ECOHYDROLOGY
Ecohydrol. **8**, 1525–1546 (2015)
Published online 23 February 2015 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/eco.1602

A comparison of hydrologic models for ecological flows and water availability

Peter V. Caldwell,[1]* Jonathan G. Kennen,[2] Ge Sun,[3] Julie E. Kiang,[4] Jon B. Butcher,[5] Michele C. Eddy,[6] Lauren E. Hay,[7] Jacob H. LaFontaine,[8] Ernie F. Hain,[9] Stacy A. C. Nelson[9] and Steve G. McNulty[3]

[1] *Center for Forest Watershed Science, USDA Forest Service, Otto, NC, USA*
[2] *US Geological Survey, Lawrenceville, NJ, USA*
[3] *Eastern Forest Environmental Threat Assessment Center, USDA Forest Service, Raleigh, NC, USA*
[4] *US Geological Survey, Office of Surface Water, Reston, VA, USA*
[5] *Tetra Tech, Research Triangle Park, NC, USA*
[6] *RTI International, Research Triangle Park, NC, USA*
[7] *US Geological Survey, Lakewood, CO, USA*
[8] *US Geological Survey, Norcross, GA, USA*
[9] *Center for Geospatial Analytics, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, 27695, USA*

ABSTRACT

Robust hydrologic models are needed to help manage water resources for healthy aquatic ecosystems and reliable water supplies for people, but there is a lack of comprehensive model comparison studies that quantify differences in streamflow predictions among model applications developed to answer management questions. We assessed differences in daily streamflow predictions by four fine-scale models and two regional-scale monthly time step models by comparing model fit statistics and bias in ecologically relevant flow statistics (ERFSs) at five sites in the Southeastern USA. Models were calibrated to different extents, including uncalibrated (level A), calibrated to a downstream site (level B), calibrated specifically for the site (level C) and calibrated for the site with adjusted precipitation and temperature inputs (level D). All models generally captured the magnitude and variability of observed streamflows at the five study sites, and increasing level of model calibration generally improved performance. All models had at least 1 of 14 ERFSs falling outside a +/−30% range of hydrologic uncertainty at every site, and ERFSs related to low flows were frequently over-predicted. Our results do not indicate that any specific hydrologic model is superior to the others evaluated at all sites and for all measures of model performance. Instead, we provide evidence that (1) model performance is as likely to be related to calibration strategy as it is to model structure and (2) simple, regional-scale models have comparable performance to the more complex, fine-scale models at a monthly time step.

KEY WORDS hydrologic models; environmental flow; calibration; uncertainty; ecosystem health; ELOHA; water supply

Received 7 July 2014; Revised 6 January 2015; Accepted 6 January 2015

INTRODUCTION

River flows are essential for sustaining the health of aquatic ecosystems and maintaining ecosystem services such as water supply for consumptive use. Human activities including regulation by dams (Graf, 1999; Poff *et al.*, 2007; Biemans *et al.*, 2011), withdrawals (Gerten *et al.*, 2008), interbasin transfers (Jackson *et al.*, 2001) and land cover change (Foley *et al.*, 2005) have significantly altered the magnitude and timing of river flows. The health and biological condition of aquatic ecosystems has declined as a result (Dudgeon *et al.*, 2006; Carlisle *et al.*, 2010; Poff and Zimmerman, 2010; USEPA, 2011). In addition to anthropogenic hydrologic alterations, future changes in climate will likely further impact river flows (Georgakakos *et al.*, 2014). Assessing the effect of flow alteration on aquatic ecosystems has been identified as a critical area of research in the Southeastern USA (SARP, 2004; SALCC, 2012; Knight *et al.*, 2013), nationally (Carlisle *et al.*, 2010), and abroad (Annear *et al.*, 2004; Arthington *et al.*, 2006; Poff *et al.*, 2010). The Southeastern USA is recognized as one of the most ecologically rich areas in the world (Masters *et al.*, 1998), making it imperative to assess ecological response to flow alteration. As a result and considering recent droughts and interstate conflict over water availability issues, many states in the Southeastern USA are investigating the implementation of regulatory controls on

*Correspondence to: Peter V. Caldwell, Center for Forest Watershed Science, USDA Forest Service, Coweeta Hydrologic Lab, 3160 Coweeta Lab Road, Otto, NC 28763, USA.
E-mail: pcaldwell02@fs.fed.us

streamflow alteration in the interest of maintaining a balance between supporting healthy aquatic ecosystems while providing ample water supplies for human use (e.g. NCEFSAB, 2013). Reliable, multi-scale hydrologic and ecosystem modelling approaches are needed to accomplish this goal (Poff *et al.*, 2010). However, error in streamflow predictions by hydrologic models and predictions of ecological response to changes in flow regime with ecological models can be significant and may be compounding, thereby confounding the determination of environmental flow requirements and exposing managers of water resources to litigation by the regulated community.

The Ecological Limits of Hydrologic Alteration (ELOHA) framework (Poff *et al.*, 2010) is often used as a basis for developing regional environmental flow requirements. One of the first steps in the ELOHA process is to develop a hydrological foundation of simulated baseline and altered streamflow hydrographs. Hydrologic models are commonly used for this purpose because they have the ability to simulate monthly, daily or sub-daily streamflow under baseline conditions and an infinite number of scenarios of flow alteration. Available hydrologic models vary in their levels of complexity, temporal and spatial resolution, and required level of calibration. Detailed and highly parameterized fine-resolution models such as distributed physically based watershed and rainfall–runoff models are well suited for smaller domains but can be computationally expensive and difficult to parameterize at larger scales. In contrast, simple, easily parameterized models such as lumped regional models are useful for assessing broad implications of streamflow alteration at a large scale and identifying potential water limited areas (i.e. 'hot spots') but may have difficulty resolving unique sub-watershed scale physical processes and associated anthropogenic effects. Regardless of the hydrologic model used, the model must reasonably replicate observed streamflow and ecologically relevant flow statistics (ERFSs) that describe the magnitude, frequency, duration, timing and rate of change of streamflow under historical and current conditions across points of interest for regulation, and predict changes in these statistics as a result of changes in climate, land use, flow regulation and water withdrawals. Error in predictions of streamflow and ecosystem response to changes in flow may compounding; therefore, error in model prediction of streamflow is often carried through the analysis and should be quantified and reported to avoid misinterpretation of modelling outcomes.

Given the myriad of hydrologic models for water supply and environmental flow studies, resource managers need to understand the relative error in streamflow predictions among commonly used hydrologic models, not only in terms of classical fit statistics of streamflow observations [e.g. bias, Nash–Sutcliffe Efficiency (NSE)] but also in terms of ERFSs. In addition to the type of hydrologic model, model inputs (e.g. climate, soils and land cover) and calibration strategy will influence the capacity of a given hydrologic model application to accurately predict observed streamflow. Unfortunately, our understanding of the differences in streamflow predictions among model applications is hampered by a lack of comprehensive model comparison studies (Knight *et al.*, 2012). The aim of this study was to address these gaps in our understanding of differences among a set of hydrologic models, inputs and calibration strategies when applied to five US Geological Survey (USGS) continuous record gauging stations in the Southeastern USA. It was not the intent of this study to identify any specific model that is better suited for streamflow prediction over another because such differences are as likely to be related to model calibration strategy, experience and personal preference as they are to differences in model structure. Further, we did not intend to separate the influence of model structure from calibration strategy because such a comparison would not be representative of how models are developed and used to make management decisions. Rather, the overarching goal of this investigation was to quantify and compare the magnitude and potential causes of error associated with predicted streamflows from seven hydrologic models of varying complexity and calibration strategy as developed to answer management questions by computing classical hydrologic model fit statistics, and bias in the prediction of ERFSs. In addition, we tested the hypotheses that (1) simple, regional-scale hydrologic models would provide poorer predictions with greater levels of uncertainty than more complex physically based models and (2) models with higher levels of calibration would perform better than those with less calibration.

METHODS

Study area

The study area is the Apalachicola–Chattahoochee–Flint (ACF) basin, which drains approximately 52 000 km^2 of Georgia, Alabama and Florida in the Southeastern USA (Figure 1). This basin has been subject to water shortages, development pressure and interstate conflicts over water availability in the past and thus has been intensely studied over the last decade with multiple modelling efforts taking place to evaluate drought and global change effects on aquatic ecosystems and water supply at multiple scales (e.g. Georgakakos *et al.*, 2010; Markstrom *et al.*, 2012; Freeman *et al.*, 2013; LaFontaine *et al.*, 2013). Five sub-basins of varying size and characteristics within the ACF basin were selected as a basis of comparison among models (Table I). The five study sites ranged in basin area from 637 to 4792 km^2 with mean annual precipitation from 1950 to 2009 ranging

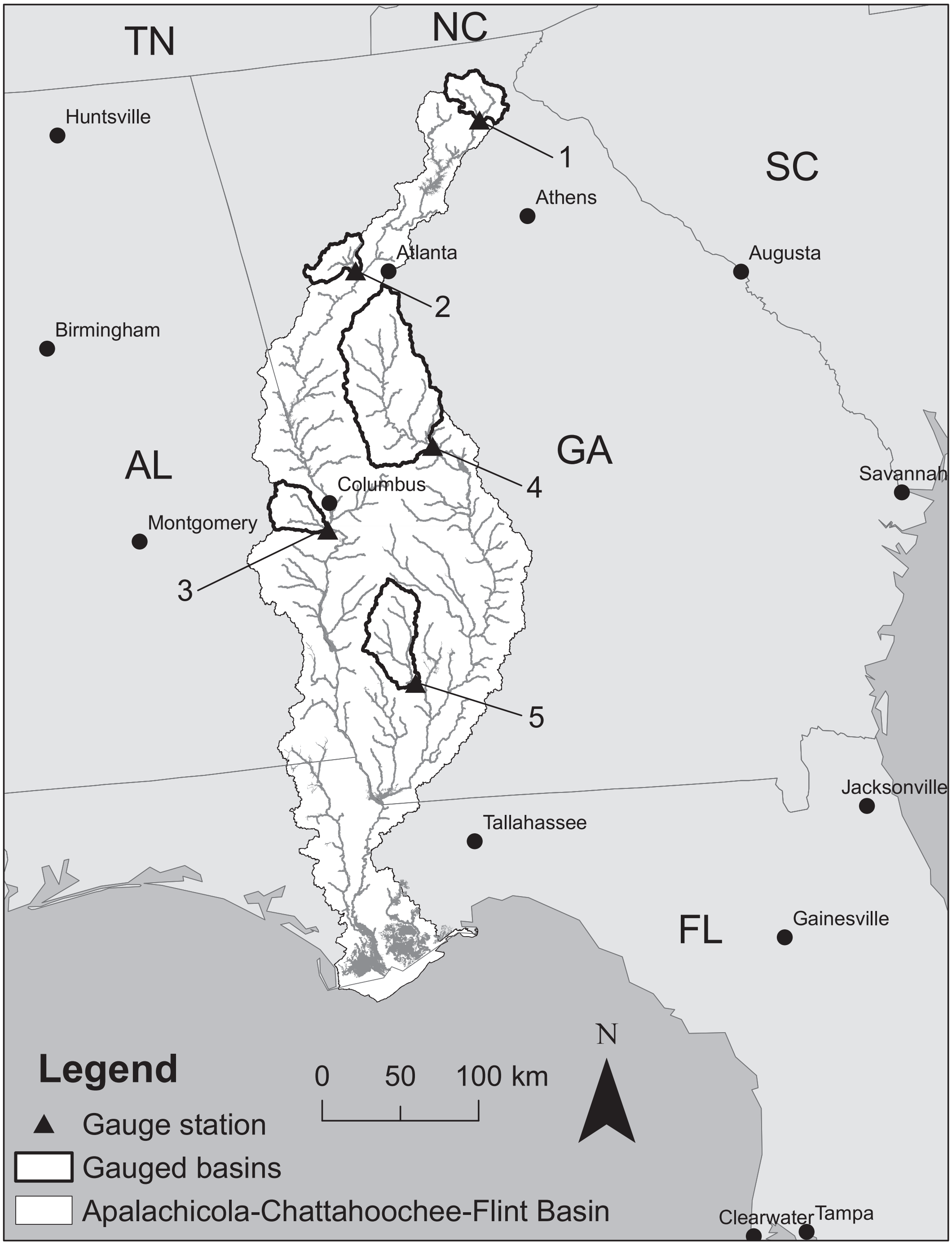

Figure 1. Locations of the five USGS gauges in the study area with continuous flow observations and predictions from all seven models,1980–1999.

from 1282 to 1728 mm and mean temperature ranging from 13.9 to 18.7 °C (Table I). The 2006 land cover (Fry *et al.*, 2011) ranged from mostly forested (site 1 – Chattahoochee River near Cornelia, GA), to developed (site 2 – Sweetwater Creek near Austell, GA), to mixed agriculture and forest along the southern extent of the ACF basin (site 5 – Ichawaynochaway Creek at Milford, GA). The amount of impervious area ranged from 0.4 to 9.4% and was significantly related to the amount of developed land in the basin ($R^2 = 0.986$, $p < 0.0001$). The five sites were selected for this study because output at these locations was available from all modelling efforts and thus allowed direct comparison among the various model applications. We compared characteristics of these sites to the 1364 other USGS streamflow gauging stations in the USGS Gages II database (Falcone *et al.*, 2010; Falcone, 2011) that were located in the

Table I. Descriptions of the five US Geological Survey continuous record gauges simulated by all models over the time period from 1980 to 1999. Screening comments for alteration, land cover percentages (based on the 2006 NLCD), and mean climate data (1950–2009 average) are based on (Falcone *et al.*, 2010; Falcone, 2011).

Site number	1	2	3	4	5
Station ID	02331600	02337000	02342500	02347500	02353500
Station name	Chattahoochee River, near Cornelia, GA	Sweetwater Creek, near Austell, GA	Uchee Creek, near Fort Mitchell, AL	Flint River, at US 19, near Carsonville, GA	Ichawaynochaway Creek, at Milford, GA
Drainage area (km^2)	816	637	834	4792	1606
Screening comments for alteration	Forested headwaters, mixed light ag/forest/suburban, many flood retention ponds on tributaries	Urban basin	Urban in headwaters of main tributary	Upstream urban	Proximate irrigated agriculture, many small ponds in headwaters, small towns on some tributaries
Developed (%)	8.8	39	8.1	13.9	3.3
Impervious (%)	1.1	9.4	1.3	3.6	0.4
Forested (%)	72	41	52	55	39
Agriculture (%)	13	11	16	16	37
Mean temperature (°C)	13.9	15.3	17.4	16.8	18.7
Mean precipitation ($mm\,yr^{-1}$)	1728	1384	1320	1282	1330
Mean runoff ($mm\,yr^{-1}$)	615	497	448	411	426
Runoff coefficient (%)	36	36	34	32	32

states of Virginia, North Carolina, South Carolina, Georgia, Alabama, Tennessee and Kentucky to insure that the five sites selected for comparison were generally representative of conditions across the region of study. We found that these five sites well represented the distribution of elevation, precipitation, air temperature, runoff, runoff/precipitation, impervious cover, developed land, forested land and agricultural land among other sites in the region. On average, percentiles of basin characteristics for five sites ranged from the 31st percentile to the 82nd percentile of the 1364 sites.

Model descriptions

All of the models evaluated were developed by different agencies, for different purposes, with different data inputs and spatial scales and were calibrated to different degrees using different objective functions. Specific models included in this study were the Hydrological Simulation Program-Fortran (HSPF), the USGS Monthly Water Balance Model (MWBM), two parameterizations of the USGS Precipitation–Runoff Modelling System (PRMS), the Soil and Water Assessment Tool (SWAT), the Generalized Watershed Loading Function (GWLF)-based Watershed Flow and Allocation model (WaterFALL®) developed by Research Triangle Institute (RTI), and the USDA Forest Service Water Supply Stress Index (WaSSI) model. The MWBM and WaSSI models are regional large-scale monthly time step models; WaterFALL® is of intermediate complexity at a fine scale and shorter time step; and HSPF, PRMS and SWAT are more complex, fine-scale shorter time step models. A summary of each model application and inputs is shown in Table II, and details regarding model calibration strategy are shown in Table III. More detail regarding each model application may be found in the cited references.

Hydrological Simulation Program-Fortran. Both the HSPF and SWAT models used in this study (Table II) were implemented by TetraTech as part of a larger study to characterize the sensitivity of streamflow, nutrient and sediment loading to a range of potential mid-21st century climate futures in 20 large, US drainage basins (Johnson *et al*., 2012; USEPA, 2013). All sub-basin delineations, input datasets and calibration objective functions used for HSPF were also used for SWAT (Tables II and III). Calibration of these models focused at the large basin scale of eight-digit hydrologic unit codes.

Hydrological Simulation Program-Fortran (Bicknell *et al*., 2005) is a water quantity and quality model commonly used for determination of total maximum daily loads to receiving waters in response to Clean Water Act requirements. HSPF is a well-documented watershed model that computes the water balance based on principles of the Stanford Watershed Model (Crawford and Linsley, 1966) in multiple surface and subsurface layers, typically at an hourly time step. The water balance is simulated on the basis of Philip's infiltration (Bicknell *et al*., 2005) coupled with multiple surface and subsurface stores [i.e. interception storage, surface storage, upper zone soil storage, lower zone soil storage, active groundwater and inactive (deep) groundwater]. Individual land units within a sub-basin in this simulation are represented using a hydrologic response unit approach that combines an overlay of land cover, soil and slope characteristics. Four model parameters were the primary focus during manual model calibration to improve model fit for hydrology with an emphasis on large basin response at a daily time scale. The calibrated model produced an hourly time series of predicted streamflow for the period from 1973 to 2003.

Soil and Water Assessment Tool. Soil and Water Assessment Tool was developed by the US Department of Agriculture to simulate the effect of land management practices on water, sediment and agricultural chemical yields in large, complex watersheds with varying soils, land use and management conditions over long periods (Neitsch *et al*., 2005). SWAT (as implemented here) uses the curve number approach (USDA Soil Conservation Service, 1972) to estimate surface runoff and then completes the water balance through simulation of subsurface flows, evapotranspiration (ET), soil storages and deep seepage losses at the daily time step. The curve number is estimated as a function of land use, cover, condition, hydrologic soil group and antecedent soil moisture. For SWAT, 11 model parameters were adjusted during manual model calibration to improve model fit. The calibrated model produced a daily time series of predicted streamflow for the period from 1973 to 2003.

Precipitation–Runoff Modelling System – Southeast Regional Assessment Project. The PRMS was applied to the ACF basin as part of a USGS Southeast Regional Assessment Project (SERAP) to provide integrated science that helps resource managers understand how ecosystems may respond to climate change (LaFontaine *et al*., 2013), hereafter PRMS-SERAP. The PRMS (Leavesley *et al*., 1983; Markstrom *et al*., 2008) is a deterministic, distributed-parameter, physical-process-based hydrologic modelling system. The model simulates daily land-surface hydrologic processes including ET, runoff, infiltration and interflow in hydrologic response units by balancing energy and mass budgets of the plant canopy, snowpack and soil zone on the basis of distributed climate information (e.g. temperature, precipitation and solar radiation) for 1951–1999. An automated parameter estimation procedure (Duan *et al*., 1994) was combined with a geographically nested approach to calibrate the PRMS-SERAP model using the Luca software (Hay and Umemoto, 2006; Hay *et al*., 2006).

Table II. General summary of model attributes.

	Model						
	HSPF	MWBM	PRMS-SERAP	PRMS-DAYMET	SWAT	WaterFALL®	WaSSI
Agency	TetraTech	USGS	USGS	USGS	TetraTech	RTI	USFS
Time step	Hourly	Monthly	Daily	Daily	Daily	Daily	Monthly
Spatial resolution	HUC10 (~410 km^2)	Aggregated NDHPlus catchments (~78 km^2)	HRU (~200 km^2)	Aggregated NDHPlus catchments (~78 km^2)	HUC10 (~410 km^2)	NHDPlus catchment (~1.0 km^2)	HUC12 (~80 km^2)
Withdrawals, flow regulation simulated	Yes	No	No	No	Yes	No	No
Land cover input	2001 NLCD (Homer *et al.*, 2007)	N/A	2001 NLCD (Homer *et al.*, 2007)	2001 NLCD (Homer *et al.*, 2007)	2001 NLCD (Homer *et al.*, 2007)	ca. 1970s USGS GIRAS (Price *et al.*, 2006)	2006 NLCD (Fry *et al.*, 2011)
Potential ET	Penman-Monteith (Monteith, 1965; Jensen *et al.*, 1990)	Hamon (Hamon, 1963)	Jensen-Haise (Jensen and Haise, 1963)	Jensen-Haise (Jensen and Haise, 1963)	Penman-Monteith (Monteith, 1965; Jensen *et al.*, 1990)	Hamon (Hamon, 1963)	Function of leaf area, precipitation, Hamon PET (Sun *et al.*, 2011a,2011b)
Climate input	Station observations (USEPA, 2008)	PRISM (PRISM Climate Group, 2013)	Maurer (Maurer *et al.*, 2002)	Daymet (Thornton *et al.*, 2013)	Station observations (USEPA, 2008)	USDA (Di Luzio *et al.*, 2008)	PRISM (PRISM Climate Group, 2013)

HSPF, Hydrological Simulation Program-Fortran; MWBM, Monthly Water Balance Model; PRMS, Precipitation–Runoff Modelling System; SWAT, Soil and Water Assessment Tool; WaSSI, Water Supply Stress Index; HRU, hydrologic response unit; HUC, hydrologic unit code; ET, evapotranspiration.

Table III. Level of model calibration by site. Calibration levels include the following: (A) uncalibrated, (B) calibrated to downstream gauge, (C) calibrated specifically for site and (D) calibrated specifically for site with adjusted precipitation, solar radiation and PET inputs.

Model	Number of parameters adjusted	Calibration objective function	Site	Level of calibration
HSPF	4	Total, seasonal, high and low streamflow within recommended ranges (Lumb *et al*., 1994; Donigian, 2000); maximize daily NSE	1	C
			2	B
			3	B
			4	B
			5	C
PRMS-SERAP	24	Minimize normalized RMSE of annual, monthly and mean monthly flow volumes and minimize normalized RMSE of daily flow timing using a 3-day moving average.	1	D
			2	D
			3	D
			4	D*
			5	D
PRMS-DAYMET	24	Minimize normalized RMSE of annual, monthly and mean monthly and daily flow volumes	1	D
			2	D
			3	D
			4	D
			5	D
SWAT	11	Total, seasonal, high and low streamflow within recommended ranges (Lumb *et al*., 1994; Donigian, 2000); maximize daily NSE	1	C
			2	B
			3	B
			4	B
			5	C
WaterFALL®	3	Minimize bias in log-transformed daily flows	1	C
			2	C
			3	B
			4	C
			5	B
MWBM	4	Minimize normalized RMSE of annual, monthly and mean monthly flow volumes	1	D
			2	D
			3	D
			4	D
			5	D
WaSSI	0	N/A	1	A
			2	A
			3	A
			4	A
			5	A

*PRMS-SERAP was calibrated to a downstream gauge at site 4, but this calibration adjusted precipitation, solar radiation and potential evapotranspiration inputs. For the purpose of comparison among calibration levels, PRMS-SERAP for this site was set at level D.
HSPF, Hydrological Simulation Program-Fortran; MWBM, Monthly Water Balance Model; PRMS, Precipitation–Runoff Modelling System; SWAT, Soil and Water Assessment Tool; WaSSI, Water Supply Stress Index; NSE, Nash–Sutcliffe Efficiency; RMSE, root mean squared error.

Precipitation–Runoff Modelling System – DAYMET. Like PRMS-SERAP, the DAYMET application of the PRMS model (hereafter PRMS-DAYMET) used the USGS's PRMS (Leavesley *et al*., 1983; Markstrom *et al*., 2008) to compute daily flows at each station, but at a different spatial resolution using different climate inputs (Table II) and different calibration objective functions within Luca (Hay and Umemoto, 2006; Hay *et al*., 2006) than PRMS-SERAP (Table III). DAYMET is a collection of gridded estimates of daily weather parameters generated by interpolation and extrapolation from daily meteorological observations (Thornton *et al*., 2013). The PRMS-DAYMET model was developed to provide a daily time series of simulated streamflows for 1980–2008 at several sites across the southeastern USA as part of a USGS National Water Census focus area study.

WaterFALL®. The application of RTI's WaterFALL® used for this study was developed to create a hydrologic foundation for detailed assessment of human and climate effects on stream and river flows, including the effects of hydrologic alterations on aquatic habitats in the Southeastern USA at the NHDPlus catchment scale (~1.0 km^2) (Kendy *et al*., 2011). The WaterFALL® simulation used in this study was developed to provide a baseline condition assessment by using land cover data representative of the

1970s and was calibrated to best match flow observations during this period. WaterFALL® employs an updated version of the well-established Generalized Water Loading Function (GWLF) hydrologic model (Haith and Shoemaker, 1987; Haith *et al.*, 1992). Updating GWLF for use in WaterFALL® included enabling the model to (1) run on EPA's enhanced National Hydrography Dataset (NHDPlus) hydrologic network, (2) use parameters from national datasets and (3) include the impacts of human alterations on streamflows. Like SWAT, surface runoff in WaterFALL® is computed on a daily basis using the curve number method across each land cover type in a catchment. Discharge from shallow groundwater is computed using a lumped parameter catchment-level water balance for unsaturated and shallow saturated zones controlled by the available water capacity of the unsaturated zone, a recession coefficient providing the rate of release from the saturated zone to the stream channel and a first-order approximation of infiltration losses to deep aquifer storage simulated using a seepage coefficient. Three model parameters were adjusted in an automated calibration process. The calibrated model produced a daily time series of predicted streamflow for the period from 1980 to 2006 for this study.

Monthly Water Balance Model. Like PRMS-DAYMET, the MWBM used in this study was developed to provide a monthly time series of simulated streamflows from 1980 to 2010 at several sites across the southeastern USA as part of a USGS National Water Census (Table II). The USGS MWBM (Hay and McCabe, 2002; McCabe and Markstrom, 2007; McCabe and Wolock, 2011) is based on the monthly Thornthwaite water balance model (Thornthwaite, 1948). When precipitation exceeds potential evapotranspiration (PET) in a given month, actual ET is equal to PET. Water in excess of PET replenishes soil-moisture storage. When soil-moisture storage reaches field capacity during a given month, the excess water becomes surplus. In a given month, some percent of the total surplus becomes runoff, and the remaining surplus is carried over to the following month. Four model parameters were adjusted during model calibration using the Shuffled Complex Evolution global search algorithm (Duan *et al.*, 1994), which is the same method that the Luca software (Hay and Umemoto, 2006; Hay *et al.*, 2006) uses for calibrating PRMS models.

Water Supply Stress Index. The monthly WaSSI model (Sun *et al.*, 2011b; Caldwell *et al.*, 2012) was developed by the US Forest Service to assess the effects of climate, land use and population change on water supply stress, river flows and aquatic ecosystems in the Eastern USA (Lockaby *et al.*, 2011; Marion *et al.*, 2013; Sun *et al.*, 2013; Tavernia *et al.*, 2013) and to examine the nexus of water and energy at the national scale (Averyt *et al.*, 2011; Averyt *et al.*, 2013). Infiltration, surface runoff, soil moisture and baseflow processes were computed using algorithms of the Sacramento Soil Moisture Accounting Model (SAC-SMA) (Burnash *et al.*, 1973; Burnash, 1995). In SAC-SMA, the soil profile is divided into a relatively thin upper layer and a much thicker lower layer, which supplies moisture to meet ET demands (Koren *et al.*, 2003). Each layer consists of tension water storage (i.e. between soil water tensions of field capacity and the plant wilting point) and free water storage (i.e. soil water tension greater than field capacity) that interact to generate surface runoff, lateral water movement from the upper soil layer to the stream (interflow), percolation from the upper soil layer to the lower soil layer, and lateral water movement from the lower soil layer to the stream (baseflow). Monthly ET in WaSSI was computed with leaf area index, precipitation and PET using an empirical equation derived from multisite eddy covariance ET measurements (Sun *et al.*, 2011a,b).WaSSI was developed to include the key ecohydrological processes that affect the water balance with off-the-shelf input datasets while having an acceptable level of predictive performance across the conterminous USA without calibration.

Evaluation and comparison of error in modelled streamflows

We compared classical hydrological model fit statistics as well as the prediction of ERFSs among model applications, sites and levels of calibration. Classical hydrological model fit statistics were computed for each model and site at the monthly time step for all seven models and at the daily time step for hourly and daily models only. Fit statistics evaluated included bias in mean streamflow, the NSE statistic (Nash and Sutcliffe, 1970), the root mean squared error and the coefficient of determination (R^2). Bias in mean streamflow within +/−25, 15 and 10% were considered indicative of satisfactory, good and very good hydrological model performance, respectively (Moriasi *et al.*, 2007). Similarly, NSE values that are greater than 0.50, 0.65 and 0.75 for prediction of monthly streamflow were considered to be indicative of satisfactory, good and very good model performance, respectively (Moriasi *et al.*, 2007). The NSE can range from negative infinity to 1.0; the closer NSE is to 1.0, the better the model fit to observations. Negative values of NSE indicate that using the mean of the observations provides a better fit than the model.

We also evaluated differences between predicted and observed ERFSs across the five hourly and daily models. The monthly MWBM and WaSSI models were not evaluated for prediction of ERFSs because many of ERFSs require a daily time step to be calculated. Predicted and observed daily mean flows were imported into the EflowStats package, an 'R' version of the National Hydrologic Assessment Tool (Henriksen *et al.*, 2006) developed for the USGS National Water Census by the USGS Center for Integrated Data Analytics and is available on GitHub at https://github.com/

USGS-R/EflowStats. This R package was developed to assist water resource professionals with characterizing the five major components of the flow regime (i.e. magnitude, frequency, duration, timing and rate of change) considered by many to be important in shaping ecological processes in streams (Olden and Poff, 2003; Henriksen *et al*., 2006; Kennen *et al*., 2007). A total of 175 ERFSs were evaluated for this study. Scatterplots were used to examine data distributions and to detect potential outliers in the ERFSs; statistics with extreme outliers or with highly limited data ranges were removed from further consideration. A Spearman rank correlation matrix (SAS Institute Inc., 1989) on the reduced set of ERFSs was then examined to eliminate any remaining redundant variables with a Spearman's rho >0.75. In cases where two statistics accounting for similar aspects of the flow regime were highly collinear, selection was based on best professional judgement. This approach was highly parsimonious and permitted the retention of important streamflow statistics, which were highly interpretable and management oriented and helped avoid the possibility of establishing significant ($p < 0.05$) correlations among a large suite of hydrologic variables simply by chance and introducing interdependencies among multiple explanatory variables (Van Sickle, 2003; King *et al*., 2005). Bias in the resulting subset of ERFSs was quantified by computing the percent difference between the predicted and observed flow statistic for each model and site. A hydrologic uncertainty of +/−30% (hereafter range of uncertainty) was used to aid in placing model prediction bias into context with inherent variability in streamflow and flow measurement (Murphy *et al*., 2013).

All model streamflow predictions were compared with USGS flow observations from October 1980 to September 1999 at each of the five study sites. After comparing classical fit statistics and prediction of ERFSs for each model, we explored differences in model structure, inputs and calibration strategy that may explain differences in predictive performance. In addition, we examined differences in model input for precipitation and model output for ET, runoff and soil moisture at site 4 (02347500, Flint River near Carsonville, GA, Table I; Figure 1), and the role these differences may play in the predicted water balance and model error. Precipitation inputs, as well as ET, and runoff outputs were directly comparable among all five models; however, soil moisture is represented differently in the hydrological models. To accommodate direct comparison, we standardized soil moisture predictions by dividing the monthly mean soil moisture by the maximum soil moisture storage for each model.

Model complexity, inputs and calibration strategy likely influenced the ability of a given model application to predict observed streamflow at the study sites. To examine the role that the level of model calibration plays in predictive performance, we defined and evaluated four levels of increasing calibration (Table III): Calibration level A models were uncalibrated (i.e. WaSSI), Calibration level B models were calibrated to a downstream gauge (i.e. some HSPF, SWAT, and WaterFALL® sites) but not the gauge of interest, Calibration level C models were calibrated specifically for that site (i.e. some HSPF, SWAT, and WaterFALL® sites), and Calibration level D models were calibrated specifically for that site, and precipitation, solar radiation and PET inputs were adjusted as part of the calibration process to account for uncertainty in gridded climate estimates (i.e. PRMS models and MWBM).

RESULTS

Classical model fit statistics

In general, all models and calibration levels satisfactorily captured the magnitude and variability of observed streamflows at the five study sites (Table IV, Figures 2 and 3). Only four of 35 combinations of model, site and calibration level yielded absolute bias of 25% or more (i.e. less than satisfactory): SWAT site 2, calibration level B (26.2%); WaterFALL® site 2 calibration level C (33.7%); WaSSI, site 2, calibration level A (25.0%); and WaterFALL® site 3, calibration level C (48.0%). Similarly, only two combinations of model, site and calibration level yielded monthly NSE less than satisfactory (i.e. <0.50): WaterFALL® site 3, calibration level B (0.28), and SWAT site 5, calibration level C (0.46). The median absolute bias across sites by model ranged from 2.5% (PRMS-DAYMET, calibration level D) to 15.4% (WaSSI, calibration level A). The median monthly NSE across all five study sites by model ranged from 0.64 (SWAT, site 2, calibration level B) to 0.87 (PRMS-SERAP, site 4, calibration level D). As expected, fit statistics for the daily models at the monthly time step were superior to the fit statistics at the daily time step for daily and sub-daily models. The median daily NSE across sites for the daily models ranged from 0.37 (HSPF, site 5, calibration level C) to 0.80 (PRMS-SERAP, site 5, calibration level D). Using model performance criteria for bias and monthly NSE (Moriasi *et al*., 2007), all models had good or better performance at most sites (Figure 3). The median bias in mean streamflow across all models by site ranged from −12% at site 3 to +9.2% at site 4, whereas the median monthly NSE across models ranged from 0.72 at site 2 to 0.89 at site 1.

Increasing calibration intensity tended to improve model fit across sites and models (Figure 4). Calibration level A (uncalibrated) included only WaSSI simulations, calibration level B (calibrated to downstream gauge) include three sites for HSPF and SWAT and two sites for WaterFALL®, calibration level C (calibrated specifically for site) included two sites for HSPF and SWAT and three sites for WaterFALL®, and calibration level D (calibrated for site

Table IV. Summary of classical model fit statistics for the seven models compared in this study.

Site	Model	Calibration level	Bias in mean (%)	Monthly NSE	Monthly RMSE	Monthly R^2	Daily NSE	Daily RMSE	Daily R^2
				–	cfs	–	–	cfs	–
1	HSPF	C	−6.4	0.85	145.5	0.87	0.66	353.2	0.67
1	PRMS-SERAP	D	−1.8	0.92	115.8	0.92	0.80	265.5	0.81
1	PRMS-DAYMET	D	−2.5	0.92	103.9	0.93	0.87	208.2	0.88
1	SWAT	C	−3.6	0.56	255.5	0.79	0.36	544.5	0.56
1	WaterFALL®	C	4.6	0.90	119.0	0.91	0.59	429.1	0.67
1	MWBM	D	−1.4	0.89	135.3	0.89	–	–	–
1	WaSSI	A	−2.8	0.68	227.4	0.79	–	–	–
2	HSPF	B	2.6	0.71	147.2	0.73	−0.23	543.6	0.20
2	PRMS-SERAP	D	1.3	0.92	81.2	0.92	0.84	207.8	0.84
2	PRMS-DAYMET	D	0.1	0.86	102.5	0.87	0.60	292.6	0.60
2	SWAT	B	−26.2	0.64	137.7	0.75	0.41	253.3	0.44
2	WaterFALL®	C	−33.7	0.62	69.5	0.87	0.31	302.5	0.37
2	MWBM	D	−2.1	0.84	114.2	0.84	–	–	–
2	WaSSI	A	25.0	0.72	123.3	0.84	–	–	–
3	HSPF	B	−24.6	0.62	211.8	0.67	−0.18	855.1	0.26
3	PRMS-SERAP	D	−12.0	0.76	135.0	0.82	0.75	296.9	0.77
3	PRMS-DAYMET	D	−2.9	0.82	158.6	0.83	0.64	400.4	0.64
3	SWAT	B	−18.2	0.65	173.8	0.70	0.50	300.3	0.55
3	WaterFALL®	B	−48.0	0.28	114.2	0.61	0.38	378.1	0.43
3	MWBM	D	−6.6	0.85	157.7	0.86	–	–	–
3	WaSSI	A	−1.7	0.75	179.8	0.75	–	–	–
4	HSPF	B	9.2	0.92	506.5	0.93	0.74	1408.9	0.75
4	PRMS-SERAP	D	16.8	0.87	615.1	0.91	0.74	1569.1	0.75
4	PRMS-DAYMET	D	−0.3	0.89	577.6	0.89	0.67	1528.3	0.68
4	SWAT	B	5.0	0.86	725.6	0.88	0.63	1499.7	0.64
4	WaterFALL®	C	11.2	0.86	691.4	0.90	0.79	1383.5	0.80
4	MWBM	D	−8.4	0.83	755.5	0.85	–	–	–
4	WaSSI	A	16.2	0.77	858.6	0.82	–	–	–
5	HSPF	C	−2.1	0.75	283.2	0.77	0.37	667.3	0.46
5	PRMS-SERAP	D	3.9	0.83	235.1	0.85	0.80	370.6	0.80
5	PRMS-DAYMET	D	−15.3	0.73	280.9	0.82	0.47	525.1	0.50
5	SWAT	C	5.8	0.46	401.9	0.77	0.24	794.5	0.52
5	WaterFALL®	B	−15.1	0.83	182.0	0.88	0.17	761.8	0.37
5	MWBM	D	−9.4	0.77	265.3	0.79	–	–	–
5	WaSSI	A	15.4	0.69	300.9	0.81	–	–	–

HSPF, Hydrological Simulation Program-Fortran; MWBM, Monthly Water Balance Model; PRMS, Precipitation–Runoff Modelling System; SWAT, Soil and Water Assessment Tool; WaSSI, Water Supply Stress Index.

with adjusted precipitation, solar radiation and PET inputs) included all PRMS-SERAP, PRMS-DAYMET and MWBM simulations (Table III). The median absolute bias in streamflow across models and sites was 15, 17, 6 and 3% for calibration levels A, B, C and D, respectively. Similarly, the monthly NSE tended to increase with increasing level of calibration; median monthly NSE across sites and models for calibration levels A, B, C and D, were 0.72, 0.68, 0.75 and 0.85, respectively. Daily NSE across sites and daily time step models also improved with increasing calibration (not shown), with median values of 0.39, 0.37 and 0.74 for calibration levels B, C and D, respectively.

Water balance components

Evaluation of the water balance components at site 4 revealed the effect of modelling assumptions and calibration strategies on model fit statistics (Figure 5). Models generally over-predicted streamflow at site 4 (Table IV, Figure 2), and thus, calibration strategies for some models were aimed at either adjusting observed precipitation (i.e. PRMS-SERAP, PRMS-DAYMET and MWBM) or increasing losses through deep seepage (i.e. WaterFALL®). Absolute streamflow over-prediction was most prevalent during the seasonally high flow months of January, February and March, but streamflow over-prediction expressed as a percentage were highest in the low-flow

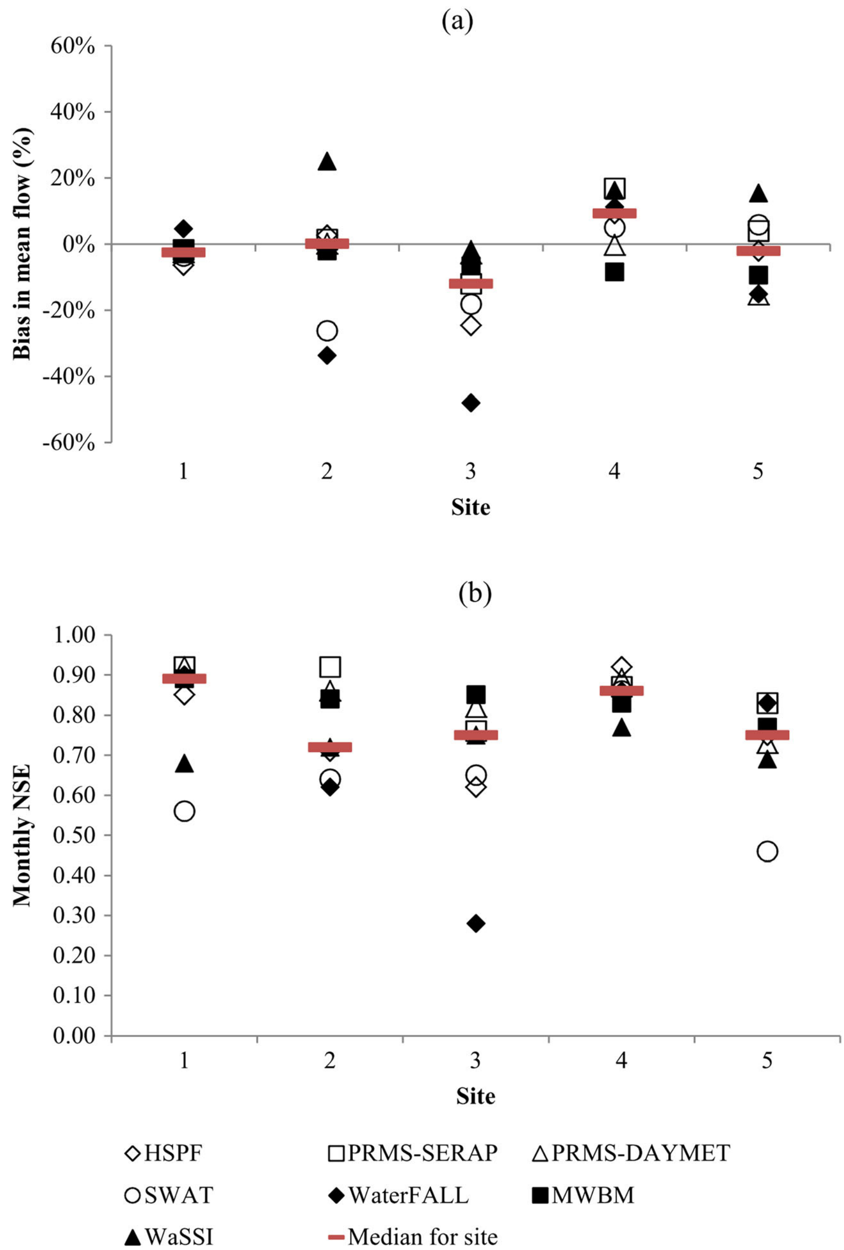

Figure 2. Distribution of selected classical model fit statistics across models by site for the flow time series from 1980 to 1999; (a) bias in mean flow and (b) monthly Nash–Sutcliffe Efficiency.

months of July, August and September (Figure 5b). Input for mean annual precipitation was similar for HSPF, SWAT, WaterFALL® and WaSSI, ranging from 1221 mm (WaterFALL®) to 1258 mm (HSPF and SWAT) for a total difference of 3% (Figure 5a). Mean annual precipitation for PRMS-SERAP, PRMS-DAYMET and MWBM was reduced for site 4 during model calibration, resulting in decreases in mean annual precipitation of approximately 9, 15 and 10%, respectively, whereas the other models did not adjust input precipitation. WaterFALL® included deep seepage losses (approximately 127 mm or 10% of precipitation) to reduce streamflow predictions to more closely match observations by adjusting the seepage coefficients in the calibration process. Had deep seepage not been included, bias in mean streamflow for WaterFALL® may have increased from 11.3 to 44.2% although other model parameters would have likely been adjusted to improve model fit. HSPF and SWAT simulations included consumptive use withdrawals based on available data that also reduced streamflow predictions, amounting to approximate 33 and 67 mm, respectively. Had consumptive use not been considered, bias in mean streamflow would have increased from 9.3 to 18.0% for HSPF and from 5.1 to 22.5% for SWAT.

Differences in seasonal runoff bias among models can be partially explained by differences in ET estimates.

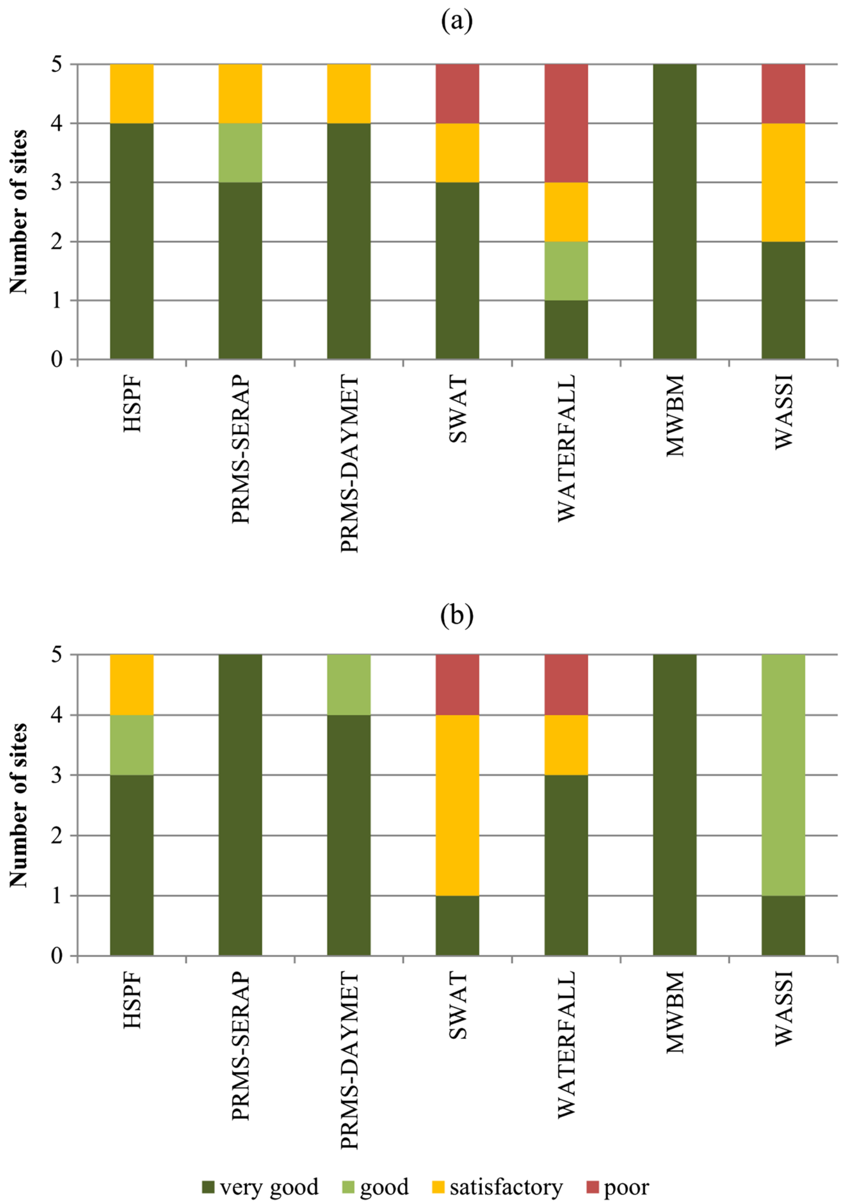

Figure 3. Distribution of sites falling into performance categories established by (Moriasi *et al*., 2007) for (a) absolute bias in mean flow and (b) monthly Nash–Sutcliffe Efficiency. Bias in mean flow within +/−25, 15 and 10% are considered to be indicative of satisfactory, good and very good model performance, respectively, and monthly NSE values that are greater than 0.50, 0.65 and 0.75 for prediction of monthly streamflow are indicative of satisfactory, good and very good model performance, respectively.

Assuming no consumptive use or deep seepage losses, ET can be estimated as the difference between long-term mean annual precipitation and runoff (Figure 5a). Using the mean precipitation across models as the mean precipitation for the basin, estimated ET, calculated as the difference between mean precipitation and runoff, was 797 mm. Bias in predicted ET was then 0.8, −14.4, −15.6, −1.4, −16.5, −5.9 and −1.7% for HSPF, PRMS-SERAP, PRMS-DAYMET, SWAT, WaterFALL®, MWBM and WaSSI, respectively. There was good agreement between HSPF, SWAT, WaSSI and the estimated observed mean annual ET, suggesting conformance to the major features of the water balance. WaterFALL® predicted less ET relative to other models and the estimated observed ET, but the deep seepage term partially compensated for this difference. Soil moisture levels were considerably lower for WaterFALL® than the other models (Figure 5b) indicating that available soil water storage and/or recession coefficients may explain the lower ET estimates. The PRMS and MWBM models also underpredicted ET relative to other models and the estimated observed ET largely because precipitation was reduced during model calibration.

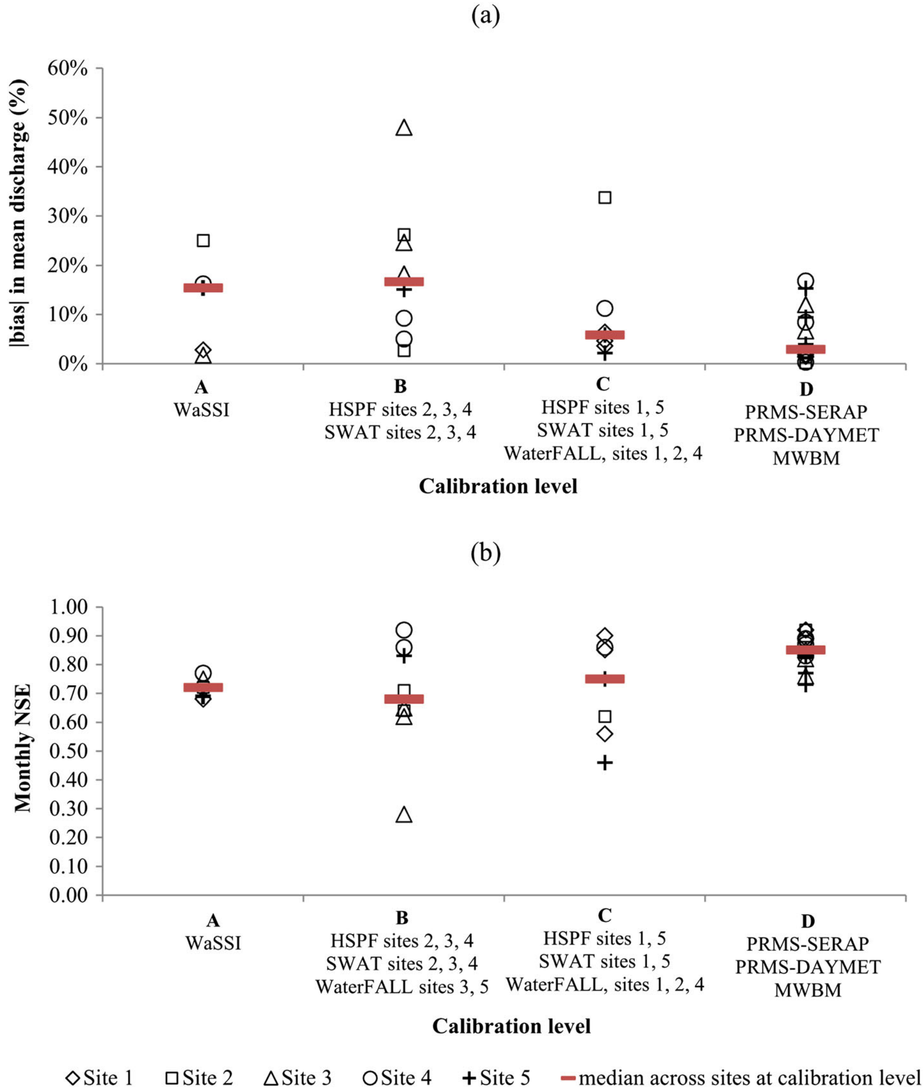

Figure 4. Distribution of selected classical model fit statistics across sites by level of calibration for the 1980–1999 flow time series; (a) absolute bias in mean flow and (b) monthly Nash–Sutcliffe Efficiency. Calibration levels include: (A) uncalibrated, (B) calibrated to downstream gauge, (C) calibrated specifically for site and (D) calibrated specifically for site with adjusted precipitation, solar radiation and PET inputs.

Prediction of ecologically relevant flow statistics

The full suite of 175 flow statistics computed with the EflowStats package in R was evaluated for redundancy and reduced to a subset of 14 ERFSs (Table V). We evaluated differences between predicted and observed ERFSs across the five hourly and daily models; the monthly MWBM and WaSSI models were not evaluated for prediction of ERFSs because many of the ERFSs require a daily time step to be calculated. Overall bias in the prediction of the ERFSs among sub-monthly time step models varied by site and by flow statistic (Table VI, Figure 6), with no model or calibration level clearly having superior predictive performance for all sites and statistics. The median absolute bias across all ERFSs and sites ranged from 18.7% (PRMS-DAYMET, calibration level D) to 31.9% (SWAT, calibration levels B and C) (Table VII). Increasing calibration tended to reduce bias for individual models overall. For example, the median absolute bias across all sites and ERFSs decreased from 22.6% for calibration level B to 16.9% for calibration level C for HSPF (Table VII). The median absolute bias across all sites and ERFSs for calibration level D (19.1%) was similar to that of all models at calibration level C (18.4%).

All hydrological models had at least one flow statistic falling outside the 30% range of hydrologic uncertainty at every site (Figures 6 and 7a). The number of ERFSs out of the total of 14 (Table V) that fell outside this range at three or more of the five sites ranged from 3 (HSPF, calibration levels B and C) to 9 (SWAT, calibration levels B and C) (Figure 7a). Some of the magnitude, frequency and duration ERFSs tended to be better represented across

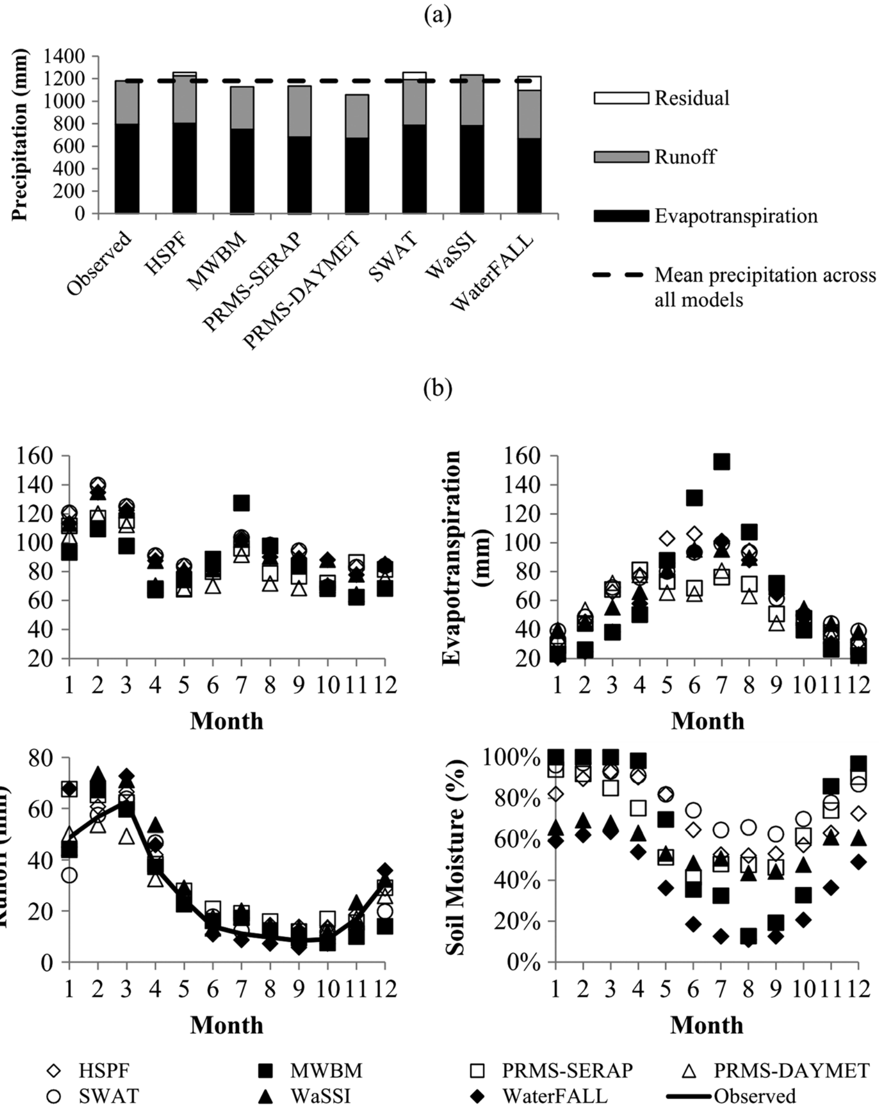

Figure 5. (a) Partitioning of mean annual precipitation into evapotranspiration and runoff and residual lost to either consumptive use or deep seepage and (b) monthly median precipitation, evapotranspiration, runoff, and soil moisture for all models at site 4: USGS gauge 02347500 Flint River at US19 near Carsonville, GA. Runoff was computed by dividing discharge by drainage area. Observed evapotranspiration was computed by taking the difference between the mean precipitation across models and the observed mean annual runoff at the gauge.

models than others. For example, MA41 (mean annual runoff), MA25 (variability of February flows), FL2 (variability in low-pulse count) and DH4 (mean of annual maximum 30-day average flow) were generally well-predicted, with bias outside the range of uncertainty at less than three of the five sites for all models. On the other hand, bias in FL1 (frequency of low flood) was outside the range of uncertainty for three or more of the five sites for all models, and bias in ML21 (variability of annual minimum flows) and DL4 (mean of annual minimum 30-day average flows) was outside the range of uncertainty for more than three or more of the five sites for four of the five models. The hydrologic models evaluated in this study generally had lower bias in the prediction of flow statistics representing mean flows (e.g. MA41) than statistics representing the extremes of flow (e.g. FL1), particularly low flow conditions. Bias was greater for many of the low flow statistics as a result of the low absolute magnitudes of these statistics (Table VI, Figure 6). This result is a fairly common outcome for many modelling studies as a result of choices made during the calibration process; however, this modelling bias may directly affect the predictive capacity of flow-ecology response models derived using ERFSs that fall outside the established range of uncertainty.

There was considerable variability in ERFS predictive performance across sites (Figures 6 and 7b). The ERFSs (Table V) with prediction bias outside the range of hydrologic uncertainty for three or more of the five models

Table V. Definitions of the reduced set of 14 ecologically relevant flow statistics used in this study to describe the magnitude, frequency, duration, timing and rate of change of flow.

Category	Statistic name	Description (unit of measurement)
Magnitude	MA41	Mean annual runoff: compute the annual mean daily streamflow and divide by the drainage area. (cubic feet per second (cfs) per square mile (cfsm))
	MA25	Variability of February flow values: compute the standard deviation for each month in each year. Divide the standard deviation by the mean for each month and take the mean of these values for each month across years. (%)
	ML6	Minimum June streamflow: Minimum June streamflow across the period of record. (cfs)
	ML9	Minimum September streamflow: minimum September streamflow across the period of record. (cfs)
	ML21	Variability of annual minimum flows: compute the standard deviation of annual minimum streamflow and divide by the mean annual minimum streamflow. (%)
	MH20	Specific mean annual maximum flow: mean annual maximum flow across the period of record divided by watershed area. (cfsm)
Frequency	FL1	Frequency of low flood: low flood pulse count. (n/year)
	FL2	Variability in low-pulse count: Coefficient of variation for the number of annual occurrences of daily flows less than the 25th percentile. (dimensionless)
Duration	DL17	Variability in low pulse duration: Standard deviation for the yearly average low-flow pulse durations (daily flow less than the 25th percentile). (%)
	DH20	High flow duration. (days)
	DL4	Mean of the annual minimum 30-day average flows. (cfs)
	DH4	Mean of the annual maximum 30-day moving average flow for the entire record. (cfs)
Timing	TH1	Average Julian date of the annual maximum flow for the entire record. (Julian day)
Rate of change	RA4	Variability of the fall rate for the entire record. (%)

included FL1 and TH1 for site 1 (2 of 14 ERFSs); ML21, FL1 and TH1 for site 2 (3 of 14); ML6, ML9, MH20, FL1, DL17 and DL4 for site 3 (6 of 14); ML6, ML9, ML21, FL2 and DL4 for site 4 (5 of 14); and ML6, ML21, FL1, DL17, DH20 and DL4 for site 5 (6 of 14). ERFSs at site 1 were generally well-predicted, likely reflecting the fact that (1) all models were calibrated for this site (i.e. calibration levels C and D) and (2) the watershed upstream of the site was mostly low intensity forested land (Table I) and likely had fewer flow alterations that may affect model performance. Although sites 1 and 3 had similar levels of urban development (Table I), ERFSs were not as well predicted at site 3. In particular, those ERFSs relating to low flows (e.g. ML6, ML9, FL1, DL17 and DL4) were generally over-predicted with bias exceeding 30% (Figure 7b). The HSPF, SWAT and WaterFALL® models were not specifically calibrated for site 3 (calibration level B), which may explain the higher bias for this site, but the PRMS-SERAP and PRMS-DAYMET models, calibrated specifically for this site, (calibration level D) also had higher bias for ERFSs at site 3 than site 1 indicating that there may be underlying natural and anthropogenic processes that are not being accounted for in the models. Most hydrological model predictions of ERFSs for site 2 were within the range of uncertainty despite some models being calibrated to a downstream gauge (level B for HSPF and SWAT) and having higher levels of urban development (Table I). The ERFSs for site 2 that were outside the range of uncertainty included ML21 (variability in annual minimum flows), FL1 (frequency of low flood), and TH1 (average Julian date of annual maximum flow). Of these ERFS, ML21 was generally under-predicted, FL1 was generally over-predicted, and TH1 was generally under-predicted. ERFSs relating to low flows were generally over-predicted by most models for site 4 (e.g. ML6, ML9, FL2 and DL4) and site 5 (e.g. ML6, ML21, FL1, DL17 and DL4).

DISCUSSION

Our hypothesis that simple, regional-scale models have poorer predictive performance than more complex fine-scaled models at the monthly timestep was not supported. We found that all models had ‘good’ or better performance (as defined in this paper) at most sites. Comparing classical model fit statistics across all sites, the simple MWBM and WaSSI applications had comparable error in predicting observed streamflows at the monthly time step as that of the more complex HSPF, PRMS, SWAT and WaterFALL® models (Figure 2). According to monthly NSE criteria (Moriasi *et al.*, 2007), the uncalibrated WaSSI model predictions would be considered ‘good’ at all of the five sites and ‘very good’ at one site, whereas the calibrated MWBM predictions would be ‘very good’ at all sites. These results support the notion that we can leverage the benefits of both simple, large-scale models at the monthly time step with more complex, high-resolution models to allow more robust climate change impact studies for maintaining a better balance between the availability of

Table VI. Observed and predicted ecologically relevant flow statistics for the five study sites.

		Magnitude						Frequency		Duration				Timing	Rate of change
		MA41	MA25	ML6	ML9	ML21	MH20	FL1	FL2	DL17	DH20	DL4	DH4	TH1	RA4
	Site/calib. level	cfsm	%	cfs	cfs	%	cfsm	n yr^{-1}	–	%	Days	cfs	cfs	Julian d	%
Observed	1	2.40	62.0	446.5	276.8	37.8	22.0	13.7	67.0	42.9	7.9	303.2	1530.5	21	342.5
HSPF	1/C	2.25	53.4	411.1	320.6	31.0	15.4	9.2	56.1	60.8	9.6	293.6	1523.7	31	249.4
PRMS-SERAP	1/D	2.36	50.3	444.1	289.9	50.7	14.9	9.2	64.0	43.2	8.0	271.8	1434.5	362	240.9
PRMS-DAYMET	1/D	2.34	52.1	465.9	309.9	42.7	18.4	9.8	61.6	45.1	5.8	297.8	1373.5	1	320.4
SWAT	1/C	2.31	57.3	127.7	119.6	92.5	16.0	11.2	55.9	36.1	4.5	113.1	1886.1	26	185.5
WaterFALL®	1/C	2.51	64.7	481.8	345.7	28.7	20.7	8.4	59.8	44.9	4.2	330.1	1524.2	13	294.9
Observed	2	1.49	90.1	92.2	51.2	62.4	17.6	9.5	40.8	45.8	5.4	59.9	1007.3	58	269.3
HSPF	2/B	1.53	82.6	110.5	74.6	31.6	23.9	13.9	38.9	42.3	3.7	84.1	993.2	50	363.8
PRMS-SERAP	2/D	1.51	91.5	99.6	42.2	41.8	16.6	13.2	38.5	27.8	4.9	67.6	994.1	35	217.6
PRMS-DAYMET	2/D	1.49	75.2	99.9	64.0	35.7	17.1	13.7	48.0	32.0	4.3	66.1	959.6	23	270.6
SWAT	2/B	1.10	49.1	86.4	56.0	92.7	7.0	5.9	41.9	38.3	14.5	32.8	797.4	48	217.0
WaterFALL®	2/C	0.99	83.8	53.8	41.7	34.5	14.8	13.4	37.4	35.1	5.8	42.3	663.9	18	430.3
Observed	3	1.26	92.1	42.6	26.5	41.3	28.0	7.1	46.9	41.5	8.4	25.5	1495.6	60	418.3
HSPF	3/B	0.95	105.7	52.2	39.0	36.8	31.9	8.0	49.7	43.0	6.2	37.8	1183.2	67	740.3
PRMS-SERAP	3/D	1.11	89.4	72.4	56.5	34.7	19.0	11.9	41.7	27.1	6.4	70.4	1112.5	59	338.0
PRMS-DAYMET	3/D	1.23	82.5	82.6	33.2	64.7	21.3	13.0	39.5	34.9	7.3	50.7	1259.5	38	386.9
SWAT	3/B	1.04	61.8	156.3	77.8	97.9	9.7	3.7	48.8	61.4	14.4	44.4	1002.9	53	273.1
WaterFALL®	3/B	0.66	95.5	46.3	37.3	44.8	17.7	15.1	37.3	25.5	5.6	44.9	713.3	38	519.1
Observed	4	1.12	73.0	475.3	302.5	42.9	14.8	6.9	30.5	44.4	9.4	342.8	6650.1	52	296.9
HSPF	4/B	1.22	56.5	723.8	531.7	30.6	12.0	9.8	41.9	33.2	8.0	532.5	6562.3	58	218.9
PRMS-SERAP	4/D	1.31	58.7	827.6	479.6	45.0	12.0	8.8	46.7	29.0	14.1	511.5	6941.4	55	265.6
PRMS-DAYMET	4/D	1.12	65.1	750.2	511.5	29.8	12.9	12.2	54.2	32.5	8.3	494.9	5948.9	44	298.4
SWAT	4/B	1.17	39.7	884.3	577.5	57.4	7.5	5.3	42.1	58.1	15.1	366.9	6511.7	63	177.9
WaterFALL®	4/C	1.24	52.8	486.2	272.2	60.5	12.0	6.7	42.0	45.0	10.6	220.7	7191.2	55	310.7
Observed	5	1.19	48.1	197.8	196.3	37.3	11.4	7.9	38.2	58.4	11.4	202.9	2014.0	47	338.0
HSPF	5/C	1.17	58.1	310.2	249.7	40.6	10.7	6.7	59.2	84.7	13.1	198.7	1973.4	38	278.1
PRMS-SERAP	5/D	1.24	42.7	287.2	234.9	24.1	8.5	7.7	47.8	34.1	10.9	232.5	1999.4	36	268.8
PRMS-DAYMET	5/D	1.01	35.9	76.2	67.3	84.1	6.0	11.8	30.4	28.9	20.5	71.5	2046.9	25	166.7
SWAT	5/C	1.26	62.8	200.8	180.8	66.2	11.0	4.9	48.6	66.9	16.5	109.1	2632.4	43	243.8
WaterFALL®	5/B	1.01	79.3	293.2	184.8	51.4	14.0	3.8	53.0	50.7	7.6	124.4	1719.4	29	454.4

HSPF, Hydrological Simulation Program-Fortran; PRMS, Precipitation–Runoff Modelling System; SWAT, Soil and Water Assessment Tool.

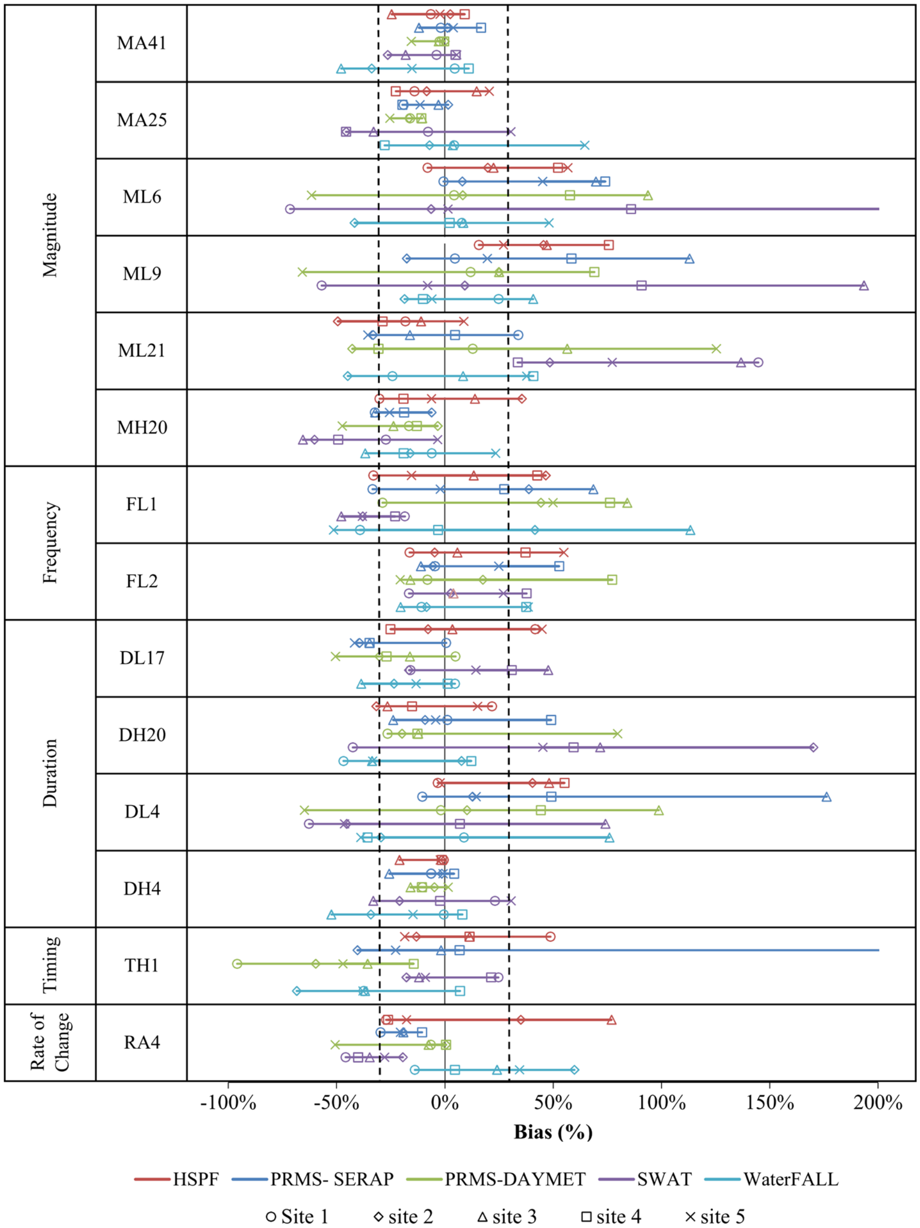

Figure 6. Bias in prediction of the 14 ecologically relevant flow statistics across the five study sites for the daily time step hydrologic models. Dashed lines show the range of hydrologic uncertainty, +/−30%.

water to support aquatic assemblages while conserving water for long-term human needs across broad regions than using either approach in isolation. For example, the WaSSI model could be used to quickly assess regional scale climate change effects and identify 'hot spots' where the combined effects of land cover change, climate change and/or streamflow alteration may threaten water resources. Fine-scale, physically based models of higher temporal resolution (e.g. HSPF, PRMS, SWAT and WaterFALL®) could then be applied to those areas of concern to provide higher resolution quantitative estimates of changes in water supply and ERFSs using more site-specific inputs. Achieving good model fit at the monthly time step with either monthly or smaller time step models generally indicates that the correct balance of precipitation and ET is represented, but it does not necessarily indicate that the separation between surface and subsurface flows is accurately represented. Thus, good model fit in monthly time step simulations may not indicate that the model would be useful in answering resource questions that require detailed information regarding surface runoff and baseflows.

We did find evidence supporting our hypothesis that increasing calibration intensity generally improved model fit across sites and models. For classical model fit statistics,

Table VII. Median absolute bias (%) across all sites and all ecologically relevant flow statistics (ERFSs) by model and calibration level for daily timestep models. The ERFSs were calculated only for the daily time step hydrologic models; therefore, WaSSI and MWBM were not included in this comparison.

Model	Calibration level B	Calibration level C	Calibration level D	Median for model across all calibration levels
HSPF	22.6	16.9		19.4
PRMS-SERAP			19.1	19.1
PRMS-DAYMET			18.7	18.7
SWAT	36.2	27.1		31.9
WaterFALL®	36.7	14.9		24.1
Median for calibration level across all models	32.9	18.4	19.1	22.6

HSPF, Hydrological Simulation Program-Fortran; PRMS, Precipitation–Runoff Modelling System; SWAT, Soil and Water Assessment Tool.

the more intensive site-specific calibrations (levels C and D) generally decreased bias and increased NSE at the monthly scale relative to uncalibrated models (level A) and models calibrated to a downstream site (level B); however, differences between calibration levels A and B were not as large (Figure 4). For ERFSs, increasing calibration tended to reduce bias for individual models overall, but no model or calibration level clearly had superior predictive performance for all sites and ERFSs. For example, bias in ERFS predictions was generally lower for sites calibrated at level C than for sites calibrated at level B for HSPF, SWAT and WaterFALL®. However, differences in ERFS bias between models calibrated for sites at level C and models calibrated for sites at level D were generally smaller. Clearly, adjusting precipitation, solar radiation and PET during model calibration (i.e. calibration level D) can result in improved model fit relative to other levels of calibration; however, caution should be used when applying models calibrated in this manner to make projections using other sources of climate input (e.g. future climate change scenarios). Often, there are no observed streamflow data available for locations where streamflow models are needed. In these cases, models of daily streamflow will likely be based on less intensive calibration (i.e. levels A or B) or parameterized on the basis of model calibration in nearby streams where streamflow observations are available. Model performance in these cases, regardless of the model framework, will likely be reflective of the models with less intensive calibration in this study. In addition to model calibration strategy, model inputs and assumptions also played a role in predictive performance. For example, the WaterFALL® model used land cover from the 1970s (with lower levels of impervious cover than in 2001 or 2006 used by other models) and tended to have negative bias for sites with some level of urbanization (e.g. site 2) because surface runoff was lower.

Similar to other ecological flow modelling studies (Wenger *et al*., 2010; Murphy *et al*., 2013), the sub-monthly time step models evaluated in this study tended to have ‘good’ predictive performance for ERFSs

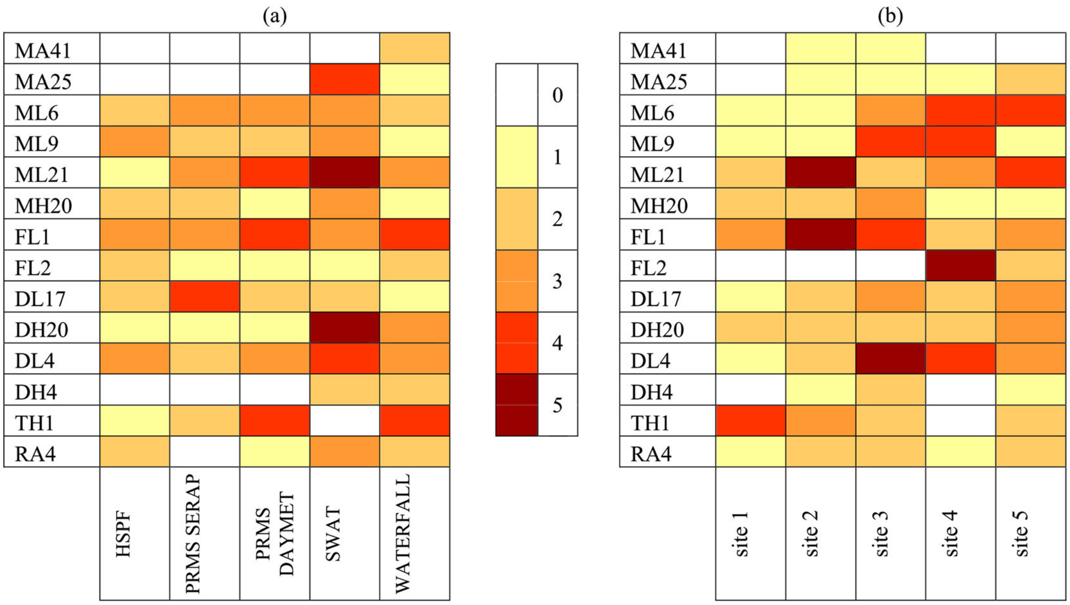

Figure 7. (a) Number of sites per model out of a total of five and (b) number of models per site out of a total of five for which bias in the 14 ecologically relevant flow statistics (ERFSs) fell outside of the +/−30% range of hydrologic uncertainty. Monthly models (MWBM and WaSSI) are not shown because computation of most ERFSs used in this study requires a daily timestep.

representing the mean of flow, but had difficulty in predicting flow statistics related to low or extreme flows (Table VI, Figures 6 and 7). The variability in prediction bias across ERFSs for the sub-monthly models is indicative not only of the variability in level of calibration across models and sites but also of the challenges associated with calibrating hydrologic models for all streamflow conditions. Model calibration is generally intended to capture the variability and mean magnitude of streamflow. It is nearly impossible to calibrate models to fit the entire range of observed streamflows because adjusting model parameters to fit a portion of the flow regime has an effect on how well the model fits observed streamflows outside of that range. For example, WaterFALL® was calibrated to log-transformed daily streamflows (Table III) to provide improved fit for low flows, but calibrating in this way can degrade fit for high flows. There was considerable variability in classical fit statistics and ERFS predictive performance across sites (Figures 2, 6 and 7). Fit statistics and ERFSs at some sites (e.g. site 1) were better predicted by all models than at other sites (e.g. site 3), illustrating the fact that model performance is site specific regardless of model framework or level of calibration. These findings may have implications for the development of flow-ecology response models because it is often the low flows (baseflows), annual-flow pulses and seasonality of high flows that provide the conditions necessary to support natural-assemblage complexity (Poff and Ward, 1989; Stanford *et al*., 1996; Poff *et al*., 1997; Richter *et al*., 1997; Matthews, 2005). Because streamflow models in this study tended to perform better when predicting mean ERFSs then when predicting low or extreme flow EFRMs, great care should be taken when using ERHMs with high prediction bias (e.g. ML6, ML9, FL1, DL17 and DL4) to develop flow-ecology response models.

All of the models evaluated were developed by different agencies, for different purposes, with different input data sets, and, in general, were calibrated to different degrees using different objective functions. From a purely scientific perspective, it may be ideal to have the ability to separate the influence of model structure from that of model calibration strategy when evaluating model performance in predicting streamflows. However, comparison of model structure alone would not be representative of how models are developed and used to make management decisions. Calibration strategies differ among modellers (even for a single model) and among models. Different model formulations predict state variables and fluxes in different ways and combinations and at different time steps, which affects how calibration is performed. One could remove some of the modeller choice bias by requiring all models to use automated calibration to a pre-specified set of metrics, but that is not how management models are typically constructed in practice. At the simplest level, if diverse models applied under diverse calibration strategies reach similar conclusions, then those conclusions might be considered to be robust and uncertainty in mean predictions is low; if the outcome of a model inference is highly dependent on the specific model or modeller strategy, then the inference is weaker and represents one realization from a broader statistical distribution of potential outcomes.

Other approaches to predicting ERFSs (e.g. regression models) have been recently shown to have better performance for low-flow statistics (Knight *et al*., 2012; Murphy *et al*., 2013). However, regression-based models are stationary, and therefore, rainfall–runoff and other physically based hydrologic models are still needed for evaluating global change and hydrologic alteration effects on aquatic ecosystems because they are more flexible and are capable of simulating scenarios of change. Additionally, even though some regression models appear to perform well in parts of the Southeastern USA, it is difficult to predict whether they will show the same level of performance or have a high level of transferability in the snowmelt driven Rocky Mountain States or in areas such as the Southwest USA where low flows predominate and where there are fewer gauges available to establish statistical relationships.

CONCLUSIONS

The primary objective of this study was to provide resource managers and environmental flow practitioners with some insight into the relative error in streamflow predictions among a subset of hydrologic models commonly used for water supply assessment, environmental flow studies and climate change predictions. All of the models evaluated were developed by different agencies, for different purposes, with different input data sets, and, in general, were calibrated to different degrees using different objective functions. It was not the intent of this study to determine whether one model was ‘better’ than another or to attempt to separate the influence of model structure from calibration strategy because such a comparison would not be representative of how models are developed and used to make management decisions. Rather, our objective was to evaluate the performance of seven hydrologic models of varying complexity and calibration strategy as developed to answer a variety of management questions. Our results do not support the hypothesis that simple, regional-scale models have less predictive power than more complex fine-scale models at a monthly time step. Differences among model predictions for specific fit statistics or ERFSs are as likely to be related to differences in model calibration strategy as they are related to differences in model structure. As a result, we do not provide recommendations of one hydrologic model over another based on the results of this study. Instead, we stress that it is incumbent upon resource managers, environmental flow practitioners and policy makers to consider the expertise of the modeller, the applicability of a model to a particular resource problem, the context to which the model is being applied, the availability of streamflow observations for model calibration for a given site and the important components of the flow regime that may be used for model calibration to minimize error across the targeted range of flows.

ACKNOWLEDGEMENTS

We are indebted to the many individuals and agencies that willingly participated in this modelling study and provided the data and flow model output that was integral to the comparisons. We thank Jerry McMahon for his guidance throughout this project and anonymous reviewers for providing timely reviews that greatly improved the final version of this manuscript. We thank North Carolina State University for providing the facilities for a model comparison workshop among agencies and modellers participating in this study. Jeremy Wyss (Tetra Tech) and Pushpa Tuppad (Texas A&M University) were responsible for calibrating the basin-scale HSPF and SWAT models, respectively. The HSPF and SWAT model applications were supported by the US EPA Office of Research and Development National Center for Environmental Assessment under the guidance of Dr Thomas Johnson; however, the views expressed in this paper represent those of the authors and do not necessarily reflect the views or policies of the US Environmental Protection Agency. Funding for the WaSSI model analysis was provided by the US Department of Agriculture Forest Service Eastern Forest Environmental Threat Assessment Center and the Coweeta Hydrologic Laboratory. Funding for this study and for a model comparison workshop was provided by the US Department of Interior, Southeast Climate Science Center. Additional support from the USGS National Water Census was greatly appreciated. This manuscript has benefitted from peer review from C. Andrew Dolloff (US Forest Service, Center for Forest Watershed Science) and Rodney R. Knight (USGS, Tennessee Water Science Center). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

REFERENCES

Annear T, Chisholm I, Beecher H, Locke A, Aarrestad P, Coomer C, Estes C, Hunt J, Jacobson R, Jobsis G, Kauffman J, Marshall J, Mayes K, Smith G, Wentworth R, Stalnaker C. 2004. *Instream Flows for Riverine Resource Stewardship, Revised Edition*. Instream Flow Council: Cheyenne.

Arthington AH, Bunn SE, Poff NL, Naiman RJ. 2006. The challenge of providing environmental flow rules to sustain river ecosystems. *Ecological Applications* **16**: 1311–1318. DOI: 10.1890/1051-0761(2006)016[1311:TCOPEF]2.0.CO;2

Averyt K, Fisher J, Huber-Lee A, Lewis A, Macknick J, Madden N, Rogers J, Tellinghuisen C. 2011. *Freshwater Use by U.S. Power Plants: Electricity's Thirst for a Precious Resource, a Report of the Energy and Water in a Warming World Initiative*. Union of Concerned Scientists: Cambridge.

Averyt K, Meldrum J, Caldwell P, Sun G, McNulty S, Huber-Lee A, Madden N. 2013. Sectoral Vulnerabilities to Changing Water Resources in the Coterminous U.S. *Environmental Research Letters* **8**: 035026. DOI: 10.1088/1748-9326/8/3/035046

Bicknell BR, Imhoff JC, Kittle JL, Jobes TH, Donigian AS. 2005. *HSPF Version 12.2 User's Manual.* U.S. Environmental Protection Agency, National Exposure Research Laboratory: Athens.

Biemans H, Haddeland I, Kabat P, Ludwig F, Hutles RWA, Heinke J, von Bloh W, Gerten D. 2011. Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water Resources Research* **47**: W03509. DOI: 10.1029/ 2009WR008929

Burnash RJC, Ferral RL, McGuire RA. 1973. *A Generalized Streamflow Simulation System – Conceptual Modeling for Digital Computers*, Technical Report, Joint Federal and State River Forecast Center. U.S. National Weather Service and California Department of Water Resources: Sacramento, CA.

Burnash RJC. 1995. The NWS river forecast system – catchment modeling. In *Computer Models of Watershed Hydrology*, Singh VP (ed.). Water Resources Publications: Littleton, CO.

Caldwell PV, Sun G, McNulty SG, Cohen EC, Moore Myers JA. 2012. Impacts of impervious cover, water withdrawals, and climate change on river flows in the conterminous US. *Hydrology and Earth System Sciences* **16**: 2839–2857. DOI: 10.5194/hess-16-2839-2012

Carlisle DM, Wolock DM, Meador MR. 2010. Alteration of streamflow magnitudes, and potential ecological consequences: a multiregional assessment. *Frontiers in Ecology and the Environment* **264**. DOI: 10.1890/100053

Crawford NH, Linsley RK. 1966. *Digital Simulation in Hydrology: Stanford Watershed Model IV*. Department of Civil Engineering, Stanford University: Stanford CA USA.

Di Luzio M, Johnson GL, Daly C, Eischeid JK, Arnold JG. 2008. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *Journal of Applied Meteorology and Climatology* **47**: 475–497. DOI: 10.1175/ 2007JAMC1356.1

Donigian AS Jr. 2000. HSPF training workshop handbook and CD. Lecture #19: calibration and verification issues. U.S. Environmental Protection Agency Headquarters, Washington Information Center, 10–14 January, 2000. Prepared for U.S. Environmental Protection Agency, Office of Water, Office of Science and Technology, Washington, DC.

Duan QY, Sorooshian S, Gupta VK. 1994. Optimal use of the SCE-UA Global Optimization Method for calibrating watershed models. *Journal of Hydrology* **158**: 265–284. DOI: 10.1016/0022-1694(94)90057-4

Dudgeon D, Arthington AH, Gessner MO, Kawabata ZI, Knowler DJ, Leveque C, Naiman RJ, Prieur-Richard AH, Soto D, Stiassny MLJ, Sullivan CA. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Review* **81**: 163–182. DOI: 10.1017/S1464793105006950

Falcone JA, Carlisle DM, Wolock DM, Meador MR. 2010. GAGES: a stream gage database for evaluating natural and altered flow conditions in the conterminous United States. *Ecology* **91**(2): 621.

Falcone JA. 2011. Geospatial attributes of gages for evaluating streamflow. Digital spatial data set. Available at: http://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml.

Foley JA, DeFries R, Asner GP, Barford C, Bonan G, Carpenter SR, Chapin FS, Coe MT, Daily GC, Gibbs HK, Helkowski JH, Holloway T, Howard EA, Kucharik CJ, Monfreda C, Patz JA, Prentice IC, Ramankutty N, Snyder PK. 2005. Global consequences of land use. *Science* **309**: 570–574. DOI: 10.1126/science.1111772

Freeman MC, Buell GR, Hay LE, Hughes WB, Jacobson RB, Jones JW, Jones SA, Lafontaine JH, Odom KR, Peterson JT, Riley JW, Schindler JS, Shea C, Weaver JD. 2013. Linking river management to species conservation using dynamic landscape-scale models. *River Research and Applications* **29**: 906–918. DOI: 10.1002/rra.2575

Fry J, Xian G, Jin S, Dewitz J, Homer C, Yang L, Barnes C, Herold N, Wickham J. 2011. Completion of the 2006 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering & Remote Sensing* **77**(9)**:** 858–864.

Georgakakos A, Zhang F, Yao H. 2010. *Climate Variability and Change Assessment for the ACF River Basin*. Southeast US. Georgia Water Resources Institute: Atlanta

Georgakakos A, Fleming P, Dettinger M, Peters-Lidard C, Richmond TC, Reckhow K, White K, Yates D. 2014. Ch. 3: water resources. In *Climate Change Impacts in the United States: the Third National Climate Assessment*, Melillo JM, Richmond TC, Yohe GW (eds). U.S. Global Change Research Program: Washington DC.

Gerten D, Rost S, von Blo W, Lucht W. 2008. Causes of change in 20th century global river discharge, *Geophysical Research Letters* **35:** L20405. DOI: 10.1029/2008GL035258

Graf WL. 1999. Dam nation: a geographic census of American dams and their large-scale hydrologic impacts. *Water Resources Research* **35**: 1305–1311.

Haith DA, Shoemaker LL. 1987. Generalized watershed loading functions for stream flow nutrients. *Water Resources Bulletin* **23**: 471–478.

Haith DA, Mandel R, Wu RS. 1992. *GWLF: Generalized Watershed Loading Functions User's Manual, Vers. 2.0*. Cornell University: Ithaca, NY.

Hamon WR. 1963. Computation of direct runoff amounts from storm rainfall. *International Association of Scientific Hydrology. Publication* **63**: 52–62.

Hay LE, McCabe GJ. 2002. Spatial variability in water-balance model performance in the conterminous United States. *Journal of the American Water Resources Association* **38**: 847–860. DOI: 10.1111/j.1752-1688.2002.tb01001.x

Hay LE, Umemoto M. 2006. Multiple-objective stepwise calibration using Luca. *U.S. Geological Survey Open-File Report* 2006-1323.

Hay LE, Leavesley GH, Clark MP, Markstrom SL, Viger RJ, Umemoto M. 2006. Step wise, multiple objective calibration of a hydrologic model for a snowmelt dominated basin. *Journal of the American Water Resources Association* **42**: 877–890. DOI: 10.1111/j.1752-1688.2006.tb04501.x

Henriksen JA, Heasley J, Kennen JG, Nieswand S. 2006. Users' manual for the Hydroecological Integrity Assessment Process software (including the New Jersey Assessment Tools). *U.S. Geological Survey Open-File Report* 2006-1093.

Homer C, Dewitz J, Fry J, Coan M, Hossain N, Larson C, Herold N, McKerrow A, VanDriel JN, Wickham J. 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering & Remote Sensing* **73**: 337–341.

Jackson RB, Carpenter SR, Dahm CN, McKnight DM, Naiman RJ, Postel SL, Running SW. 2001. Water in a changing world. *Ecological Applications* **11**: 1027–1045. DOI: 10.2307/3061010

Jensen ME, Haise HR. 1963. Estimating evapotranspiration from solar radiation. *Journal of Irrigation and Drainage Division* **89**: 15–41.

Jensen ME, Burman RD, Allen RG. 1990. *Evapotranspiration and Irrigation Water Requirements*. ASCE: New York, NY.

Johnson T, Butcher J, Weaver C. 2012. Investigating the implications of climate change for U.S. stream water quality: the EPA Global Change Research Program's "20 Watersheds" project. *Journal of Water Resources Planning and Management* **138**(5): 453–464.

Kendy E, Cutrofello M, Fekadu M. 2011. Unified Hydrologic Model for Assessing Human and Climate Impacts on Streamflows at Multiple Geographic Scales. Available at: http://data.southatlanticlcc.org/TNC_flows_for_web.pdf. [Accessed 9 December 2013].

Kennen JG, Henriksen JA, Nieswand SP. 2007. Development of the Hydroecological Integrity Assessment Process for Determining Environmental Flows for New Jersey Streams. *U.S. Geological Survey Scientific Investigations Report* 2007-5206.

King RS, Baker ME, Whigham DF, Weller DE, Jordan TE, Kazyak PF, Hurd MK. 2005. Spatial consideration for linking watershed land cover to ecological indicators in streams. *Ecological Applications* **15**: 137–153. DOI: 10.1890/04-0481

Knight RR, Gain WS, Wolfe WJ. 2012. Modelling ecological flow regime: an example from the Tennessee and Cumberland River basins. *Ecohydrology* **5**: 613–627. DOI: 10.1002/eco.246

Knight RR, Murphy JC, Wolfe WJ, Saylor CF, Wales AK. 2013. Ecological limit functions relating fish community response to hydrologic departures of the ecological flow regime in the Tennessee River basin, United States. *Ecohydrology*. DOI: 10.1002/eco.1460

Koren V, Smith M, Duan Q. 2003. Use of a priori parameter estimates in the derivation of spatially consistent parameter sets of rainfall–runoff models. In *Calibration of Watershed Models Water Science and Applications*, Vol. **6**, Duan Q, Sorooshian S, Gupta H, Rosseau H, Turcotte H (eds). AGU: Washington, D.C.

LaFontaine JH, Hay LE, Viger RJ, Markstrom SL, Regan RS, Elliott CM, Jones JW. 2013. Application of the Precipitation–Runoff Modeling System (PRMS) in the Apalachicola–Chattahoochee–Flint River Basin in the Southeastern United States. *U.S. Geological Survey Scientific Investigations Report* 2013-5162.

Leavesley GH, Lichty RW, Troutman BM, Saindon LG. 1983. Precipitation–Runoff Modeling System – User's manual. *U.S. Geological Survey Water-Resources Investigations Report* 83-4238.

Lockaby G, Nagy C, Vose JM, Ford CR, Sun G, McNulty S, Caldwell P, Moore Myers JA. 2011. Water and Forests. In *The Southern Forest Futures Project: Technical Report*, Wear DN, Greis JG (eds). USDA Forest Service Southern Research Station: Asheville, NC.

Lumb AM, McCammon RB, Kittle JL Jr. 1994. User's manual for an expert system (HSPEXP) for calibration of the hydrological simulation program-FORTRAN. U.S. Geological Survey Water Resources Investigation Report 94-4168. U.S. Geological Survey, Reston, VA.

Marion DA, Sun G, Caldwell PV, Ford CR, Ouyang Y, Amatya DM, Clinton BD, Conrads PA, Gull Laird S, Dai Z, Clingenpeel JA, Liu Y, Roehl EA, Moore Myers JA, Trettin C. 2013. Managing forest water quantity and quality under climate change in the Southern US. In *Climate Change Adaptation and Mitigation Management Options*, Vose J (ed.). CRC Press: Boca Raton, FL.

Markstrom SL, Niswonger RG, Regan RS, Prudic DE, Barlow PM. 2008. GSFLOW – Coupled groundwater and surface-water flow model based on the integration of the Precipitation–Runoff Modeling System (PRMS) and the Modular Ground-Water Flow Model (MODFLOW – 2005). In U.S. Geological Survey Techniques and Methods, book 6.

Markstrom SL, Hay LE, Ward-Garrison CD, Risley JC, Battaglin WA, Bjerklie DM, Chase KJ, Christiansen DE, Dudley RW, Hunt RJ, Koczot KM, Mastin MC, Regan RS, Viger RJ, Vining KC, Walker JF. 2012. Integrated watershed-scale response to climate change for selected basins across the United States. *U.S. Geological Survey Scientific Investigations Report* 2011-5077.

Masters LL, Flack SR, Stein BA. 1998. *Rivers of Life: Critical Watersheds for Protecting Freshwater Biodiversity*. The Nature Conservancy: Arlington, VA.

Matthews R. 2005. A six-step framework for ecologically sustainable water management. *Journal of Contemporary Water Research and Education* **131**: 60–65. DOI: 10.1111/j.1936-704X.2005.mp131001010.x

Maurer EP, Wood AW, Adam JC, Lettenmaier DP, Nijssen B. 2002. A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States. *Journal of Climate* **15**: 3237–3251.

McCabe GJ, Markstrom SL. 2007. *A Monthly Water-balance Model Driven by a Graphical User Interface*. U.S. Geological Survey Open-File Report 2007-1088: Reston, VA.

McCabe GJ, Wolock DM. 2011. Independent effects of temperature and precipitation on modeled runoff in the conterminous United States. *Water Resources Research* **47**: W11522. DOI: 10.1029/2011WR010630

Monteith JL. 1965. Evaporation and the environment. *The State and Movement of Water in Living Organisms XIXth Symposium, Swansea. Soc Exp Biology*. Cambridge University Press: Cambridge.

Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the American Society of Agricultural and Biological Engineers* **50**(3): 885–900.

Murphy JC, Knight RR, Wolfe WJ, Gain WS. 2013. Predicting ecological flow regime at ungaged sites: a comparison of methods. *River Research and Applications* **29**: 660–669. DOI: 10.1002/rra.2570

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models, I, a discussion of principles. *Journal of Hydrology* **10**: 282–290.

NCEFSAB. 2013. Recommendations for estimating flows to maintain ecological integrity in streams and rivers in North Carolina. Available at:

http://www.ncwater.org/files/eflows/sab/EFSAB_Final_Report_to_NCDENR.pdf. [Accessed 6 December 2013].

Neitsch SL, Arnold JG, Kiniry JR. 2005. Soil and water assessment tool, theoretical documentation. Temple, TX, Grassland, Soil and Water Research Laboratory, U.S. Department of Agriculture, Agricultural Research Service.

Olden JD, Poff NL. 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications* **19**: 101–121. DOI: 10.1002/rra.700

Poff NL, Ward JV. 1989. Implications of streamflow variability and predictability for lotic community structure – a regional analysis of streamflow patterns. *Canadian Journal of Fisheries and Aquatic Sciences* **46**: 1805–1818. DOI: 10.1139/f89-228

Poff NL, Zimmerman JKH. 2010. Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. *Freshwater Biology* **55**: 194–205. DOI: 10.1111/j.1365-2427.2009.02272.x

Poff NL, Allan JD, Bain MB, Karr JR, Prestegaard KL, Richter BD, Sparks RE, Stromberg JC. 1997. The natural flow regime – a paradigm for river conservation and restoration. *BioScience* **47**: 769–784. DOI: 10.2307/1313099

Poff NL, Olden JD, Merritt DM, Pepin DM. 2007. Homogenization of regional river dynamics by dams and global biodiversity implications. *Proceeding of the National Academy of Science* **104**: 5732−5737. DOI: 10.1073/pnas.0609812104

Poff NL, Richter BD, Arthington AH, Bunn SE, Naiman RJ, Kendy E, Acreman M, Apse C, Bledsoe BP, Freeman MC, Henriksen J, Jacobson RB, Kennen JG, Merritt DM, O'keeffe JH, Olden JD, Rogers K, Tharme RE, Warner A. 2010. The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards. *Freshwater Biology* **55**: 147–170. DOI: 10.1111/j.1365-2427.2009.02204.x

Price CV, Nakagaki N, Hitt KJ, Clawges RM. 2006. Enhanced historical land-use and land-cover data sets of the US Geological Survey. US Geological Survey Data Series 240.

PRISM Climate Group. 2013. PRISM Climate Data. Available at: http://www.prism.oregonstate.edu/ [Accessed 01 July 2013].

Richter BD, Baumgartner JV, Wigington R, Braun DP. 1997. How much water does a river need? *Freshwater Biology* **37**: 231–249. DOI: 10.1046/j.1365-2427.1997.00153.x

SALCC. 2012. South Atlantic Landscape Conservation Cooperative Science Assessment 2012. Available at: http://www.southatlanticlcc.org/page/science-assessment.

SARP. 2004. *Conserving Our Southeastern Aquatic Habitat: The Southeast Aquatic Habitat Plan*. Available at: http://southeastaquatics.net/documents/southeast-aquatic-habitat-plan.

SAS Institute Inc. 1989. *SAS/STAT Users Guide. Version 6*, Vol. **2**, 4th edn. SAS Institute Inc.: Cary, NC.

Stanford JA, Ward JV, Liss WJ, Frissell CA, Williams RN, Lichatowich JA, Coutant CC. 1996. A general protocol for restoration of regulated rivers. *Regulated Rivers: Research and Management* **12**: 391–414.

Sun G, Alstad K, Chen J, Chen S, Ford CR, Lin G, Lu N, McNulty SG, Noormets A, Vose JM, Wilske B, Zeppel M, Zhang Y, Zhang Z. 2011a. A general predictive model for estimating monthly ecosystem evapotranspiration. *Ecohydrology* **4**: 245–255. DOI: 10.1002/eco.194

Sun G, Caldwell P, Noormets A, Cohen E, McNulty S, Treasure E, Domec JC, Mu Q, Xiao J, John R, Chen J. 2011b. Upscaling key ecosystem functions across the conterminous United States by a water-centric ecosystem model. *Journal of Geophysical Research* **116**: G00J05. DOI: 10.1029/2010JG001573

Sun G, Caldwell PV, Georgakakos AP, Arumugam S, Cruise J, McNider RT, Terando A, Conrads PA, Feldt J, Misra V, Romolo L, Rasmussen TC, McNulty SG, Marion DA. 2013. Impacts of climate change and variability on water resources in the Southeastern US. In *Southeast Regional Technical Report to the National Climate Change Assessment, Water Resources*, Ingram KT, Dow K, Carter L (eds). Island Press: Washington DC.

Tavernia BG, Nelson MD, Caldwell P, Sun G. 2013. Water stress projections in the Northeastern and Midwestern United States in 2060: anthropogenic and ecological consequences. *Journal of the American Water Resources Association* **49**: 938–952. DOI: 10.1111/jawr.12075

Thornthwaite CW. 1948. An approach toward a rational classification of climate. *Geographical Review* **38**: 55–94.

Thornton PE, Thornton MM, Mayer BW, Wilhelmi B, Wei Y, Cook RB. 2013. Daymet: Daily Surface Weather on a 1 km grid for North America, 1980–2012. Available at: http://daymet.ornl.gov/

USDA Soil Conservation Service. 1972. *Hydrology Guide for Use in Watershed Planning. National engineering handbook, section 4: hydrology, supplement A*. U.S. Department of Agriculture, Natural Resources Conservation Service: Washington, D.C.

USEPA. 2008. *Using the BASINS Meteorological Database – Version 2006*. Office of Water: Washington, D.C.

USEPA. 2011. *Healthy Watersheds Initiative: National Framework and Action Plan*. USEPA Office of Water: Washington, D.C.

USEPA. 2013. *Watershed Modeling to Assess the Sensitivity of Streamflow, Nutrient, and Sediment Loads to Potential Climate Change and Urban Development in 20 U.S. Watersheds*. National Technical Information Service: Washington, D.C.

Van Sickle J. 2003. Analyzing correlations between stream and watershed attributes. *Journal of the American Water Resources Association* **39**: 717–7126. DOI: 10.1111/j.1752-1688.2003.tb03687.x

Wenger SJ, Luce CH, Hamlet AF, Isaak DJ, Neville HM. 2010. Macroscale hydrologic modeling of ecologically relevant flow metrics. *Water Resources Research* **46**: W09513. DOI: 10.1029/2009WR008839

