daily time step using the three conceptual models calibrated for the periods 1949–1959 and validated for the periods 1960–1970.

The nine datasets introduced in the methodology section are referred to as test datasets. The control data (e.g. observed streamflow or streamflow simulated with observed rainfall) are defined as a reference dataset, against which the test datasets are evaluated. In this study, the test datasets are evaluated by analyzing the following: (1) discrepancies between the observed streamflow and the outputs of the model forced with the test data (model error hereafter) and (2) discrepancies between streamflow simulated with observed rainfall and the output obtained from the model forced with test data (model-propagated error hereafter).

The control data are the model forcing data used for calibrating the parameter of RR models so that the differences between observed and simulated streamflow values are minimal. In this study, the closeness of the fit is evaluated quantitatively using the following objective criteria: (a) the Nash–Sutcliffe efficiency index (NSE) (Equation 3), which reflects the overall agreement of the shape of the hydrograph, and (b) volume error (VE) (Equation 4).

$$NSE = 1 - \sum_{i=1}^{n} \left(Q_{\text{obs}\ i} - Q_{\text{sim}\ i}\right)^2 / \sum_{i=1}^{n} \left(Q_{\text{obs}\ i}\right)^2 \tag{3}$$

$$VE = \sum_{i=1}^{n} \left(Q_{\text{sim}\ i} - Q_{\text{obs}\ i}\right) / Q_{\text{obs}\ i} \tag{4}$$

where Q_{obs} and Q_{sim} are the observed and simulated stream flows. Q_{obs} refers to the measured streamflow while estimating model error and the model-simulated flow with control input while estimating model-propagated error. Control flow is the model simulated flow for the evaluation period with measured forcing.

DOI: 10.1002/hyp

The values of parameters estimated through model calibration are sensitive to the period and climatic condition for which the model is calibrated (Bastola *et al.*, 2011a). Therefore, the simulation for a period different from the one used for model calibration is likely to result in biases. This study adopts evaluation of the datasets based on model propagation error (i.e. comparison of a simulation with the control simulation). If the differences between the model's propagated error and the model error are small, then the bias associated with the calibrated model parameter is minimal and can be neglected. However, if the bias is markedly high, then the model needs to be recalibrated with the dataset and the period that is consistent with the test data period.

RESULTS AND DISCUSSION

Table IV summarizes the performances for the three models and the multimodel mean. For the watershed used in this study, the uncertainty in prediction associated with TANK model is the highest among models. Similarly, for NAM model, the uncertainty in the model is the lowest and the reliability in prediction is the highest. For both calibration and validation, combining the output from three models resulted in wider prediction intervals (PI; ~23% increase in PI) and consequently resulted in improved reliability (~17% increase in reliability).

Figure 2 shows the median NSE, PI and count efficiency (CE) for the selected watersheds and hydrological model for both the calibration and the validation periods. The CE is estimated as the percentage of observation included within the PI, and the PI is estimated as the average width of the PI. It is apparent from the figure that the combination of output from different model resulted in improvement in model performance and reliability of model (measured as CE).

The average model performance (i.e. NSE) is nearly 72%, and the average VE is around 10% for both calibration and validation (not shown). For watersheds 2236000 (FL) and 216500 (NC), the reliability of a 90% PI in encapsulating the observation is far less than for the other watersheds, indicating that the output is less reliable. For calibration, the model is run several times, testing with a large number (~20 000) of parameters that are randomly generated from their range assuming that their prior distribution follows uniform distribution. All parameter sets that result in performance greater than 0.5 are considered behavioural parameter sets. The selection of 0.5 for performance is made to accommodate the fact

Table IV. Performance of hydrological models (aggregated across watersheds) for calibration and validation period

SN	Scheme	NSE	CE	PI	Period
1	Multimodel	0.706	0.836	0.819	Calibration
2	HYMOD	0.685	0.697	0.659	
3	NAM	0.666	0.726	0.615	
4	TANK	0.651	0.724	0.739	
5	Multimodel	0.689	0.833	0.944	Validation
6	HYMOD	0.679	0.714	0.765	
7	NAM	0.643	0.689	0.703	
8	TANK	0.634	0.723	0.839	

NSE, Nash–Sutcliffe efficiency; CE, count efficiency; PI, prediction interval

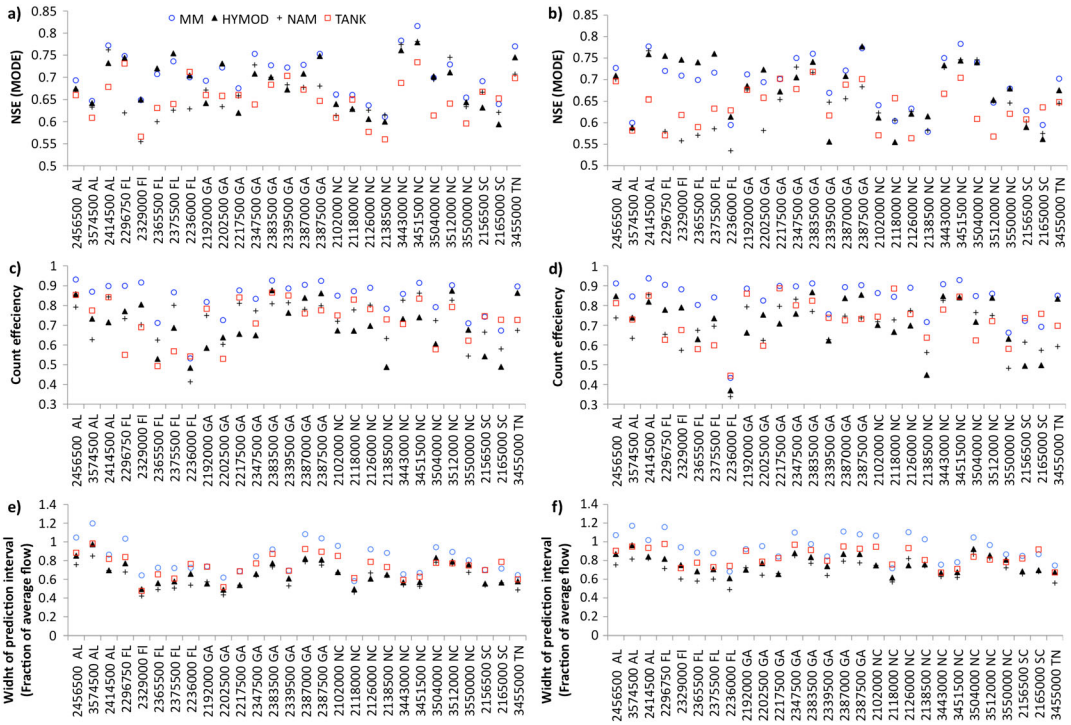

Figure 2. Performances of the three hydrological models, namely HYMOD, including multimodal mean during calibration and validation (a and b) Nash–Sutcliffe efficiency index (NSE) for calibration and validation period, (c and d) count efficiency, i.e. fraction of observations that are encapsulated within the prediction range for calibration and validation period, and (e and f) average width of the prediction range expressed as a fraction of average streamflow for calibration and validation

DOI: 10.1002/hyp

that the marginal increase in the threshold value does not significantly decrease the reliability of the PI in encapsulating the observations.

The uncertainty associated with prediction, mainly arising from parameter and model selection, is on average (across all 28 watersheds) nearly two thirds of the average flow for both calibration and validation period. A seasonal streamflow plot (1948–1968), for two watersheds, i.e. 3574500 and 2236000, showing ensemble average and upper 95% and lower 5% stream flow for each of the model and ensemble mean of the models is shown in Figure 3. For each of the individual model, the inability of the 90% PI to encapsulate 90% of observation points reflects that parameter uncertainty alone is not sufficient to explain the uncertainty in simulation. The uncertainty grew with the size of the watershed, whereas the relationship of uncertainty of model response to total precipitation and the ratio of precipitation to evapotranspiration were not clear (not shown). For all chosen watersheds (Figure 1), however, the uncertainty in seasonal streamflow prediction shows the tendency to decrease with an increase in latitude.

The seasonal flow and PI (shaded grey) estimated for both the calibration and the validation periods for the selected watersheds are shown in Figure 4. On the basis of model performances, a set of behavioural model parameters for each model and catchment is identified. Subsequently, calibrated models are then forced with five different types of datasets [i.e. global (OriR2 and OriERA); dynamically downscaled (DSR2 and DSERA); and bias-corrected global (BC_OriR2 and BC_OriERA); bias corrected dynamically downscaled (BC_DSR2 and BC_DSERA); and synthetically generated data using WGEN]. For the synthetic data, the parameters of WGEN are estimated using meteorological data (spatially average at watershed scale) from all the selected watersheds spanning a 30-year period. The WGEN is capable of generating large sets of statistically consistent time series; therefore, 100 sets of precipitation time series are generated, and the median hydrological response is used for evaluation purpose.

For comparing the different types of datasets, we used nine indices from the statistical and regional dynamical downscaling of extremes for European regions project. We compare mean climatological precipitation (pav; mm/day), 90th percentile of rain day amounts (pq90; mm/day), the greatest 5-day total rainfall (px5d), simple daily intensity (pint), mean wet day/dry day persistence (ppww/ppdd), % of total rainfall from events greater than long-term P90 (pfl90), number of events greater than long-term 90th percentile of rain days (pnl90) and maximum number of consecutive dry days (pxcdd)(Figure 5). Compared with oriR2 and OriERA, some of the characteristics such as average rainfall, 90th percentile rainfall, wet day persistence, the greatest 5-day total, average intensity, 90th percentile rain day amounts and number of events greater than long-term are well reproduced by regionally downscaled data. The downscaled data however overestimated the maximum number of consecutive dry days. With regard to WGEN output, the average precipitation, wet day intensity, 90th percentile rainfall amount, and wet and dry day persistence are reproduced reasonably well for the selected watersheds. WGEN, however, poorly simulated the maximum number of consecutive dry days (not shown).

Figure 6 shows the seasonal rainfall and corresponding flow simulated with different model forcing datasets (over six watersheds of SEUS). For all watersheds except those in Florida, the precipitation from R2 is associated with a stronger seasonal cycle and with overestimation during summer and underestimation during winter. However, the bias associated with ERA is less than the bias associated with R2. Bias in simulated flow is reduced for DSR2 and DSERA as precipitation associated with smaller scale phenomena are accounted for in the high-resolution dynamically downscaled regional reanalysis data. The statistical bias correction applied to dynamically downscaled regional reanalysis results in significant improvement (see upper panel in Figure 6). Moreover, the seasonal cycle of (both dynamically and statistically) bias-corrected rainfall is in closer agreement with the observation than OriR2 and OriERA. At watershed scale,

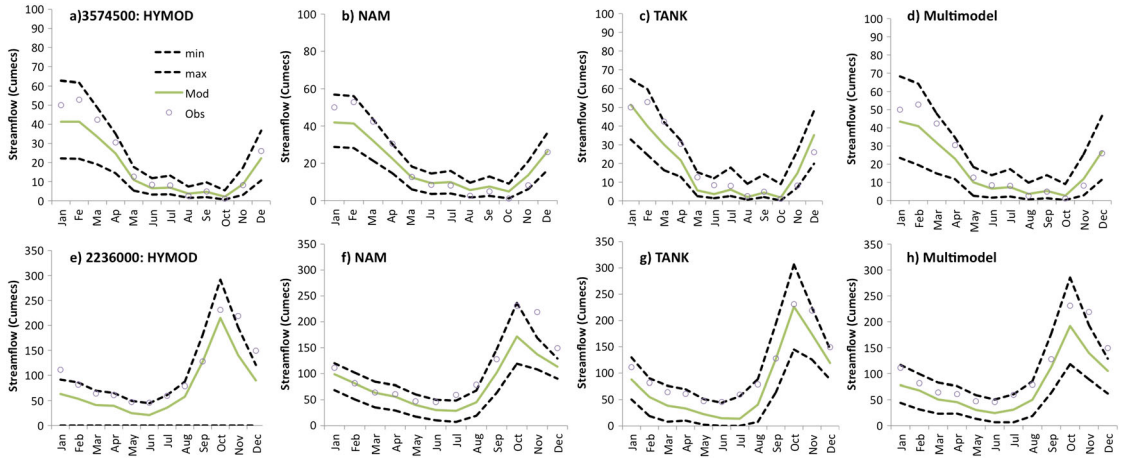

Figure 3. The upper 95%, lower 5% and median value of simulated streamflow (1948–1968) for four schemes, i.e. HYMOD, NAM, TANK and multimodel, and two watersheds (a–d) 3574500 and (e–h) 2236000

DOI: 10.1002/hyp

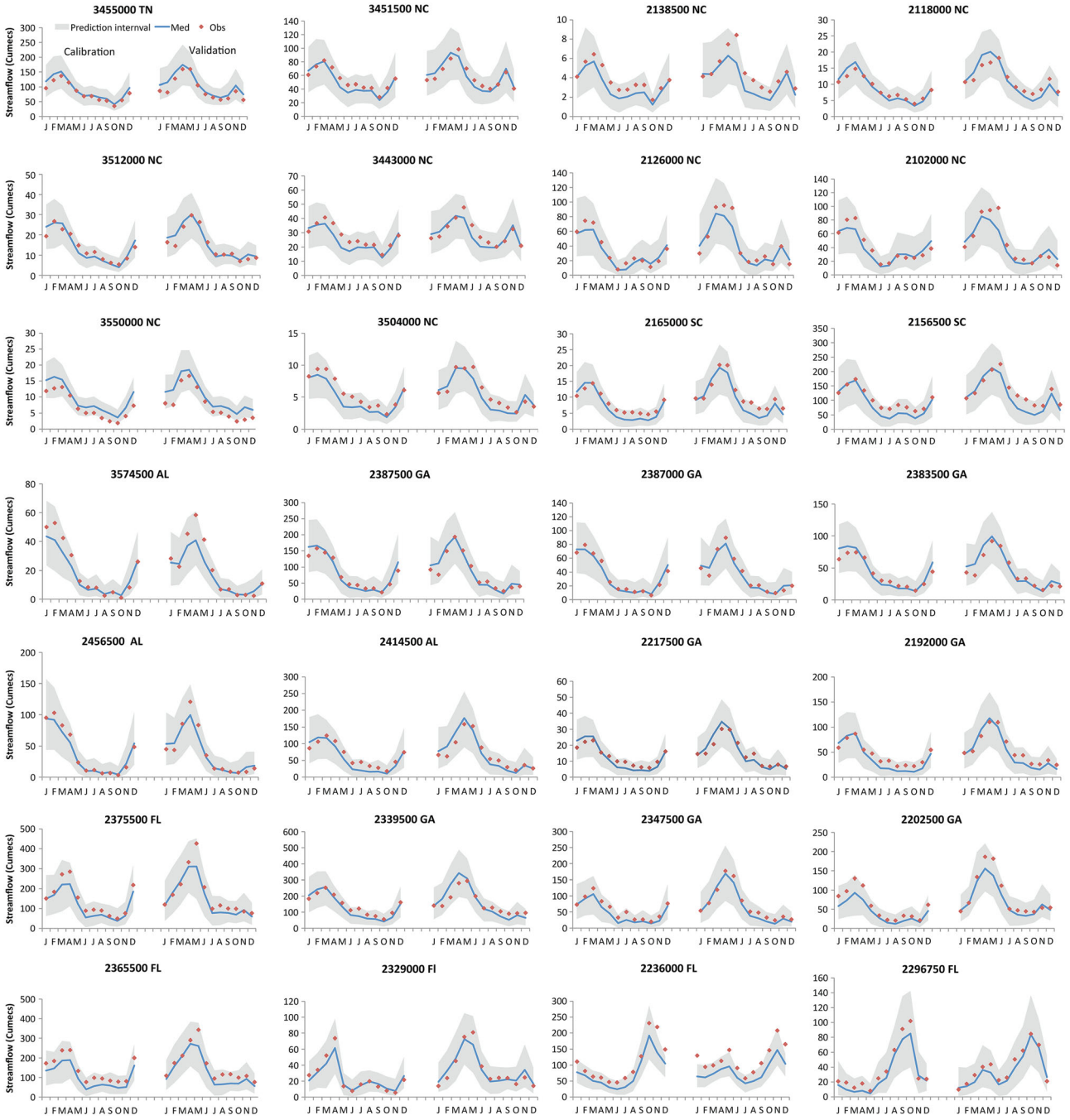

Figure 4. Observed and simulated seasonal streamflow and prediction uncertainties (shaded grey) for the 28 selected watersheds in the southeastern United States during model calibration and validation

the downscaling does reduce the bias in precipitation, a primary input of hydrological model.

In addition to the qualitative assessment, quantitative comparisons based on the NSE and overall VE is made on flows simulated with a suite of hydrological models. The best-performing dataset is the one that results in simulations that are closer to the observed streamflow (or the flow simulated with observed forcing), i.e. near the ideal performance point ($NSE = 1$ and $VE = 0$).

As shown in Table V, which shows the aggregate performances of different datasets in simulating the streamflow measured with respect to observation and control simulation, on average OriR2 overestimates the flow, whereas OriERA underestimates the flow (measured in terms of VE). Compared with the bias in OriERA and OriR2, the VE in response to WGEN forcing is significantly smaller and is comparable with the bias in dynamically downscaled data. However, in general, compared with WGEN, dynamically downscaled data result in more realistic hydrologic simulation. The performance improves with the use of the downscaled reanalysis dataset, both in terms of NSE and VE. Statistical bias correction results in further improvement in performance of the dynamically downscaled reanalysis data.

Figure 7(a, b) (Figure (7c, d)) shows the comparison, for each of the watershed, with respect to observed flow

DOI: 10.1002/hyp

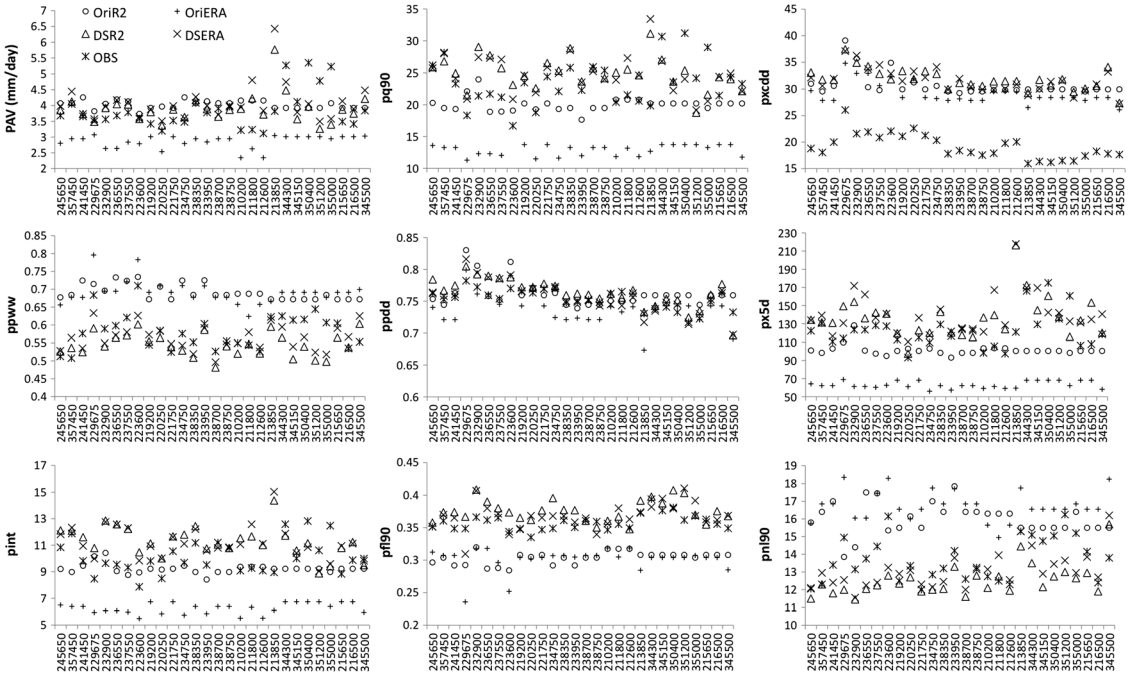

Figure 5. Comparison of the reanalysis, dynamically downscaled reanalysis and observed daily rainfall using indices from the statistical and regional dynamical downscaling of extremes for European regions project: mean climatological precipitation (pav; mm/day), 90th percentile of rain day amounts (pq90; mm/day), the greatest 5-day total rainfall (px5d), simple daily intensity (pint), mean wet day/dry day persistence (ppww/ppdd), % of total rainfall from events greater than long-term P90 (pfl90), number of events greater than long-term 90th percentile of rain days (pnl90) and maximum number of consecutive dry days (pxcdd)

(flow simulated with observed forcing). It is apparent from Figure 7(a, c) that the points in performance space are much more scattered and farther away from ideal points for OriR2 (can also be seen from Table V). The flow simulated with OriR2 overestimates the control flow. Apart from the seasonal average value, the OriR2 tends to overestimate (underestimate) heavy (light) rainfall event (Stefanova *et al*., 2012a). On the other hand, the performance of OriERA for all 28 watersheds are similar but with systematic underestimation of the flow in all the selected watersheds. The underestimation of the flow can also be attributed to the fact that, over our study region, OriERA tends to overestimate (underestimate) the frequency of light (heavy) rainfall (Stefanova *et al*., 2012a). Furthermore, Figure 7(b, d) shows improvement in performance, as points move closer towards the ideal point (i.e. 1.0), with bias correction. For comparing dynamically downscaled regional reanalysis data with data extracted from synthetically generated dataset, i.e. WGEN, we used monthly output instead of daily output.

The result based on monthly simulated flow (Table V) reflects that the NSE improved when model performance is evaluated on the basis of monthly aggregated result instead of daily. In average, the downscaled and bias-corrected global reanalysis data clearly out formed other dataset at when model performance is evaluated on the basis of daily simulated flow.

Figure 8(a, b) shows the performance based on monthly aggregated simulated flow, for the entire selected watershed, of the statistically downscaled and dynamically downscaled regional reanalysis dataset derived on the basis of R2 and ERA, respectively. It is apparent from the figure that WGEN performed better than OriR2 and OriERA (Table V). But the points for DSR2_BC and DSERA_BC are closer to the ideal point compared with WGEN [the average NSE for DSR2_BC and DSERA_BC is greater than WGEN (Table V)]. In comparison with OriR2 and OriERA, the performance of WGEN is better for hydrological simulation in the SEUS. Moreover, the performance of bias-corrected, dynamically downscaled regional reanalysis data is superior to the performance of bias-corrected global reanalysis data, dynamically downscaled regional reanalysis data without bias correction and synthetically generated data.

The watersheds selected for this study (Table I) vary in size and other characteristics. Therefore, the VE and performance of simulation corresponding to each of the test input vary from watershed to watershed. The bubble plot (Figure 9) shows the spatial distribution of a value of VE and NSE for the 28 watersheds for ERA-40 (a and b) and NCEP-R2 (c and d). The performance is measured with respect to flow simulated with observed model forcing. For VE, the smaller the size of the bubble, the better the simulation, and for NSE, the larger the size of the bubble, the better the simulation. For the five watersheds located within Florida, bias-corrected regionally downscaled reanalysis data perform better than the global reanalysis and corresponding regionally downscaled reanalysis. For watersheds located within North Carolina, ERA-40 performs better than NCEP-R2. In these watersheds, the bias correction shows marked improvement in model simulation.

DOI: 10.1002/hyp

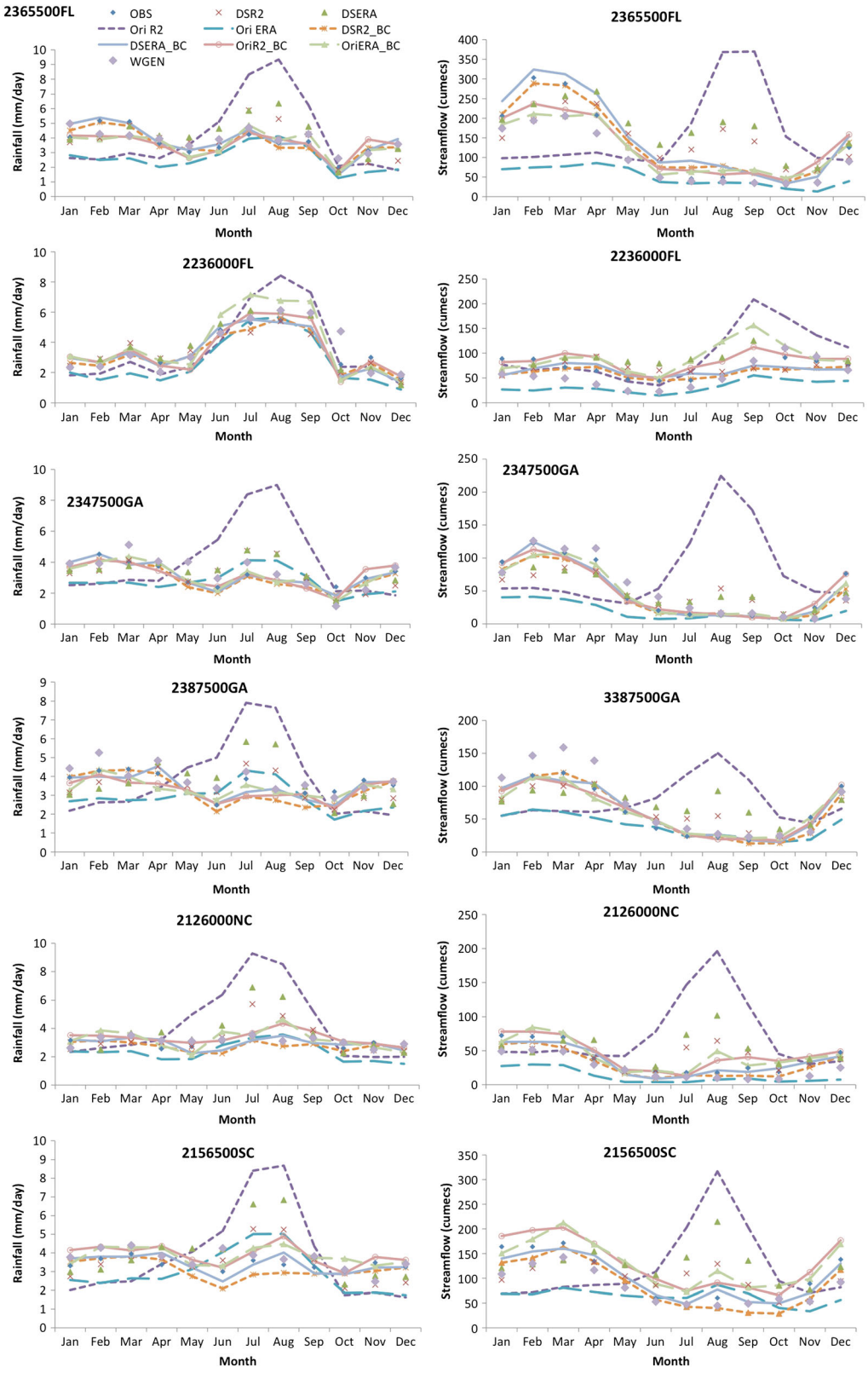

Figure 6. Seasonal aerial average precipitation for the selected watersheds derived from different test precipitation datasets: global reanalysis (OriR2 and OriERA), dynamically downscaled reanalysis (DSR2 and DSERA), bias-corrected, dynamically downscaled reanalysis (BC_DSR2 and BC_DSERA) and synthetically generated (WGEN) and corresponding simulated streamflow

CONCLUSION

In this study, precipitation data derived from regional (downscaled) reanalysis are evaluated for their skill in driving a hydrological model of the SEUS. The watersheds analyzed in this study were selected because of their relatively minimal anthropogenic influence on streamflow regulation in the SEUS. This study shows

DOI: 10.1002/hyp

Table V. summary of model's performance

SN	Precipitation dataset	Based on daily simulated flow				Based on monthly simulated flow	
		Model error		Model propagated error		Model propagated error	
		NSE	VE	NSE	VE	NSE	VE
1	OriR2	−0.927	0.185	−1.138	0.252	−1.952	0.252
2	OriERA	−0.092	−0.599	0.056	−0.563	−0.142	−0.563
3	DSR2	−0.021	−0.120	−0.004	−0.060	−1.293	−0.060
4	DSERA	−0.341	0.077	−0.578	0.160	−1.085	0.160
5	OriR2_BC	0.025	−0.040	0.190	0.030	0.246	0.030
6	OriERA_BC	−0.139	−0.047	−0.035	0.016	0.045	0.016
7	DSR2_BC	0.166	−0.241	0.393	−0.189	0.423	−0.189
8	DSERA_BC	0.138	−0.182	0.339	−0.123	0.477	−0.123
9	WGEN	−0.108	−0.160	−0.610	0.060	−0.185	0.060

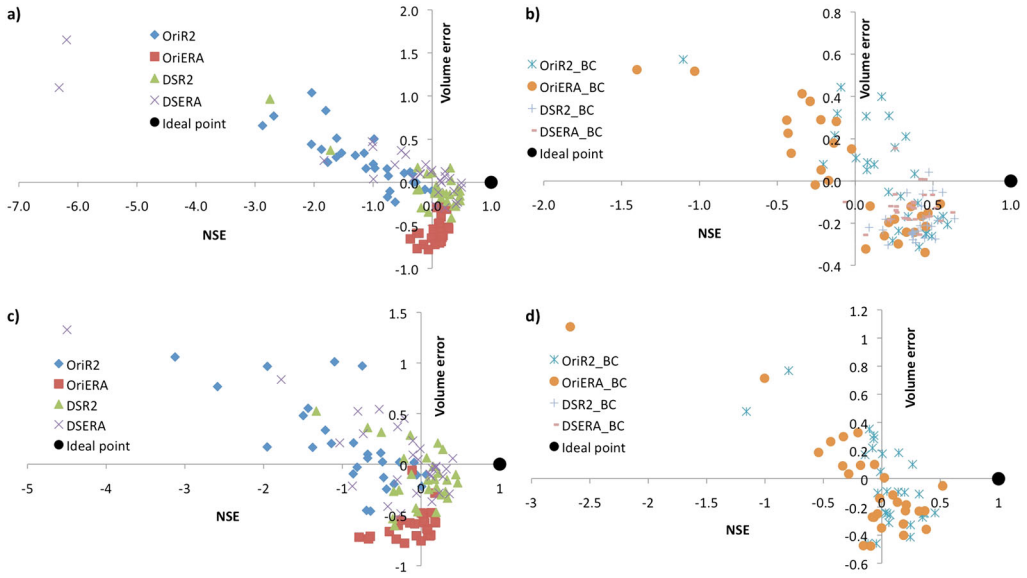

Figure 7. Performance of original and downscaled precipitation datasets in terms of model error, i.e. measured with respect to observed quantity (a and b) and model propagated error (c and d), i.e. measured with respect to the quantity simulated with control model forcing

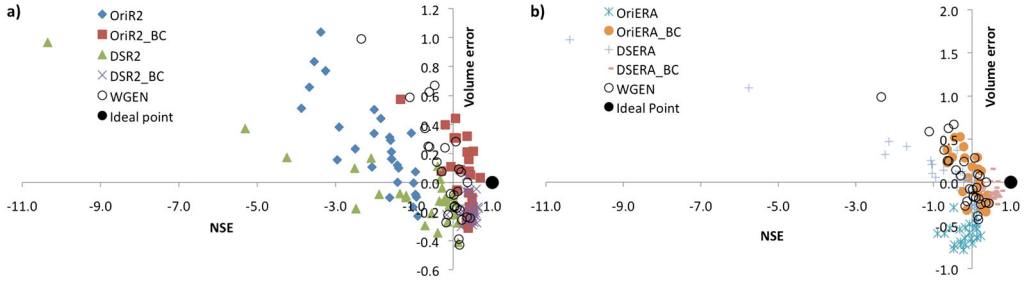

Figure 8. Performance of original and downscaled precipitation datasets in terms of model propagated error, i.e. measured with respect to quantity simulated with control model forcing. Observed quantity: (a) variants of R2 and WGEN and (b) variants of ERA and WGEN

that the global reanalysis of ERA40 tends to underestimate streamflow, whereas NCEP-R2 tends to overestimate it in these SEUS watersheds. We further show that forcing data for the RR models derived from the bias correction following the methodology of Wood *et al.* (2002) on the surface meteorology of these global reanalysis seem to outperform the corresponding regionally downscaled reanalysis product. Our work, however, showed that the regionally downscaled reanalysis without any bias correction did far better than the corresponding global reanalysis. Furthermore, this study also shows that the regional dynamically downscaled data with bias correction produced more realistic hydrologic simulation than synthetically generated data using WGEN. For

DOI: 10.1002/hyp

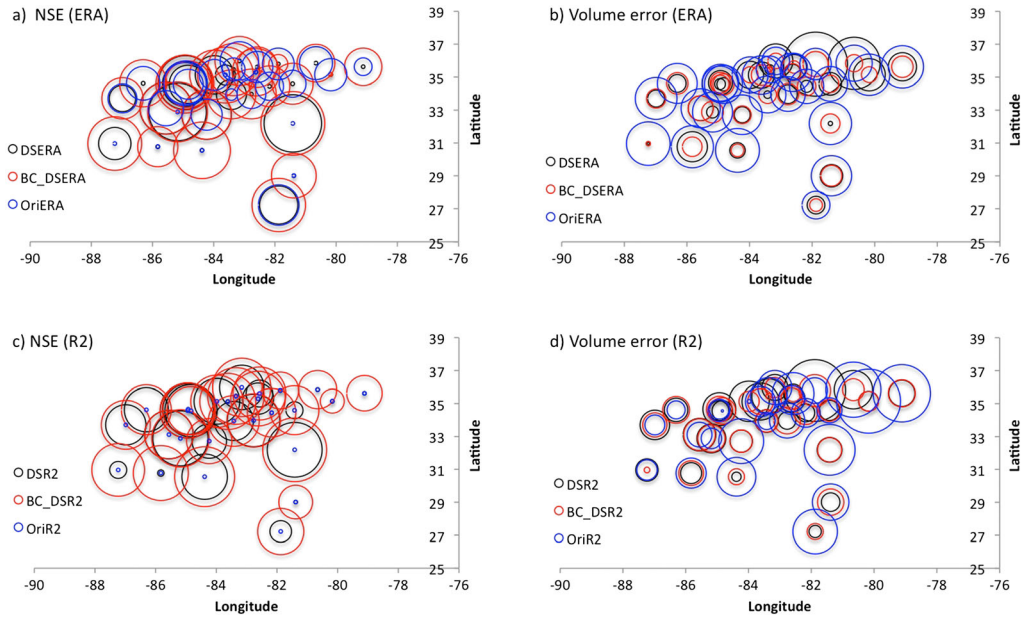

Figure 9. Bubble plot showing the value of Nash–Sutcliffe efficiency index (NSE) (a and c) and volume error (b and d) measured with respect to control simulation (i.e., model propagation error)

hydrological simulation, bias-corrected dynamically downscaled data from global reanalysis are superior to global reanalysis, bias-corrected global reanalysis, and dynamically downscaled global reanalysis without bias correction. From our comprehensive analysis of using multiple sources of meteorological forcing, deploying multiple hydrological models over several (28) watersheds, we conclude that for hydrological model application studies over the SEUS, there is merit in dynamically downscaling coarser global reanalysis.

Even though the weather generator resulted in better reproduction of seasonal cycle of precipitation and consequently of streamflow, the model performance evaluated at a daily time scale is as poor as that of the global reanalysis without bias correction. Such a significant loss in performance of hydrological model with the use of precipitation products other than used for model calibration implies that application that requires the use of output from GCM, e.g. hydrologic impact of climate change, should consider focusing on the seasonal to annual response from watershed instead of attempting to simulate the dynamics of river flow at a daily time step.

ACKNOWLEDGEMENTS

The authors acknowledge Kathy Fearon of COAPS for the help with editing the manuscript. This work was supported by NOAA grant NA07OAR4310221 and USGS grant 06HQGR0125. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the acknowledged funding agencies.

REFERENCES

Ajami NK, Duan Q, Gao X, Sorooshian S. 2006. Multimodel combination techniques for analysis of hydrological simulations: application to Distributed Model Intercomparison Project results. *Journal of Hydrometeorology* **7**: 755–768.

Bastola S, François D. 2012. Temporal extension of meteorological records for hydrological modelling of Lake Chad Basin (Africa) using satellite rainfall data and reanalysis datasets. *Meteorological Applications*. **19**(1): 57–70. DOI: 10.1002/met.257.

Bastola S, Murphy C, Sweeney J. 2011a. Evaluation of the transferability of hydrological model parameters for simulations under changed climatic conditions. *Hydrological Earth System Science Discussion* **8**: 5891–5915. DOI:10.5194/hessd-8-5891-2011.

Bastola S, Murphy C, Sweeney J. 2011b. The role of hydrological modelling uncertainties in climate change impact assessments of Irish river catchments. *Advance in water Resources* **34**(5): 562–576. DOI:10.1016/j.advwatres.2011.01.008, 2011.

Beven K. 2005. A manifesto for the equifinality thesis. *Journal of Hydrology* **20**: 1–19.

Beven K, Binley A. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrological Process* **6**: 279–298.

Beven K, Freer J. 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems. *Journal of hydrology* **249**: 11–29.

Boyle D 2001. Multicriteria calibration of hydrological models. PhD dissertation. Tucson, AZ: Department of Hydrology and Water Resources, University of Arizona.

Chan S, Misra V. 2010. A diagnosis of the 1979–2005 extreme rainfall events in the southeast US with isentropic moisture tracing. *Monthly Weather Review* **138**: 1172–1185.

Duan Q, Sorooshian S, Gupta VK. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* **28**(4): 1015–1031.

Fowler HJ, Blenkinsopa S, Tebaldi C. 2007. Review: linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology* **27**: 1547–1578.

Frei C, Christensen JH, Déqué M, Jacob D, Jones RG, Vidale PL. 2003. Daily precipitation statistics in regional climate models: evaluation and intercomparison for the European Alps. *Journal of Geophysical Research* **108**(D3): 4124. DOI: 10.1029/2002JD002287.

Georgakakos KP, Seo DJ, Gupta HV, Schake J, Butts MB. 2004. Characterizing streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology* **298**: 222–241.

Gupta HV, Beven K, Wagener T. 2003. In *Model Calibration and Uncertainty Estimation*, Anderson MG (ed.). John Wiley and Sons: Chichester, United Kingdom.

Hagemann,S, Machenhauer B, Jones R, Christensen OB, Déqué M, Jacob D, Vidale PL. 2004. Evaluation of water and energy budgets in regional climate models applied over Europe. *Climate Dynamics* **23**: 547–567.

HEPEX 2006. Hydrologic ensemble prediction experiment. Available at http://www.gewex.org/PAN-GEWEX-MTG/Pan-GEWEX_HEPEX-2006.pdf

Hongwei Y, Wang B, Wang B. 2012. Reduction of systematic biases in regional climate downscaling through ensemble forcing. *Climate Dynamics* **39**:2523–2532. DOI:10.1007/s00382-011-1006-4.

Kanamitsu M, Kanamaru H. 2007. Fifty-seven-year reanalysis downscaling at 10 km (CaRD10). Part I: system detail and validation with observations. *Journal of Climate* **20**:5553–5571

Kanamitsu M, Ebisuzaki W, Woollen J, Yang S-K, Hnilo JJ, Fiorino M, Potter GL. 2002. NCEP–DOE AMIP-II Reanalysis (R-2). *Bulletin of the American Meteorological Society* **83**: 1631–1643

Kuczera G. 1997. Efficient subspace probabilistic parameter optimization for catchment models. *Water Resources Research* **33**(1): 177–185.

Li H, Sheffield J, Wood EF. 2010. Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *Journal of Geophysical Research* **115**: D10101. DOI:10.1029/2009JD012882, 201

Madsen H. 2000. Automatic calibration of a conceptual rainfall–runoff model using multiple objectives. *Journal of Hydrology* **235**: 276–288

Mesinger F, DiMego G, Kalnay E, et al.. 2006. North American Regional Reanalysis. *Bulletin of the American Meteorological Society* **87**: 343–360.

Misra V, Moeller L, Stefanova S, Chan S, O'Brien JJ, Smith III TJ, Plant N. 2011. The influence of the Atlantic Warm Pool on the Florida panhandle sea breeze. *Journal of Geophysical Research-Atmospheres* **116**, D00Q06, DOI:10.1029/2010JD01.

Mooney PA, Fealy R, Mulligan FJ. 2011. Comparison of ERA-Interim, ERA-40 and NCEP/NCAR reanalysis data with observed surface air temperatures over Ireland. *International Journal of Climatology* **31**: 545–557. DOI: 10.1002/joc.2008.

Piani C, Haerter JO 2012. Two dimensional bias correction of temperature and precipitation copulas in climate models. *Geophysical Research Letters* (39): L20401, 6 pp. doi:10.1029/2012GL053839.

Richardson CW, Wright DA. 1984. WGEN: a model for generating daily weather variables. U.S. Dep. of Agric., Agric. Res. Service, ARS-8: Springfield, Virginia; 83.

Ropelewski CF, Halpern MS. 1987. The Southern Oscillation. Part V: the anomalies in the lower stratosphere of the Northern Hemisphere in winter and a comparison with the quasi-biennial oscillation. *Monthly Weather Review* **115**: 357–369.

Schaake J, Cong S, Duan Q. 2006. U.S MOPEX datasets, IAHS publication series (https://e-reports-ext.llnl.gov/pdf/333681.pdf).

Seager R, Burgman R, Kushnir Y, Clement A, Cook ER, Naik N, Miller J. 2008. Tropical Pacific forcing of North American Medieval megadroughts: testing the concept with an atmosphere model forced by coral-reconstructed SSTs. *Journal of Climate* **21**:6175–6190.

Slack JR, Lumb AM, Landwehr JM. 1993. Hydroclimatic Data Network: streamflow data set, 1874–1988, USGS Water Resources Investigations Report: 93–4076.

Stefanova L, Misra V, Chan S, Griffin M, O'Brien JJ, SmithIII TJ. 2012a. A proxy for high-resolution regional reanalysis for the southeast United States: assessment of precipitation variability in dynamically downscaled reanalyses. *Climate Dynamics* **38**: 2449–2446. DOI:10.1007.s00382-011-1230-y

Stefanova L, Misra V, O'Brien JJ, Chassignet EP, Hameed S. 2012b. Hindcast skill and predictability for precipitation and two-meter air temperature anomalies in global circulation models over the Southeast United States. *Climate Dynamics* **38**(1–2): 161–173. DOI:10.1007/s00382-010-0988-7.

Sugawara M. 1995. Tank model. In: Singh VP, editor. *Computer Models of Watershed Hydrology*. Littleton, CO: Water Resource Publication: 165–214.

Tolson BA, Shoemaker CA. 2008.Efficient prediction uncertainty approximation in the calibration of environmental simulation models. *Water resource research* **44** , p. WR005869

Uppala SM, et al. 2006. The ERA40 re-analysis. *The Quarterly Journal of the Royal Meteorological Society* **131**(612): 2961–3012. DOI: 10.1256/qj.04.176.

Von Storch H, Langenbert H, Feser F. 2000. A spectral nudging technique for dynamical downscaling purpose. *Monthly Weather Review* **128**: 3664–3673.

Wagener T, Boyle DP, Lees MJ, Wheater HS, Gupta HV, Sorooshian S. 2001. A framework for development and application of hydrological models. *Hydrology and Earth System Science* **51**: 13–26.

Wallis JR, Lettenmaier DP, Wood EF. 1991. A daily hydroclimatological data set for the continental United States. *Water Resource Research* **27**(7): 1657–1663.

Wang H, Fu R, Kumar A, Li WH. 2010. Intensification of summer rainfall variability in the southeastern United States during recent decades. *Journal of Hydrometeorology* **11**: 1007–1018.

Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO. 2004. The guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the Intergovernmental Panel on Climate Change (IPCC), prepared on behalf of Task Group on Data and Scenario Support for Impacts and Climate Analysis (TGICA).

Wilby RL, Hay LE, Gutowski Jr WJ, Arritt RW, Takle ES, Pan Z, Leavesley,GH, Clark MP. 2000. Hydrological responses to dynamically and statistically downscaled climate model output. *Geophysical Research Letter* **27**(8): 1199–1202.

Wood AW, Leung LR, Sridhar V, Lettenmaier DP. 2004. Hydrologic implications of dynamical and statistical approaches to downscale climate model outputs. *Climatic Change* **62**: 189–216.

Wood AW, Maurer EP, Kumar A, Lettenmaier D. 2002. Long-range experimental hydrologic forecasting for the eastern United States. *Journal of Geophysical Research* **107**(D20): 4429, doi:10.1029/2001JD000659, 2002.

Yao H, Georgakakos A. 2001. Assessment of Folsom Lake response to historical and potential future climate scenarios. *Journal of Hydrology* **249**: 176–196.

Zhang F, Georgakakos A. 2011. Joint variable spatial downscaling. *Climatic Change* **111**(3–4): 945–972. DOI: 10.1007/s10584-011-0167-9

Xu CY. 1999. From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. *Progress in Physical Geography* **23**(2): 229–249.