HYDROLOGICAL PROCESSES
Hydrol. Process. **28**, 3505–3520 (2014)
Published online 18 June 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.9890

Comparison of radar and gauge precipitation data in watershed models across varying spatial and temporal scales

Katie Price,* S. Thomas Purucker, Stephen R. Kraemer, Justin E. Babendreier and Chris D. Knightes

Ecosystems Research Division, National Exposure Research Laboratory, Office of Research and Development, US Environmental Protection Agency, 960 College Station Rd, Athens, GA, 30605, USA

Abstract:

Precipitation is a key control on watershed hydrologic modelling output, with errors in rainfall propagating through subsequent stages of water quantity and quality analysis. Most watershed models incorporate precipitation data from rain gauges; higher-resolution data sources are available, but they are associated with greater computational requirements and expertise. Here, we investigate whether the Multisensor Precipitation Estimator (MPE or Stage IV Next-Generation Radar) data improve the accuracy of streamflow simulations using the Soil and Water Assessment Tool (SWAT), compared with rain gauge data. Simulated flows from 2002 to 2010 at five timesteps were compared with observed flows for four nested subwatersheds of the Neuse River basin in North Carolina (21-, 203-, 2979-, and 10 100-km^2 watershed area), using a multi-objective function, informal likelihood-weighted calibration approach. Across watersheds and timesteps, total gauge precipitation was greater than radar precipitation, but radar data showed a conditional bias of higher rainfall estimates during large events (>25–50 mm/day). Model parameterization differed between calibrations with the two datasets, despite the fact that all watershed characteristics were the same across simulation scenarios. This underscores the importance of linking calibration parameters to realistic processes. SWAT simulations with both datasets underestimated median and low flows, whereas radar-based simulations were more accurate than gauge-based simulations for high flows. At coarser timesteps, differences were less pronounced. Our results suggest that modelling efforts in watersheds with poor rain gauge coverage can be improved with MPE radar data, especially at short timesteps. Published 2013. This article is a U.S. Government work and is in the public domain in the USA.

KEY WORDS Stage IV NEXRAD; MPE; SWAT; scale; spatiotemporal

Received 22 November 2012; Accepted 2 May 2013

INTRODUCTION

Watershed hydrologic models are widely used to simulate streamflow for water quality and ecosystem services and to evaluate alternative future scenarios associated with land use and climate change. Simulation of streamflow, sediment, and dissolved constituents requires climatic forcing data, minimally precipitation and temperature. Precipitation is often identified as the key variable in modelling watershed hydrology (Attorre *et al*., 2007; Schuurmans and Bierkens, 2007), and reducing the uncertainty of precipitation data has been shown to have a substantial stabilizing effect on model parameterization and calibration (Andréassian *et al*., 2001; Moulin *et al*., 2009; Looper *et al*., 2012). Temperature can be reasonably estimated from a sparse network of stations within and surrounding the study watershed (Attorre *et al*., 2007); however, capturing precipitation spatial and temporal variability successfully from available resources has been a challenge for hydrologic modelling. Failure to incorporate such variability introduces unquantified uncertainty to hydrologic and fate-and-transport modelling efforts (Jordan *et al*., 2000; Andréassian *et al*., 2001; Kalinga and Gan, 2006; Schuurmans and Bierkens, 2007; Villarini *et al*., 2008a; Golden *et al*., 2010).

The main sources of precipitation data are rain gauges, Doppler weather radar, and satellite sensors, with many subcategories of each, and each data source has associated specific considerations of instrumental uncertainty and spatio-temporal resolution (Moreau *et al*., 2009). Although satellite data sources have the unique advantage of providing global coverage, the spatio-temporal resolution is coarser compared with North American and European weather radar networks and, as an emergent technology, has not been adequately validated. At present, rain gauge and weather radar data remain the best options for most watershed modelling applications within the USA and Europe and are thus the focus of this comparison.

*Correspondence to: Katie Price, Ecosystems Research Division, National Exposure Research Laboratory, Office of Research and Development, US Environmental Protection Agency, 960 College Station Rd, Athens, GA 30605, USA.
E-mail: price.katie.m@gmail.com

Rain gauge precipitation data

Most commonly used watershed models are designed to readily incorporate data from rain gauges. Although gauge data can be assumed to represent the most accurate source of information at the exact location of the gauge, precipitation is known to be highly spatially variable, and areal totals are often poorly represented by point gauges (Strangeways, 2007; Starks and Moriasi, 2009; Tobin and Bennett, 2009). Most agencies providing publicly available rainfall data employ tipping-bucket rain gauges (Habib *et al*., 2001; Molini *et al*., 2005). These gauges collect falling rain in a bucket of known volume that tips and drains when full, and the time of each bucket tip is recorded. Rain gauges are typically treated as the actual amount of rainfall by hydrologists, although there are known instrumental errors associated with rain gauges (Rinehart, 2004; Sieck *et al*., 2007). Many studies have shown a conditional negative bias of rain gauges during heavy rainfall, as there is a tendency for water loss due to wind and erratic behaviour of the mechanical aspects of the gauge during intense rainfall (Molini *et al*., 2005; Lanza and Stagi, 2008).

Radar precipitation data

Operated by the National Weather Service (NWS), the Next-Generation Radar (NEXRAD) is a network of 178 Doppler weather radars (WSR-88D) across the USA, from which high-spatial-resolution precipitation estimates are derived. The predominant source of uncertainty in radar precipitation estimates is introduced by the assumed relationship between reflectivity and rainfall amount ('Z/R relationship'), which varies by precipitation type (Rinehart, 2004). Because precipitation is sensed well above the ground surface, rainfall detected by the radar may move large distances downwind or evaporate before reaching the ground. Further uncertainty arises from the radar technology itself, such as beam blockage and bright band errors (Rinehart, 2004). Radar precipitation estimates that have been adjusted on the basis of gauge data (gauge-conditioned) are consistently superior to unadjusted radar data (Yilmaz *et al*., 2005; Young and Brunsell, 2008). Gauge conditioning of NEXRAD Stage II data is performed using the Multisensor Precipitation Estimation (MPE) algorithm to create the NEXRAD Stage IV data, with rain gauge values taken from weather stations within the NWS (2012) Hydrometeorological Automated Data System network. MPE data are obtained in grid format at 4-km pixel resolution, provided at an hourly timestep (NCAR Earth Observing Laboratory, 2011). Although available since the 1990s, gauge-conditioned radar data are considered most accurate after 2002, when the MPE algorithm was implemented (Westcott *et al*., 2008). The high spatial resolution of MPE data makes them appealing in hydrologic modelling, particularly when there are no rain gauges within the watershed boundaries. Most watershed models were not designed to incorporate gridded precipitation data, however, and a computationally intensive data conversion process is currently required to use MPE data, although streamlined processes can be expected in coming years.

It is not presently understood how the accuracy of streamflow simulations using radar and gauge data sources differ, particularly across ranges of spatial and temporal scales, which is, thus, the impetus and contribution of this study. Whereas existing studies have compared rain gauge and radar streamflow simulations in other contexts (e.g., Moon *et al*., 2004, Kalin and Hantush, 2006; Palanisamy, 2006; Golden *et al*., 2010), results have not been consistent, and it remains unresolved how simulations compare across watersheds of different size, with varied rain gauge coverage, and across timesteps. In the absence of a very dense, tightly quality-controlled network of rain gauges, it is inappropriate to treat any rainfall data source as actual precipitation because of known instrument uncertainty and spatial variability concerns (Rinehart, 2004; Jayakrishnan *et al*., 2005; Villarini *et al*., 2008b; Stisen and Sandholdt, 2010). An alternative method is to use streamflow simulations from a watershed model as an indirect assessment of the precipitation data accuracy, because watershed models spatially and temporally integrate rainfall data (e.g. Moon *et al*., 2004; Kalin and Hantush, 2006; Schuurmans and Bierkens, 2007; Sexton *et al*., 2010; Biggs and Atkinson, 2011; Bitew *et al*., 2012). In this approach, deviations between simulated and observed flows are attributed to errors in the precipitation data. Although it is acknowledged that model structure and other input data sources introduce uncertainty to simulated streamflows, everything other than precipitation is held constant in this type of comparative framework, isolating precipitation as the driver of differences between simulation results.

Objectives

Many authors have emphasized the need for better accounting of spatial variability of rainfall in watershed modelling (e.g. O'Connell and Todini, 1996; Bell and Moore, 2000; Arnaud *et al*., 2002; Schuurmans and Bierkens, 2007), which encourages the use of radar data in place of rain gauge data. It remains unclear whether the higher spatial resolution of radar rainfall data is associated with equal or greater accuracy of average precipitation depth than that of gauge data, or at which scales and timesteps higher spatial resolution precipitation data produce better model calibrations and streamflow simulations. This is especially important given the greater computational demands of gridded precipitation data, compared with gauge data in most watershed models. The objective of this research is to go beyond an accuracy assessment of rainfall measurements, by addressing the use of varied precipitation data types as forcing data at multiple spatial and temporal scales of model operation.

METHODS

We compared Soil and Water Assessment Tool (SWAT) streamflow simulations with observed flows for accuracy assessment in four watersheds, at five temporal aggregations. Comparisons were performed by separately calibrating and simulating streamflow for each watershed with both gauge and radar precipitation, holding other model inputs constant. Spatial scales of analysis for the study include four nested sub-basins of the Neuse River watershed, ranging from 21 to 10 100 km^2. Temporal scales of analysis included daily, weekly, monthly, quarterly, and annual precipitation and streamflow totals.

Site selection and study area

The comparison of gauge and radar precipitation data was conducted in four nested sub-basins of the Neuse River in the North Carolina Piedmont and Coastal Plain (Figure 1): Mountain Creek (MC, 21 km^2), Little River (LR, 203 km^2), Neuse River near Clayton (NC, 2979 km^2), and Neuse River near Fort Barnwell (FB, 10 100 km^2). The Neuse River basin was selected for study because of multi-agency concerns related to excessive upland nutrient loading and eutrophication in the Albemarle-Pamlico estuary (Johnston *et al.*, 2011). Topographic relief in this region is not pronounced, precluding concerns about topographic radar shadowing. This study area is representative of much of the Southeastern and Mid-Atlantic regions of the USA, in terms of both watershed hydrologic processes and precipitation types and distributions.

Watershed selection was determined by multiple criteria: (1) Each watershed outlet needed to coincide with a US Geological Survey (USGS) streamflow gauge; (2) We sought to create a nest of watersheds spanning multiple orders of magnitude in an area; (3) Watersheds with reservoirs covering a >25% area were avoided; and (4) USGS gauges located <50 km downstream of a reservoir outlet were excluded. Criteria 3 and 4 were included because of the artificial effects of reservoirs on watershed rainfall–run-off response.

Piedmont portions of the Neuse watershed are characterized by moderate-relief, crystalline bedrock and saprolite, with deep, weathered soils (Mills *et al.*, 1987). Coastal Plain portions are underlain by clastic alluvial and marine sediments (Walker and Coleman, 1987; Leigh,

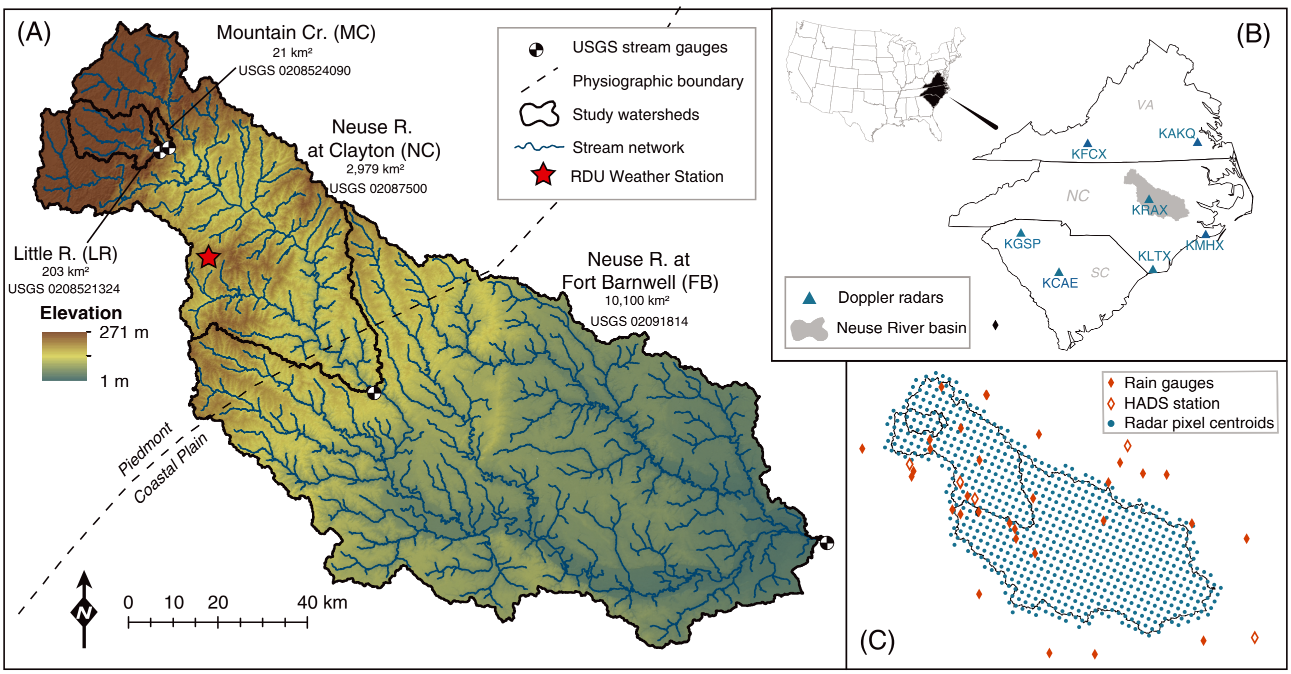

Figure 1. Study area. (A) Four subwatersheds of the Neuse River were included in this study. (B) The location of the Neuse River watershed in North Carolina and the surrounding NEXRAD Doppler radar stations used to create Multisensor Precipitation Estimator (MPE) precipitation data used in this study. (C) Locations of National Climate Data Center and North Carolina CRONOS rain gauges and the centroids of MPE pixels. Hydrometeorological Automated Data System (HADS) rain gauges are a subset of National Weather Service gauges used for gauge correction of radar data to produce the MPE data product. Note the much higher density of radar pixels compared with that of rain gauges

2008). There are no significant areas of karst within the Neuse system. Climate normals (1981–2010) from the Raleigh–Durham International Airport (RDU) weather station (Figure 1A) are average annual precipitation of 1101 mm and average January and July temperatures of 5.0 and 26.7 °C, respectively (National Climate Data Center, NCDC, 2011). Snowfall is not a dominant part of regional hydrology. Land use in the study area is predominantly forest and agriculture (Table I).

Watershed modelling

We used the 2009 version of the SWAT for streamflow simulation (Neitsch *et al.*, 2011). SWAT is a semidistributed model, in the sense that all areas having the same combination of soil, land use, and slope are lumped together as hydrologic response units (HRUs), and water is not directly routed from one landscape unit to the next downslope unit. SWAT uses the curve number II (CN_2) for partitioning overland flow and infiltration for each HRU (Hawkins, 2009). We used the ArcGIS 9 version of the ArcSWAT extension (Winchell *et al.*, 2007) for data preprocessing and model set-up. Daily streamflows were simulated for all four watersheds from 2001 to 2010, with 2001 values discarded as model spin-up. All default options for SWAT operation were used, except that the Hargreaves equation for evapotranspiration was substituted for the default Penman–Monteith equation. Hargreaves produced much better uncalibrated fits between simulated and observed flows, particularly with low flows, as has been observed in other SWAT applications (Wang *et al.*, 2006; Setegn *et al.*, 2008).

SWAT input data

SWAT input requirements include (1) meteorological data (minimally daily total precipitation and maximum and minimum temperature), (2) spatial layers representing soil type, elevation, and land use, and (3) observed streamflow for model calibration.

All meteorological variables (precipitation, temperature, etc.) are assigned to each SWAT sub-basin on the basis of the closest station to the geographic centre of each. Daily maximum and minimum temperature data for the study period (2001–2010) were obtained from the NCDC (2011). Rain gauge precipitation data were obtained from the NCDC and North Carolina CRONOS databases (NCDC, 2011; State Climate Office of North Carolina, 2011). Stations missing more than 10% of daily observations were excluded, and missing values below this threshold were filled from the nearest station (Ngongondo *et al.*, 2011). Hourly Doppler radar-derived, gauge-adjusted MPE precipitation data were obtained from the NCAR Earth Observing Laboratory (2011), aggregated to daily timestep, and converted from daily grids to time series for SWAT (Price *et al.*, 2011). The centroid of each MPE pixel was treated as a 'virtual' precipitation gauge in SWAT (Figure 1C).

The US General Soil Map (STATSGO2) is included in SWAT's database and was used here (Soil Survey Staff, 2011). The 2006 National Land Cover Dataset was combined with the 2009 Cropland Data Layer to represent land use (USGS, 2011; NASS, 2011), and a 30-m digital elevation model was used. Observed daily streamflow time series for model calibration and validation were

Table I. Watershed characteristics

		MC	LR	NC	FB
Area (km^2)		21	203	2979	10 100
Total relief (m)		98	126	249	270
Sub-basins		4	18	249	890
HRUs		46	241	4850	15 813
Land use[a] (%)					
Forest	Deciduous	55.4	63.6	44..2	18.8
	Evergreen	1.2	3.3	13.0	12.8
	Mixed	0.0	0.0	1.4	3.5
Agriculture	Pasture/grass	34.2	26.6	12.6	9.1
	Row crop	0.0	0.0	0.1	29.2
Developed	Low intensity	8.5	5.0	20.6	11.8
	Medium intensity	0.1	0.1	2.9	1.4
	High intensity	0.0	0.0	0.7	0.4
Wetland		0.3	1.0	2.2	11.5
Open water		0.3	0.4	2.3	1.5

MC, Mountain Creek; LR, Little River; NC, Neuse River near Clayton; FB, Neuse River near Fort Barnwell; HRU, hydrologic response unit.
[a] Classification from 2006 National Land Cover Database (*USGS*, 2011).

obtained from four USGS gauges, located at the outlet of each study watershed (Figure 1A).

SWAT spatial discretization

During preprocessing in ArcSWAT, each study watershed was subdivided in two hierarchical ways. First, each watershed was divided into topographically defined sub-basins. The size of sub-basins is determined by a user-defined flow accumulation area threshold and/or user specification of sub-basin outlets. Meteorological variables are assigned uniformly across each sub-basin, and sub-basins were delineated to ensure that all radar pixels were incorporated as precipitation stations (to maximally utilize the higher-resolution precipitation data). Sub-basins are further discretized into HRUs, in which all pixels sharing the same combined land use, soils, and slope class are assigned a uniform flow response.

Calibration

We used a multi-objective, iterative, informal, likelihood approach for model calibration (Price *et al.*, 2012). A total of eight separate calibrations were performed, one for each precipitation data type, for each of the four watersheds. Initial sensitivity analyses and manual calibrations indicated the influence of 12 calibration parameters (Table II). A split-sample approach was used for calibration and validation (Klemes, 1986), where the 9-year simulation period was divided into a calibration period (2002–2007) and validation period (2008–2010). Our multi-objective function, the composite likelihood (CL), equally emphasized the Nash–Sutcliffe efficiency (Nash and Sutcliffe, 1970), modified Nash–Sutcliffe efficiency (Krause *et al.*, 2005), and standard deviation ratios of observed and simulated flows (Moriasi *et al.*, 2007).

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{n}\left(\mathrm{O_t} - S_t\right)^2}{\sum_{t=1}^{n}\left(\mathrm{O}_t - \overline{\mathrm{O}}\right)^2} \tag{1}$$

$$\mathrm{MNS} = 1 - \frac{\sum_{t=1}^{n}\left|\mathrm{O_t} -, \mathrm{S_t}\right|}{\sum_{t=1}^{n}\left|\mathrm{O_t} - \overline{\mathrm{O}}\right|} \tag{2}$$

$$RSD = \frac{\sigma_s}{\sigma_o} \tag{3}$$

O and S indicate observed and simulated flow values, respectively, timesteps t range from 1 to n, and σ represents the standard deviation.

The CL index is designed to incorporate extreme flows, median flows, and flow variability in model calibration, as opposed to emphasizing a single flow modality by using a single objective function (Price *et al.*, 2012). Each individual objective function (Equations (1–3)) was calculated in R statistical software (R Development Core Team, 2011), using the hydroGOF package (Bigiarni, 2010). The Nash–Sutcliffe efficiency, modified Nash–Sutcliffe efficiency, and standard deviation ratios of observed and simulated flows were truncated or scaled to range from 0 to 1 (Equations (4–6)) and then combined in a simple average to calculate CL (Equation (7)):

$$\theta_{\mathrm{NSE}} = \max(0, NSE) \tag{4}$$

$$\theta_{\mathrm{MNS}} = \max(0, MNS) \tag{5}$$

$$\theta_{\mathrm{RSD}} = 1 - \min(1, |1 - RSD|) \tag{6}$$

$$\mathrm{CL} = \overline{\theta} \tag{7}$$

Table II. Soil and Water Assessment Tool calibration parameters

Parameter	Definition[a]	Scale	Adjustment[b]	Initial range
CN_2	SCS curve number, moisture condition II	HRU	x	−1 to 1
Alpha_Bf	Baseflow alpha factor (days)	Watershed	=	0 to 1
GW_Revap	Groundwater evaporation coefficient	Watershed	=	0.02 to 0.2
CH_N2	Channel Manning's coefficient	Watershed	=	0 to 0.3
CH_K2	Channel hydraulic conductivity (mm/h)	Watershed	=	0 to 150
GWQMN	Groundwater depth for return flow (mm)	Watershed	=	0 to 1000
SOL_AWC	Available soil water capacity	HRU	x	−0.2 to 0.6
SOL_K	Soil hydraulic conductivity (mm/h)	HRU	x	−0.5 to 0.5
SOL_Z	Soil depth (mm)	HRU	x	−0.25 to 0.25
CANMX	Maximum canopy index	HRU	=	0 to 25
ESCO	Soil evaporation compensation factor	HRU	=	0.001 to 1
SURLAG	Surface run-off lag coefficient	Watershed	=	0 to 24

HRU, hydrologic response unit.
[a] Source: Neitsch *et al.* (2009).
[b] 'x' indicates that the original value is multiplied by the adjustment factor, '=' indicates that the original value is replaced by the adjustment factor.

After identifying sensitive parameters and establishing initial ranges for each, we used the SWAT Calibration and Uncertainty Procedure subroutine SUFI-2 for Latin hypercube sampling (LHS) to generate multidimensional parameter value combinations across a set of 2001 SWAT simulations runs. SUFI-2 was also used to execute parallelized model runs (Abbaspour, 2009). CL fit scores were calculated for each simulation, and informal likelihoods were assigned to each parameter set using R statistical software. The 16th and 84th percentile values (±1 standard deviation) were extracted from the weighted likelihoods for each parameter and used as input ranges for the next pass of LHS and 2001 simulations. Iteration was stopped when the average CL score failed to improve by at least 5%. The parameter ranges associated with the best calibration pass were used for the validation period, for which separate fit scores and summary statistics were calculated. A detailed explanation of the calibration method can be found in Price *et al*. (2012).

Model simulations were parallelized using the SuperMUSE computing facility (Babendreier and Castleton, 2005). The iterative calibration process for all four watersheds required 30 626 total processor hours (equivalent to 3 days of full-capacity SuperMUSE runtime for single-threaded Windows operations).

Analysis

R software was used for all statistical analyses and data manipulation. Daily, weekly, monthly, quarterly, and annual precipitation totals for each sub-basin were compared for the radar and gauge datasets, using pairwise difference of means tests and Spearman rank correlation analyses, after all daily values for which both data types reported zero precipitation were removed (Yoo *et al*., 2010). Per cent difference was calculated for each of the four study watersheds for each day and month, and locally weighted regression (LOESS) was fit through all observed differences. Paired *t*-tests were conducted for total watershed precipitation using the area-weighted average sub-basin precipitation. Local polynomial regression fitting (using the 'LOESS' function in the base R statistics package) was used to evaluate changes in precipitation data bias due to event size. Daily streamflows were aggregated to weekly, monthly, quarterly, and annual timesteps for comparison of simulated and observed flows at multiple temporal scales using the eXtensible Time Series package (Ryan and Ulrich, 2012). The 5th, 50th, and 95th percentile flows were calculated for simulated and observed flows in each watershed, to represent low, medium, and high flows, respectively.

RESULTS

Precipitation differences

Gauge total precipitation was greater than radar total precipitation across the entire study period (Figure 2), and this positive bias was statistically significant at all timesteps and spatial aggregation scenarios considered (Table III). Although gauges recorded greater precipitation overall than radar, the LOESS results indicate a conditional negative bias for gauges during larger events

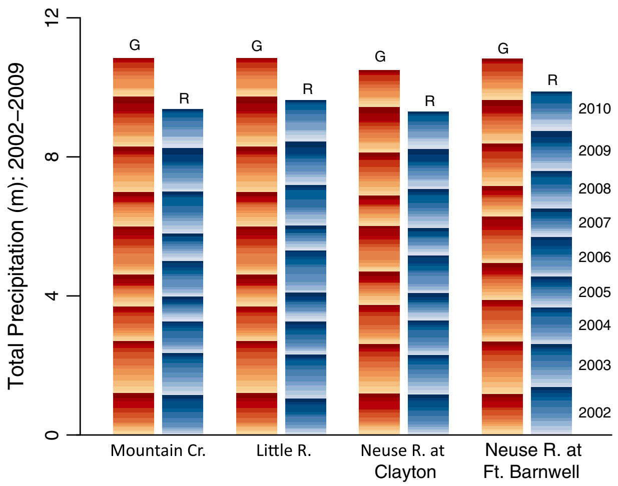

Figure 2. Total gauge (G) and radar (R) measured precipitation over the entire study period. Precipitation totals represent 9-year total depth per unit watershed area (m), determined by area-weighted Soil and Water Assessment Tool sub-basin precipitation totals. Light to dark bands represent January to December of each year of observation. Total precipitation measured by the rain gauge network is higher than radar totals in all four watersheds. The year 2007 stands out as an extreme drought year across the study area. The numbers of National Climate Data Center gauges used in area weighting are as follows: Mountain Creek = 1, Little River = 1, Neuse River near Clayton = 16, and Neuse River near Fort Barnwell = 29

Table III. Precipitation differences over study watersheds

	MC	LR	NC	FB
All sub-basins				
Daily	6.6*	13.8*	46.1*	64.6*
Monthly	7.9*	16.4*	55.0*	67.6*
Annual	7.7*	18.1*	42.1*	46.8*
Watershed average				
Daily	3.4*	2.6*	6.4*	5.8*
Monthly	4.1*	3.0*	6.9*	4.4*
Annual	4.0 (0.004)	3.5 (0.009)	4.4 (0.002)	2.3 (0.049)

Gauge *versus* radar paired *t*-statistics (with *p*-value in parentheses).
MC, Mountain Creek; LR, Little River; NC, Neuse River near Clayton; FB, Neuse River near Fort Barnwell.
$^*p < 0.001$.

because radar totals are greater when mean daily totals exceed 25–50 mm (Figure 3). The much higher frequency of light events, for which gauge precipitation totals are generally greater, leads to overall positive gauge bias when comparing these datasets. Biases presented for these datasets are evaluated relative to each other.

Model calibration and parameterization differences

Although it is clear that gauge precipitation totals were generally higher than radar totals (Figure 2), with the exception of a conditional negative bias in gauge precipitation during heavy rainfall (Figure 3), these differences do not translate directly to differences in total predicted streamflow volume. This results because the calibration process results in different parameterization for the two precipitation datasets (Figure 4), which in turn leads to important differences in water budget partitioning. The most sensitive parameter in these SWAT simulations was CN_2 (Soil Conservation Service (SCS)-modified curve number), which partitions rainfall into run-off and infiltration (Hawkins, 2009). Each HRU is assigned an empirically based parameter given the soil, land use, and slope, but these parameters can be scaled by a factor ranging from −1 to 1 as part of SWAT calibration. As shown in Figure 4, the calibration for radar data led to a positive skew for CN_2 (acting to increase overland flow), whereas a negative skew was observed in the rain gauge calibration (acting to increase infiltration). No other calibration parameters showed such clearly systematic differences between the two datasets, although the minimal groundwater depth for return flow and soil evaporation compensation factor both skewed slightly higher in radar calibrations than in gauge calibrations (Figure 4). Given that land use, soils, and topography were identical in simulations using the two precipitation datasets, it is noteworthy that CN_2 and other watershed characteristics systematically varied between gauge and radar calibrations. This was due to both the overall magnitude differences and intensity-conditional biases in the precipitation datasets, which showed greater radar precipitation during heavy events and greater gauge precipitation during light events. It is noteworthy that all three differing parameters relate to infiltration and subsurface storage. It is evident the calibration process and the model's parameter flexibility allow these factors to compensate for errors in precipitation during large and small storms, adjusting overall water budgets to match simulated streamflows to observed data.

Range widths systematically differ between precipitation data types (Figure 4). Range width is influenced by the number of iterations required to meet the stopping rule during calibration, because narrowing of the possible parameter range occurs prior to each new iteration. 'Winning' iterations for each watershed and data type are shown in Table IV. The differences resulted in wider parameter ranges for the radar calibrations in MC and LR across all parameters, whereas ranges were wider for gauge calibrations in NC and FB.

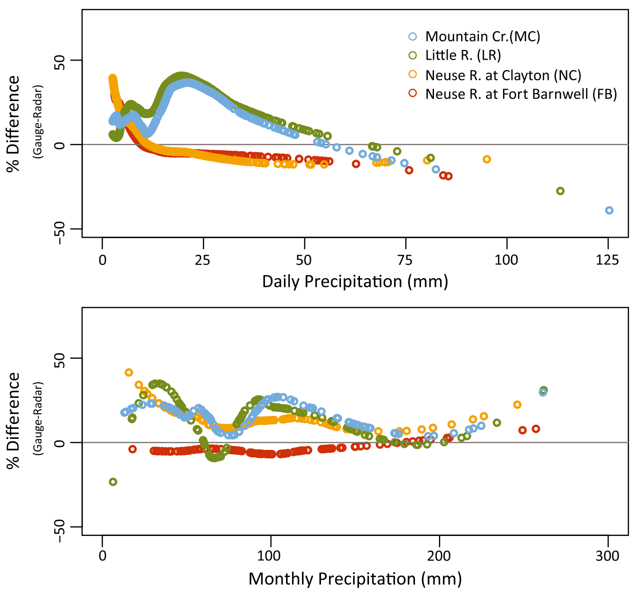

Figure 3. Differences between gauge and radar precipitation totals as a function of daily and monthly totals. Per cent difference was calculated for each of the four study watersheds as (gauge − radar) * (mean of gauge and radar)$^{-1}$ * 100 for each day and month (for each watershed, $n = 3287$ for daily precipitation and $n = 108$ for monthly). The points shown on these graphs represent locally weighted regression through all observed differences for each watershed and show the highest density of observations with light daily rainfall (<25 mm/day) and medium monthly rainfall (~100 mm/month). These results show that gauge precipitation is generally higher than radar precipitation. This figure additionally shows that the relationship for daily precipitation actually reverses for all four watersheds when there is relatively heavy rainfall within a given day. That is, on days when the mean of the precipitation totals exceeds 25–50 mm, radar records higher totals. At monthly and coarser timesteps (e.g. Figure 2), the much greater frequency of light rainfall events leads to consistently greater gauge precipitation totals, compared with radar

Streamflow differences

Precipitation is the only model input variable that differs between the simulation scenarios compared in this study, prior to calibration. Although we acknowledge that there are many sources of uncertainty in watershed modelling, this comparative framework was designed to isolate the effects of using radar *versus* gauge precipitation values as model forcing data, both in terms of parameterization and in streamflow simulation accuracy. A visual comparison of gauge-simulated and radar-simulated streamflows indicates that SWAT simulations are roughly equal in agreement with observed flows, with flashiness exaggerated in the radar-based calibrations and underestimated in the gauge-based calibrations (Figure 4).

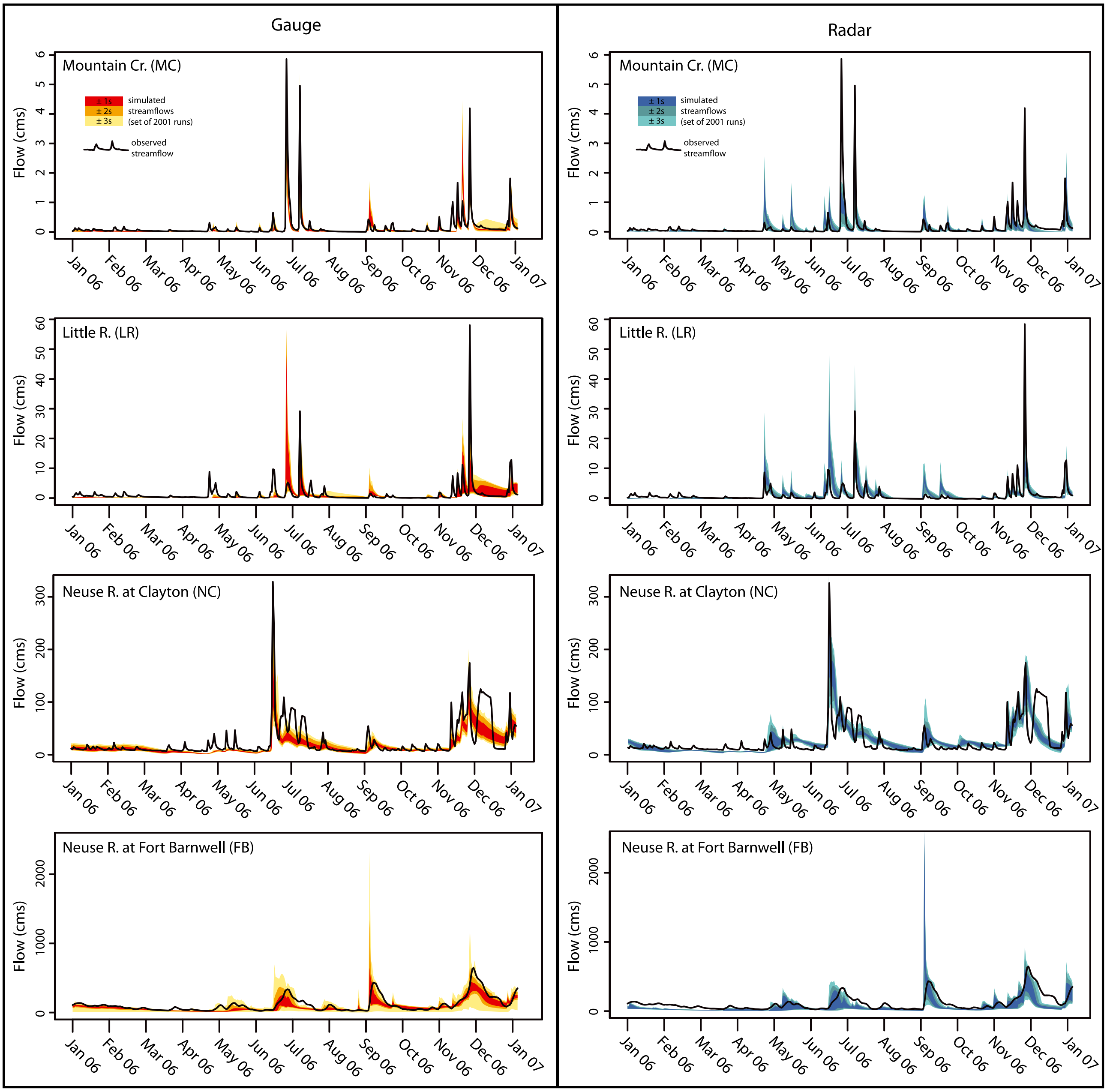

Figure 4. Example of simulated *versus* observed flows. One representative year (2006) of simulated and observed streamflows is shown for comparison of precipitation data types. Calibrated flows are shown with bands of graded colour, representing the first (darkest) to third (lightest) standard deviations of the simulation sets

Table IV. Number of calibration iterations required

	MC	LR	NC	FB
Gauge	3	4	3	4
Radar	2	2	5	5

2001 simulations per iteration, stopping rule ≤ 5% improvement in composite likelihood in the next iteration.
MC, Mountain Creek; LR, Little River; NC, Neuse River near Clayton; FB, Neuse River near Fort Barnwell.

This is a direct reflection of differences in calibrated parameters, as discussed in the previous section. Although basic water budget volumes are realistic for systems in this region using both datasets, radar simulations demonstrate more overland flow and lower baseflows than seen with gauge simulations.

Agreement between SWAT-simulated flows and USGS-observed flows was quantified using the CL index. CL scores for daily streamflow simulations are presented

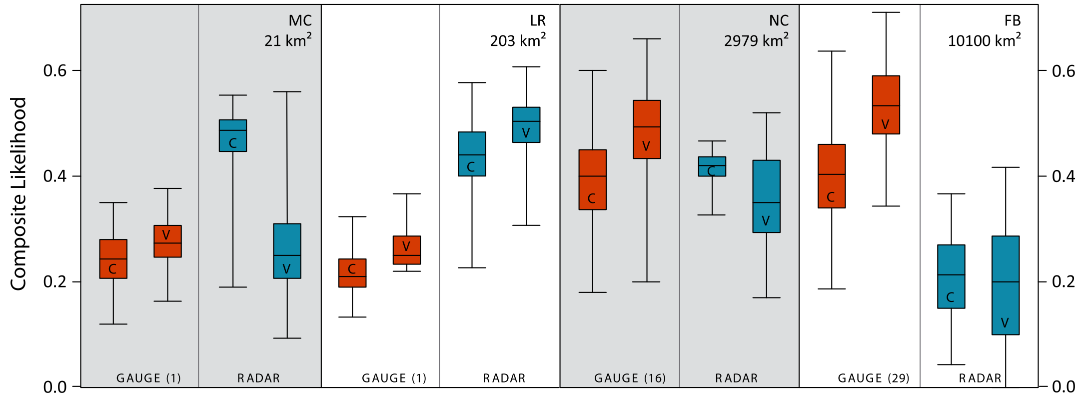

Figure 5. Simulated *versus* observed fit scores for streamflow simulated with gauge and radar data in the four study watersheds. Each box-and-whisker plot represents the distribution of composite likelihood scores for 2001 simulations. The left box in each pair presents data from the calibration period (C; 2002–2007), whereas the right box presents the validation period (V; 2008–2010). Possible CL scores range from 0 to 1, with higher values indicating a better fit to observed data. This multi-objective function is explained in Section 2. Radar-generated streamflow simulations produced better fits overall for the two smaller watersheds, Mountain Creek (MC) and Little River (LR); no clear trend emerged in the third-largest watershed, Neuse River near Clayton (NC); and gauge data produced more accurate streamflow simulations in the largest watershed, the Neuse River near Fort Barnwell (FB)

for the best simulation set (Figure 5). Results show that in the two smaller watersheds (MC and LR), the median and top fit scores for the 2001 simulations in the winning pass are better with radar precipitation data than with gauge data. In NC, CL scores are similar between the two precipitation data types, and in the largest watershed, FB, simulations using gauge data outperform simulations using radar. CL scores were calculated for each watershed and precipitation data type at daily, weekly, monthly, quarterly, and annual timesteps. The mean CL score of the simulation set ($n = 2001$) at each timestep was calculated for each precipitation type (Figure 6). Simulation differences between gauge and radar precipitation data were most apparent at a daily timestep and decreased as timesteps became longer, as a result of averaged positive and negative errors.

Simulated low flows (fifth percentile), median flows, and high flows (95th percentile) were compared with observed flows for each data type and watershed (Figure 7). SWAT simulations using gauge precipitation

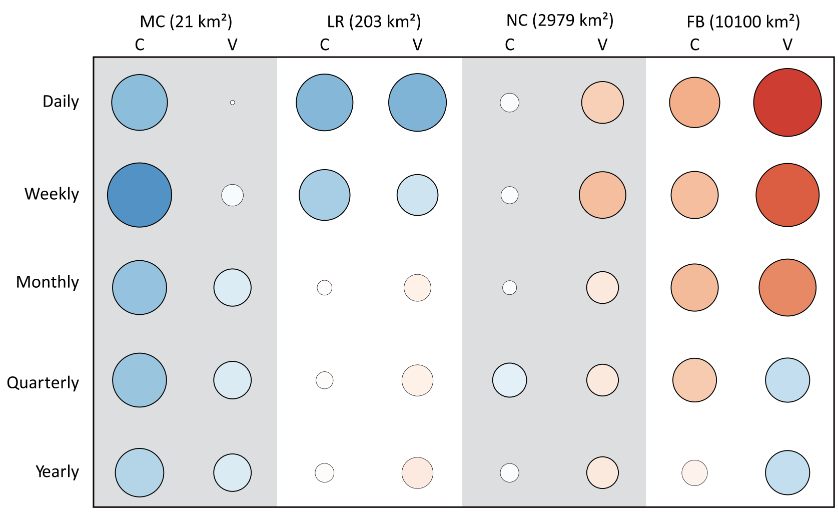

Figure 6. Differences between mean composite likelihood (CL) scores in streamflow simulation sets across timesteps. For each timestep (daily, weekly, quarterly, monthly, and yearly), the average CL score was computed for each calibration (C) and validation (V) simulation set, for each watershed. Means are calculated from 2001 runs in each simulation set. The size and colour of the circle are a function of the magnitude of difference between model fits using gauge *versus* radar data. Red circles indicate better model performance using gauge data, whereas blue circles indicate better model performance using radar data. Larger differences are associated with darker shades and larger circles. The figure shows that radar data generally produce better streamflow in smaller watersheds that contain no rain gauges within the boundary [Mountain Creek (MC) and Little River (LR)], whereas gauge data tend to produce better streamflows in the larger watersheds containing many rain gauges [Neuse River near Clayton (NC) and Neuse River near Fort Barnwell (FB)]. Gauge locations are shown in Figure 1. Also evident from this figure is the decreasing difference between simulation accuracy using the two precipitation datasets as timestep is increased

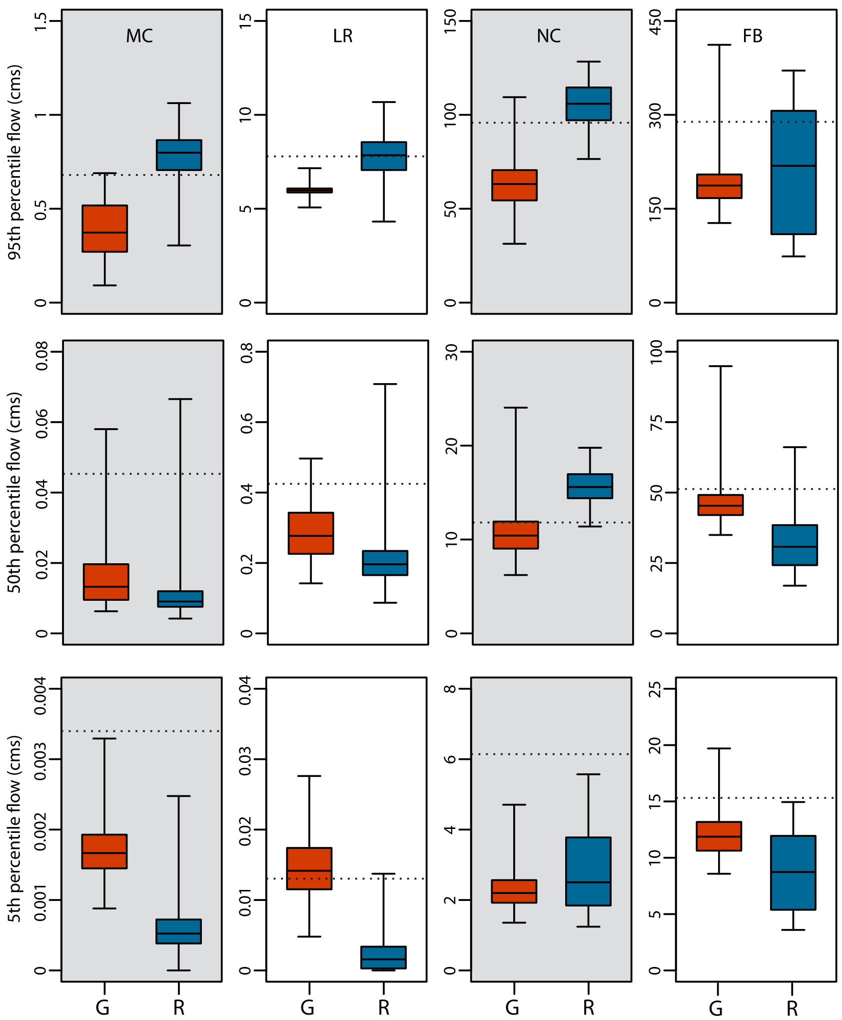

Figure 7. Low, medium, and high flows simulated with gauge (G) and radar (R) precipitation data. Low (fifth percentile), medium (50th percentile), and high (95th percentile) flow magnitudes are calculated from each of 2001 simulations during the validation period (2008–2010). The dotted line represents the observed flows. MC, Mountain Creek; LR, Little River; NC, Neuse River near Clayton; FB, Neuse River near Fort Barnwell

data tended to underestimate low, median, and high flows. Radar simulations showed even greater underestimation of low and median flows but demonstrated much greater accuracy than gauge data at simulating high flows. These flow results relate to the aforementioned biases in the precipitation data, where gauges generally measured greater precipitation during light events (corresponding to low and median flow simulations), whereas radar tended to report greater precipitation during heavier events (corresponding to high-flow simulations). Additionally, model parameterization differences resulting from the basic precipitation differences contribute to these differences in calibrated streamflows (Figure 8). By accepting the assumption that observed streamflows are accurate, it appears that radar data in fact underestimate light precipitation, whereas gauge data underestimate heavy precipitation (Figure 7).

DISCUSSION

The objective of this research was to compare gauge and radar precipitation data at multiple spatial and temporal scales of watershed model operation. Our results suggest a conditional bias in that radar results are superior to rain gauge data during large events, whereas rain gauge data generate more accurate streamflow response to smaller events.

Implications for watershed modelling practice

Our results suggest that increasing precipitation gauge coverage improves simulation accuracy. Neither the MC nor the LR catchment contained any rain gauges, and only one gauge (located outside both watersheds) was used by SWAT to represent the entire watershed. Using a single, external rain gauge not only raises questions about

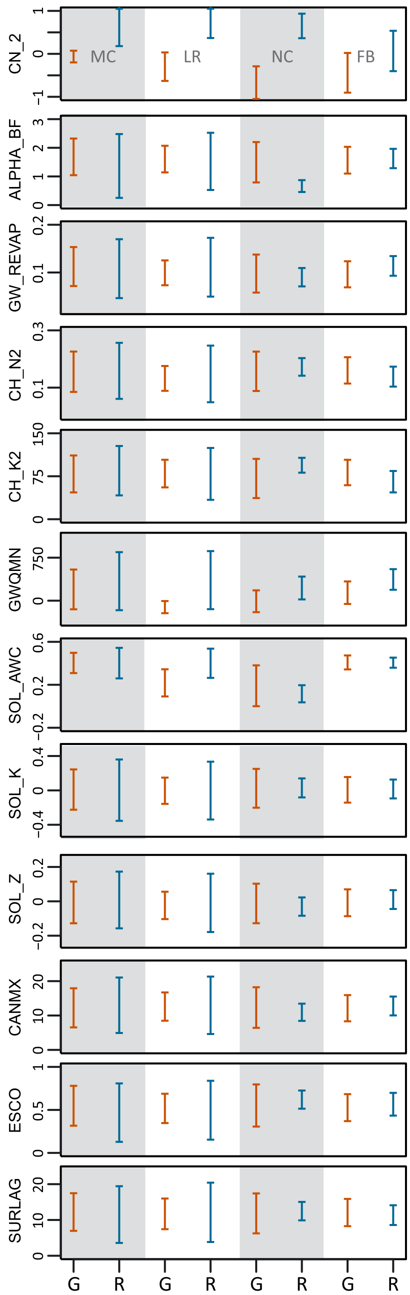

Figure 8. Calibrated Soil and Water Assessment Tool parameter ranges for gauge and radar streamflow simulations. Abbreviations are explained in Table I. *Y*-axes span the initial ranges for each parameter, which were narrowed during the calibration process (explained in detail in Section 2). Curve number adjustments skewed positively in the radar calibrations and they skewed negatively in the gauge calibrations, whereas in reality, there should be no difference in watershed characteristics. This phenomenon is an important reminder that sensitive calibration parameters may act mostly to accommodate error in input data sources (in this case, precipitation) when calibrating simulations to match observed streamflow data

extrapolation from the location of the gauge to the study watershed but also precludes incorporation of any information on spatial variability. It is clear from gauge CL scores across the four watersheds that the average fit improves with increasing watershed area (Figure 6), which is at least partially due to increasing the number of rain gauges inside the watershed boundary. Given that the average performance of radar data decreases with increasing watershed size, it may be inferred that gauge data are indeed more accurate than radar data, but it is advisable to use radar data for watersheds containing few or no rain gauges inside their boundaries. Given the known systematic and random uncertainties of radar data (Habib *et al.*, 2008), it is not surprising that good rain gauge coverage would perform better in watershed modelling. It should also be noted that there is better coverage of Hydrometeorological Automated Data System stations near the smaller watersheds, which may indicate differences in gauge-conditioned radar accuracy among the watersheds.

Other studies have shown a similar dependence on gauge coverage when comparing radar-driven and gauge-driven streamflow simulations. In separate studies in Oklahoma and Maryland, radar precipitation data were shown to produce far more accurate streamflow simulations than rain gauge data, in watersheds that did not contain a gauge within the watershed boundary (Looper *et al.*, 2012; Sexton *et al.*, 2010). However, studies with multiple gauges inside or very near the watershed boundaries showed little difference between gauge-driven and radar-driven simulations (Kalin and Hantush, 2006; Golden *et al.*, 2010). Further research that explores spatial and temporal thresholds of necessary gauge coverage is needed, below which modellers should turn to radar precipitation.

Temporally aggregating streamflow simulations improves simulation accuracy by integrating errors (Madsen, 2000; Knox and Anagnostou, 2009). Studies comparing streamflow simulations over monthly or greater timesteps have not shown consistent superiority of one data type over another. Where present, differences in streamflow accuracy tend to be small, especially when gauge-conditioned radar data are used, and when more than one gauge is included in the comparison (e.g. Borga *et al.*, 2002; Neary *et al.*, 2004; Jayakrishnan *et al.*, 2005). Our results corroborate these findings in that monthly, quarterly, and yearly simulation accuracy showed little difference between gauge and radar data, whereas daily and weekly simulation accuracy was more sensitive to precipitation data type (Figure 7).

It is evident that the model calibration process largely acts to correct error or bias in precipitation data via parameter adjustment, creating differences in basic water budget partitioning. This is especially disconcerting in comparative studies such as this, when there is no physical logic for such differences (Starks and Moriasi, 2009; Strauch *et al.*, 2012). For example, CN_2 systematically varied in the calibrations across all four watersheds in this study. Because watershed characteristics were identical within a given watershed, in all simulations using the two precipitation datasets, CN_2 theoretically should not change. However, our gauge calibrations consistently favoured decreases to default CN_2 values, whereas radar calibrations consistently favoured increases (Figure 8). Given that the model and all other inputs were held constant, these CN_2 differences resulted from overall magnitude and conditional biases in the precipitation datasets, which showed greater radar precipitation during heavy events and greater gauge precipitation during light events. Studies evaluating parameter uncertainty within a single model, but in different regions of the world, have shown very different sensitivities. This suggests that parameter uncertainty significantly depends on the catchments studied and their data attributes, even when model structure is held constant (Seibert, 1997; Uhlenbrook *et al*., 1999; Merz *et al*., 2009). As used here and in other recent studies, calibration approaches that rely on likelihood distributions, as opposed to single best-fit parameter optimizations, can help to avoid skewing parameter values unrealistically to fit observed streamflow data (Beven, 2006; Strauch *et al*., 2012; Price *et al*., 2012).

Implications regarding precipitation data accuracy

Although the primary objective of this study was to compare precipitation data types for use in watershed models, our results also suggest interesting trends in precipitation estimates. Gauge precipitation totals were greater for light rainfall events (<25–50 mm/day), whereas radar precipitation totals were greater for heavier rainfall events (>25–50 mm/day). Because of the much more frequent occurrence of light rainfall events, total precipitation is greater in gauge records than radar records for the 2002–2010 study period (Figure 2). Accordingly, rain gauge-derived simulations of high streamflows underestimate high flows recorded at USGS gauging stations, compared with radar-derived high-streamflow simulations. This difference has been observed in other parts of the USA (e.g. Moon *et al*., 2004) and agrees with the observed instrument error of rain gauges during storms, due to rapid bucket tipping and catching errors (Molini *et al*., 2005; Lanza and Stagi, 2008; Westcott *et al*., 2008). Molini *et al*. (2005) and Lanza and Stagi (2008) reported a conditionally negative bias of gauge data during heavy events, and radar data have been previously shown to underestimate light precipitation (Kalinga and Gan, 2006).

Further research needs

Our results continue the efforts of others to understand the complex relationships between model structure, precipitation data resolution and accuracy, and importance of spatial and temporal scaling considerations in watershed modelling (e.g. Kirchner, 2006; Buytaert *et al*., 2008; Santhi *et al*., 2008; Sivapalan, 2009; Golden *et al*., 2010; Moriasi and Starks, 2010; Thampi *et al*., 2010, and many others). Despite the breadth of ongoing work to resolve these issues, there are still many important questions to answer and methodologies to explore:

1. *Model structure* – Perhaps most imperatively, the modelling community needs a better understanding of the role of watershed model structure in utilizing information on spatial distributions of rainfall. In semidistributed and lumped models, the spatial aggregation of processes may eliminate all information provided by high-resolution data, rendering its increased processing demands superfluous. The relationship between model structure and input data resolution needs to be systematically explored (Kirchner, 2006; Buytaert *et al*., 2008).
2. *User-defined spatial discretization* – Model users typically have some control over the level of spatial discretization. For example, in SWAT (a semidistributed model), the user can choose the number of discrete sub-basins and HRUs in the simulated watershed. This choice is based on modelling goals but is especially important in the context of this discussion, as precipitation spatial variability can only be accommodated to the level of initial discretization. As modellers make choices to more finely discretize the watershed (for more accurate representation of watershed processes), the result may be much longer runtimes, which can limit calibration options. The modelling community would benefit from a better understanding of how much accuracy is gained from varied levels of discretization across all model structures (e.g. Gong *et al*., 2010).
3. *Storm events* – Because various landscape attributes are important to the partitioning of overland flow *versus* infiltration and recharge, inaccurate spatial distribution of rainfall may lead to erroneous streamflow simulations, especially for individual storm events (Cranston and Black, 2006; Gourley *et al*., 2012; Looper and Vieux, 2012; Mohamoud and Prieto, 2012). The fact that the majority of sediment and associated contaminants are moved during event flows underscores the importance of investigating rainfall data accuracy during different storm types and sizes (Inamdar *et al*., 2006; Das *et al*., 2008).
4. *Precipitation ensembles and multimodel frameworks* – It is highly likely that there are distinctions in accuracy of precipitation data (from all sources) when different types of storm structures and precipitation types are considered. A promising methodological advancement is the use of ensembles of precipitation types, either to bound uncertainty (Amengual *et al*., 2008; Strauch *et al*., 2012) or to weight individual estimates by the accuracy associated with the specific precipitation type. Classification of precipitation types and event scales, as devised by

atmospheric science experts, could serve as a framework for conditional calibration schemes (Gamble and Meentemeyer, 1997; Ashley and Ashley, 2008; Dunkerley, 2008). Model ensembles and model averaging approaches could find the likelihood of simulated streamflows under varied model structures and discretization schemes (Thielen and Schaake, 2008; Sahoo *et al*., 2008; Golden *et al*., 2012). In the larger scheme of water quantity, quality, and habitat suitability modelling, uncertainty could be reduced and options enhanced through further development of integrated modelling frameworks, providing users with options among model combinations (Babendreier and Castleton, 2005; Moore and Tindall, 2005; Athanasiadis *et al*., 2009; Matott *et al*., 2009; Johnston *et al*., 2011; Laniak *et al*., 2013).

CONCLUSIONS

Understanding relationships between input data resolution and streamflow simulation accuracy at varied spatial and temporal scales is critical for efficient and effective watershed modelling. In this study, we compared the accuracy of SWAT streamflow simulations across four spatial and five temporal scales, using two types of publicly available precipitation data: NCDC gauge data and MPE radar data. Comparisons were performed by separately using both precipitation datasets to simulate streamflow within each watershed and evaluating both sets of results with USGS-observed streamflow. All other model inputs were held constant. Using USGS-observed streamflows, we calibrated each model scenario using a multi-objective function and a likelihood-weighted approach. Radar data produced more accurate simulated daily streamflows in the 21- and 203-km^2 watersheds (MC and LR, respectively, one rain gauge each); gauge and radar precipitation produced equally good simulations in the 2979-km^2 watershed (NC, 16 gauges); and gauge data produced more accurate simulated streamflows in the 10 100-km^2 watershed (FB, 29 gauges). There was a conditional bias in which gauge data tended to underestimate high precipitation totals and, by extension, high streamflows; otherwise, gauge precipitation estimates tended to be higher than radar totals. As timestep was increased from daily to annual, discrepancies between gauge-driven and radar-driven streamflow simulations decreased. Calibrations using the two data sources resulted in different parameterizations, despite the fact that all watershed characteristics were the same in the simulation scenarios. This highlights concerns that parameterization may be overly influential in streamflow simulations, even beyond equifinality issues.

The results of this study and others discussed here suggest that there is at present no universally superior precipitation product for watershed modelling. The choice of whether to use radar, gauge, or other precipitation data should be based on the spatial and temporal scales of interest, the availability of high-quality gauge data within or near the watershed, the difficulty of processing and manipulating gridded precipitation data (as opposed to the more straightforward point-based time series of gauge data), and the availability of computational resources for use of higher-resolution products. There are also highly important considerations of model structure. Here, we used a semidistributed model, run at a relatively fine level of spatial discretization to use each radar pixel; it may well be the case, however, that lumped models do not benefit from high-resolution precipitation data, whereas fully distributed models may show even more pronounced differences than we observed with SWAT. The impacts of precipitation data choice extend beyond streamflow quantity modelling because any uncertainty or bias introduced in the water quantity will propagate through subsequent modelling and analyses of sediment, nutrients, contaminants, and habitat suitability (Chaplot *et al*., 2005; Wei *et al*., 2009).

Further research is needed to understand how the modelling community may optimize the use of various precipitation datasets in different model structures and spatial discretization scenarios and to develop regional guidance on minimum rain gauge density, below which the use of radar-derived products would be advised. Data resolution is always increasing, but this is generally associated with larger file sizes and more spatially explicit modelling, which can substantially increase model runtime. Minimizing runtime can enable researchers to explore automated calibration schemes that require many thousands of model runs, unachievable if the spatial input is too complex. Furthermore, the use of finer-resolution precipitation data does not necessarily improve run-off simulations, conditional upon model structure, watershed size, and the density of rain gauge coverage. It is to our advantage to identify thresholds of maximized gain to use high-resolution data more precisely, and a key element of this will be to identify opportunities for conditional use and/or averaging of multiple precipitation data sources and watershed model ensembles.

ACKNOWLEDGEMENTS

This research would not have been possible without collaboration and input from Ellen Cooter, Robin Dennis, and John Johnston. Heather Golden, Rhett Jackson, Roy Sidle, and two anonymous reviewers provided generous and insightful feedback on drafts of this manuscript. This paper has been reviewed in accordance with the US

Environmental Protection Agency's peer and administrative review policies and has been approved for publication. Approval does not signify that that the contents necessarily reflect the views and policies of the agency, nor does the mention of trade names or commercial products constitute endorsement or recommendation for use.

REFERENCES

Abbaspour K. 2009. SWAT-CUP2: SWAT Calibration and Uncertainty Programs – A User Manual. Department of Systems Analysis, Integrated Assessment and Modeling (SIAM), Dübendorf, Switzerland: 95 pp. http://www.eawag.ch/forschung/siam/software/swat/index.

Amengual A, Romero R, Alonso S. 2008. Hydrometeorological ensemble simulations of flood events over a small basin of Majorca Island, Spain. *Quarterly Journal of the Royal Meteorological Society* **134**: 1221–1242.

Andréassian V, Perrin C, Michel C, Usart-Sanchez I, Lavabre J. 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *Journal of Hydrology* **336**: 206–223.

Arnaud P, Bouvier C, Cisneros L, Dominguez R. 2002. Influence of rainfall spatial variability on flood prediction. *Journal of Hydrology* **260**: 216–230.

Ashley S, Ashley W. 2008. The storm morphology of deadly flooding events in the United States. *International Journal of Climatology* **28**: 493–503.

Athanasiadis I, Rizzoli A, Janssen S, Andersen E, Villa F. 2009. Ontology for seamless integration of agricultural data and models. *Proceedings of Third International Conference on Metadata and Semantics Research*, Springer-Verlag.

Attorre F, Alfo M, De Sanctis M, Francesconi F, Bruno F. 2007. Comparison of interpolation methods for mapping climatic and bioclimatic variables at regional scale. *International Journal of Climatology* **27**: 1825–1843.

Babendreier J, Castleton K. 2005. Investigating uncertainty and sensitivity in integrated, multimedia environmental models: tools for FRAMES-3MRA. *Environmental Modelling & Software* **20**: 1043–1055.

Bell V, Moore R. 2000. The sensitivity of catchment runoff models to rainfall data at different spatial scales. *Hydrology and Earth System Sciences* **4**: 653–667.

Beven K. 2006. A manifesto for the equifinality thesis. *Journal of Hydrology* **320**: 18–36.

Biggs E, Atkinson P. 2011. A comparison of gauge and radar precipitation data for simulating an extreme hydrological event in the Severn Uplands, UK. *Hydrological Processes* **25**: 795–810.

Bigiarni M. 2010. R Package 'hydroGOF': goodness-of-fit functions for comparison of simulated and observed hydrological time series. http://cran.r-project.org/web/packages/hydroGOF.

Bitew M, Gebremichael M, Ghebremichael L, Bayissa Y. 2012. Evaluation of high-resolution satellite rainfall products through streamflow simulation in a hydrological modeling of a small, mountainous watershed in Ethiopia. *Journal of Hydrometeorology* **13**: 338–350.

Borga M, Tonelli F, Moore R, Andrieu H. 2002. Long-term assessment of bias adjustment in radar rainfall estimation. *Water Resources Research* **38**: 1226, doi:10.1029/2001WR000555.

Buytaert W, Reusser D, Krause S, Renaud J. 2008. Why can't we do better than Topmodel? *Hydrological Processes* **22**: 4175–4179.

Chaplot V, Saleh A, Jaynes D. 2005. Effect of the accuracy of spatial rainfall information on the modeling of water, sediment, and NO_3-N loads at the watershed level. *Journal of Hydrology* **312**: 223–234.

Cranston M, Black A. 2006. Flood warning and the use of weather radar in Scotland: a study of flood events in the Ruchill water catchment. *Meteorological Applications* **13**: 43–52.

Das T, Bardossy A, Zehe E, He Y. 2008. Comparison of conceptual model performance using different representations of spatial variability. *Journal of Hydrology* **356**: 106–118.

Dunkerley D. 2008. Rain event properties in nature and in rainfall simulation experiments: a comparative review with recommendations for increasingly systematic study and reporting. *Hydrological Processes* **22**: 4415–4435.

Gamble D, Meentemeyer V. 1997. A synoptic climatology of extreme unseasonable floods in the southeastern United States, 1950–1990. *Physical Geography* **18**: 496–524.

Golden H, Knightes C, Cooter E, Dennis R, Gilliam R, Foley K. 2010. Linking air quality and watershed models for environmental assessments: analysis of the effects of model-specific precipitation estimates on calculated water flux. *Environmental Modelling & Software* **25**: 1722–1737.

Golden H, Knightes C, Conrads P, Davis G, Feaster T, Journey C, Benedict S, Brigham M, Bradley P. 2012. Characterizing mercury concentrations and fluxes in a Coastal Plain watershed: insights from dynamic modeling and data. *Journal of Geophysical Research* **117**: G01006.

Gong Y, Shen Z, Liu R, Wang X, Chen T. 2010. Effect of watershed subdivision on SWAT modeling with consideration of parameter uncertainty. *Journal of Hydrologic Engineering* **15**: 1070–1074.

Gourley J, Erlingis J, Hong Y, Wells E. 2012. Evaluation of tools used for monitoring and forecasting flash floods in the United States. *Weather and Forecasting* **27**: 158–173.

Habib E, Krajewski W, Kruger A. 2001. Sampling errors of tipping-bucket rain gauge measurements. *Journal of Hydrologic Engineering* **6**: 159–166.

Habib E, Aduvala A, Mesehle E. 2008. Analysis of radar-rainfall error characteristics and implications for streamflow simulation uncertainty. *Hydrological Sciences Journal* **53**: 56–587.

Hawkins R. 2009. Curve Number Hydrology: State of the Practice. ASCE: Reston, VA; 107.

Inamdar S, O'Leary N, Mitchell M, Riley J. 2006. The impact of storm events on solute exports from a glaciated forested watershed in western New York, USA. *Hydrological Processes* **20**: 3423–3439.

Jayakrishnan R, Srinivasan R, Santhi C, Arnold J. 2005. Advances in the application of the SWAT model for water resources management. *Hydrological Processes* **19**: 749–762.

Johnston J, McGarvey D, Barber M, Laniak G, Babendreier J, Parmar R, Wolfe K, Kraemer S, Cyterski M, Knightes C, Rashleigh B, Suarez L, Ambrose R. 2011. An integrated modeling framework for performing environmental assessments: application to ecosystem services in the Albemarle-Pamlico basins (NC and VA, USA). *Ecological Modelling* **222**: 2471–2484.

Jordan P, Seed A, Austin G. 2000. Sampling errors in radar estimates of rainfall. *Journal of Geophysical Research – Atmospheres* **105**: 2247–2257.

Kalin L, Hantush M. 2006. Hydrologic modeling of an eastern Pennsylvania watershed with NEXRAD and rain gauge data. *Journal of Hydrologic Engineering* **11**: 555–569.

Kalinga O, Gan T. 2006. Semi-distributed modelling of basin hydrology with radar and gauged precipitation. *Hydrological Processes* **20**: 3725–3746.

Kirchner J. 2006. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research* **42**: W03S04.

Klemes V. 1986. Operational testing of hydrological simulation-models. *Hydrological Sciences Journal* **31**: 13–24.

Knox R, Anagnostou E. 2009. Scale interactions in radar rainfall estimation uncertainty. *Journal of Hydrologic Engineering* **14**: 944–953.

Krause P. 2005. Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* **5**: 89–97.

Laniak G, Olchin G, Goodall J, Voinov A, Hill M, Glynn P, Whelan G, Gellar, G, Quinn N, Blind M, Peckham S, Reaney S, Gaber N, Kennedy R, Hughes A. 2013. Integrated environmental modeling: a vision and roadmap for the future. *Environmental Modelling and Software* **39**: 3–23.

Lanza L, Stagi L. 2008. Certified accuracy of rainfall data as a standard requirement in scientific investigations. *Advances in Geosciences* **16**: 43–48.

Leigh D. 2008. Late Quaternary climates and river channels of the Atlantic Coastal Plain, Southeastern USA. *Geomorphology* **101**: 90–108.

Looper J, Vieux B. 2012. An assessment of distributed flash flood forecasting accuracy using radar and rain gauge input for a physics-based distributed hydrologic model. *Journal of Hydrology* **412**: 114–132.

Looper J, Vieux B, Moreno M. 2012. Assessing the impacts of precipitation bias on distributed hydrologic model calibration and prediction accuracy. *Journal of Hydrology* **418**: 110–122.

Madsen H. 2000. Automatic calibration of a conceptual rainfall–runoff model using multiple objectives. *Journal of Hydrology* **235**: 76–288.

Matott L, Babendreier J, Purucker S. 2009. Evaluating uncertainty in integrated environmental models: a review of concepts and tools. *Water Resources Research* **45**: W06421.

Merz R, Parajka J, Blöschl G. 2009. Scale effects in conceptual hydrological modeling. *Water Resources Research* **45**: W09405.

Mills H, Brakenridge G, Jacobson R, Newell W, Pavich M, Pomeroy J. 1987. Appalachian mountains and plateaus. In Geomorphic Systems of North America, Graf WL (ed.) Geological Society of America: Tempe, AZ; 5–50.

Mohamoud Y, Prieto L. 2012. Effect of temporal and spatial rainfall resolution on HSPF predictive performance and parameter estimation. *Journal of Hydrologic Engineering* **17**: 377–388.

Molini A, Lanza L, La Barbera P. 2005. The impact of tipping-bucket raingauge measurement errors on design rainfall for urban-scale applications. *Hydrological Processes* **19**: 1073–1088.

Moon J, Srinivasan R, Jacobs J. 2004. Stream flow estimation using spatially distributed rainfall in the Trinity River Basin, Texas. *Transactions of ASAE* **47**:1445–1451.

Moore R, Tindall C. 2005. An overview of the open modelling interface and environment (the OpenMI). *Environmental Science and Policy* **8**: 279–286.

Moreau E, Testud J, Le Bouar E. 2009. Rainfall spatial variability observed by X-band weather radar and its implication for the accuracy of rainfall estimates. *Advances in Water Resources* **32**: 1011–1019.

Moriasi D, Starks P. 2010. Effects of the resolution of soil dataset and precipitation dataset on SWAT2005 streamflow calibration parameters and simulation accuracy. *Journal of Soil and Water Conservation* **65**: 63–78.

Moriasi D, Arnold J, Van Liew M, Bingner R, Harmel R, Veith T. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE* **50**: 885–900.

Moulin L, Gaume E, Obled C. 2009. Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations. *Hydrology and Earth System Sciences* **13**: 99–114.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models, Part 1: a discussion of principles. *Journal of Hydrology* **10**: 282–290.

NASS. 2011. 2010 North Carolina Cropland Data Layer. http://datagateway.nrcs.usda.gov/.

NCAR Earth Observing Laboratory. 2011. GCIP/EOP surface: precipitation NCEP/EMC 4KM gridded data (GRIB) Stage IV data. http://data.eol.ucar.edu/codiac/dss/id=21.093.

NCDC. 2011. Monthly climate station summaries. 1981–2010, www.ncdc.noaa.gov.

Neary V, Habib E, Fleming M. 2004. Hydrologic modeling with NEXRAD precipitation in middle Tennessee. *Journal of Hydrologic Engineering* **9**: 339–349.

Neitsch S, Arnold J, Kiniry J, Williams J. 2011. Soil and Water Assessment Tool theoretical documentation version 2009. Texas Water Resources Institute Technical Report 406, Texas A&M University System, College Station, TX.

Ngongondo C, Xu C, Gottschalk L, Alemaw B. 2011. Evaluation of spatial and temporal characteristics of rainfall in Malawi: a case of data scarce region. *Theoretical and Applied Climatology* **106**: 79–93.

NWS. 2012. HADS Hydrometeorological Automated Data System. http://www.nws.noaa.gov/oh/hads/.

O'Connell P, Todini E. 1996. Modelling of rainfall, flow and mass transport in hydrological systems: an overview. *Journal of Hydrology* **175**: 3–16.

Palanisamy B. 2006. Evaluation of SWAT model – subdaily runoff prediction in Texas watersheds. MS Thesis, Texas A&M University, College Station, TX.

Price K, Purucker S, Kraemer S. 2011. Multi-scale comparison of Stage IV NEXRAD (MPE) and gauge precipitation data for watershed modeling. *Proceedings of Georgia Water Resources Conference*, Athens, GA, April 11-13, 2011.

Price K, Purucker S, Kraemer S, Babendreier J. 2012. Tradeoffs among calibration targets for watershed modeling. *Water Resources Research* **48**: W10542, doi:10.1029/2012WR012005WR012005.

R Development Core Team. 2011. R: a language and environment for statistical computing. R Foundation for Statistical Computing, www.R-project.org.

Rinehart R. 2004. Radar for Meteorologists, 4th edn. Rinehart Publications: Columbia, MO; 482.

Ryan J, Ulrich J. 2012. R Package 'xts': eXtensible Time Series. http://cran.r-project.org/web/packages/xts/xts.pdf.

Sahoo A, Dirmeyer P, Houser P, Kafatos M. 2008. A study of land surface processes using land surface models over the Little River Experimental Watershed, Georgia. *Journal of Geophysical Research – Atmospheres* **113**: D20.

Santhi C, Kannan N, Arnold J, Di Luzio M. 2008. Spatial calibration and temporal validation of flow for regional scale hydrologic modeling. *Journal of the American Water Resources Association* **44**: 829–846.

Schuurmans J, Bierkens M. 2007. Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model. *Hydrology and Earth System Sciences* **11**: 677–693.

Seibert J. 1997. Estimation of parameter uncertainty in the HBV model. *Nordic Hydrology* **28**: 247–262.

Setegn S, Srinivasan R, Daraghi B. 2008. Hydrological modelling in the Lake Tana Basin, Ethiopia using SWAT model. *The Open Hydrology Journal* **2**: 49–62.

Sexton A, Sadeghi A, Zhang X, Srinivasan R, Shirmohammadi A. 2010. Using NEXRAD and rain gauge precipitation data for hydrologic calibration of SWAT in a Northeastern watershed. *Transactions of the ASABE* **53**: 1501–1510.

Sieck L, Burges S, Steiner M. 2007. Challenges in obtaining reliable measurements of point rainfall. *Water Resources Research* **43**: W01420.

Sivapalan M. 2009. The secret to 'doing better hydrological science': change the question!. *Hydrological Processes* **23**: 1391–1396.

Soil Survey Staff. 2011. U.S. General Soil Map (STATSGO2). http://soildatamart.nrcs.usda.gov.

Starks P, Moriasi D. 2009. Spatial resolution effect of precipitation data on SWAT calibration and performance: implications for CEAP. *Transactions of the ASABE* **52**: 1171–1180.

State Climate Office of North Carolina. 2011. NC Climate Retrieval and Observations Network of the Southeast (CRONOS) Database. http://www.nc-climate.ncsu.edu/cronos.

Stisen S, Sandholdt I. 2010. Evaluation of remote-sensing-based rainfall products through predictive capability in hydrologic runoff modeling. *Hydrological Processes* **24**: 879–891.

Strangeways I. 2007. Precipitation: Theory, Measurement and Distribution, Cambridge University Press, Cambridge, UK; 302.

Strauch M, Bernhofer C, Koide S, Volk M, Lorz C, Makeschin F. 2012. Using precipitation data ensemble for uncertainty analysis in SWAT streamflow simulation. *Journal of Hydrology* **414**: 413–424.

Thampi S, Raneesh K, Surya T. 2010. Influence of scale on SWAT model calibration for streamflow in a river basin in the humid tropics. *Water Resources Management* **24**: 4567–4578.

Thielen J, Schaake J. 2008. Special issue: HEPEX Workshop: Stresa, Italy, June 2007. *Atmospheric Science Letters* **9**: 27–28.

Tobin K, Bennett M. 2009. Using SWAT to model streamflow in two river basins with ground and satellite precipitation data. *Journal of the American Water Resources Association* **45**: 253–271.

Uhlenbrook S, Seibert J, Leibundgut C, Rodhe A. 1999. Prediction uncertainty of conceptual rainfall–runoff models caused by problems in identifying model parameters and structure. *Hydrological Sciences Journal* **45**: 779–797.

USGS. 2011. NLCD 2006 Land Cover. www.mrlc.gov/nlcd2006_downloads.php.

Villarini G, Serinaldi F, Krajewski W. 2008a. Modeling radar-rainfall estimation uncertainties using parametric and non-parametric approaches. *Advances in Water Resources* **31**: 1674–1686.

Villarini G, Mandapaka P, Krajewski W, Moore R. 2008b. Rainfall and sampling uncertainties: a rain gauge perspective. *Journal of Geophysical Research* **113**: D11102.

Walker H, Coleman M. 1987. Atlantic and Gulf Coastal Province. In Geomorphic Systems of North America, Graf WL (ed). Geological Society of America: Tempe, AZ; 51–110.

Wang X, Melesse A, Yang W. 2006. Influences of potential evapotranspiration estimation methods on SWAT's hydrologic simulation in a northwestern Minnesota watershed. *Transactions of the ASABE* **49**: 1755–1771.

Wei W, Chen L, Fu B. 2009. Effects of rainfall change on water erosion processes in terrestrial ecosystems: a review. *Progress in Physical Geography* **33**: 307–318.

Westcott N, Knapp H, Hilberg S. 2008. Comparison of gage and multi-sensor precipitation estimates over a range of spatial and temporal scales in the Midwestern United States. *Journal of Hydrology* **35**:1–12.

Winchell M, Srinivasan R, Di Luzio M, Arnold J. 2007. ArcSWAT Inferface for SWAT 2005: User's Guide. Grassland Soil and Water Research Laboratory: Temple, TX.

Yilmaz K, Hogue T, Hsu K, Sorooshian S, Gupta H, Wagener T. 2005. Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologic forecasting. *Journal of Hydrometeorology* **6**: 497–517.

Yoo C, Kim K, Choi J, Ha E. 2010. Importance of no-rain measurements on the comparison of radar and rain gauge rain rate. *Hydrological Processes* **24**: 924–933.

Young C, Brunsell N. 2008. Evaluating NEXRAD estimates for the Missouri River basin: analysis using daily raingauge data. *Journal of Hydrologic Engineering* **13**: 549–553.