Received: 18 October 2017 | Revised: 11 March 2018 | Accepted: 22 August 2018

DOI: 10.1111/jfr3.12492

CIWEM Chartered Institution of Water and Environmental Management WILEY Journal of Flood Risk Management

ORIGINAL ARTICLE

A comparison of SAC-SMA and Adaptive Neuro-fuzzy Inference System for real-time flood forecasting in small urban catchments

Babak K. Roodsari[1] | David G. Chandler[1] | Christa Kelleher[1,2] | Charles N. Kroll[3]

[1]Department of Civil and Environmental Engineering, Syracuse University, Syracuse, New York

[2]Department of Earth Sciences, Syracuse University, Syracuse, New York

[3]Department of Environmental Resources Engineering, State University of New York College of Environmental Science and Forestry, Syracuse, New York

Correspondence
Babak K. Roodsari, Department of Civil and Environmental Engineering, Syracuse University, Syracuse, New York.
Email: bkasaeer@syr.edu

Funding information
National Science Foundation, Grant/Award Number: SBE-1417542SBE-1444755

Growing urbanisation and imperviousness have augmented magnitudes of peak flows, resulting in flooding especially during extreme events. Flood forecast of extreme events can rely on real-time ensemble flood forecasting systems. Such systems often use predictions from physical models and precipitation ensembles to predict downstream urban flood hydrographs. However, these methods are seldom used in small catchments, where flood predictions may assist emergency management. We explore the relative utility of two models, the Sacramento Model (SAC-SMA) and an adaptive neuro-fuzzy inference system (ANFIS) for ensemble flood prediction for nine small urban catchments located near New York City. The models were used to reforecast streamflow for Hurricane Irene (160 mm) and a 35 mm storm across lead times from 3 to 24 hr. Differences in performance between models were small for short (3 hr) lead times, and were similar for the 35 mm storm. Reforecasts of hurricane Irene at 24-hr lead times show strong performance for SAC-SMA, but a decline in performance for ANFIS. Model performance did not vary systematically with either catchment size or imperviousness. Our results suggest that model selection is especially important when reforecasting large rain events with longer lead times in small urban catchments.

KEYWORDS
ANFIS, ensemble flood forecasting, hurricane Irene, hydrologic modelling, imperviousness, runoff peak flows, SAC-SMA, urbanisation

1 | INTRODUCTION

As a consequence of rapid urbanisation and increased surface imperviousness, many urban watersheds worldwide are threatened by greater frequency and depth of flooding (Du et al., 2012; Liu, De Smedt, Hoffmann, & Pfister, 2005; Nirupama & Simonovic, 2007; Qaiser, Yuan, & Lopez, 2012; Suriya & Mudgal, 2012; Wang & Yang, 2013; Zope, Eldho, & Jothiprakash, 2015). Urban floods endanger human lives, damage property, and initiate a cascade of environmental and health impacts (Jha, Bloch, & Lamond, 2012; World Bank, 2013). To mitigate this damage, emergency management authorities may rely on real-time flood forecast systems to provide sufficient lead time for evacuation and asset protection in urban watersheds during extreme rainfall events. However, developing these systems is complicated by spatio-temporal variations and uncertainty in rainfall distributions alongside complex rainfall–runoff relationships. As such, flood forecasting remains one of the most challenging tasks in hydrology (Chang, Chiang, & Chang, 2007).

Flood forecasting often requires two key decisions: (a) how to treat and represent precipitation forecasts and uncertainty in these forecasts, and (b) appropriate model selection for best streamflow response simulation. Recently, application of ensemble streamflow prediction (ESP) systems for real-time flood forecasting have gained popularity as an approach to represent the inherent uncertainty associated with rainfall predictions for flood

forecasts (Cloke & Pappenberger, 2009; Cloke, Wetterhall, He, Freer, & Pappenberger, 2013; Day, 1985; Emerton et al., 2016; Gouweleeuw, Thielen Del Pozo, Franchello, De Roo, & Buizza, 2005). ESP systems use an ensemble of rainfall forecast scenarios to represent a range of future weather assumptions. The rainfall forecast ensembles are used to generate a series of future flood hydrographs, also called spaghetti hydrographs (Emerton et al., 2016). This procedure is performed on a real-time basis using a flood forecast model that is continuously calibrated up to the current time from historical weather and streamflow observations. A great benefit of an ESP approach over the traditional single-run deterministic modelling approach is the generation of an ensemble of predicted flood hydrographs that facilitate uncertainty analyses (Day, 1985). A spaghetti hydrograph can be used to inform emergency managers about possible future flooding scenarios and can guide strategies for evacuation and rescue.

Historically, local ESP systems have utilised a wide variety of physical models and numerical weather prediction (NWP) data sources for simulating rainfall–runoff processes during extreme events such as flash floods or hurricanes (Liechti, Zappa, Fundel, & Germann, 2013). Overall, previous studies on ensemble flood prediction agree that this approach is a great benefit over the traditional deterministic modelling approach for relatively large river basins (>100 km^2) (Amengual, Homar, & Jaume, 2015; Hally et al., 2015; Saleh, Ramaswamy, Georgas, Blumberg, & Pullen, 2016). The effectiveness of the ensemble flood forecasting approach for relatively small urban catchments (<100 km^2) remains unclear due to the limited research at this scale. Furthermore, nearly all ESP systems simulate streamflow using physically based hydrologic models. While physically based hydrologic models are useful tools for flood forecasting, they also have large input requirements (topography, land use, meteorological data, and soil characteristics), many degrees of freedom, and are therefore subject to the known problems of over-parameterisation and equifinality (Beven, 2006).

Artificial intelligence (AI) models are an alternative approach to physical models (Napolitano, See, Calvo, Savi, & Heppenstall, 2010). AI models have reduced degrees of freedom due to fewer parameters, and therefore may be at lower risk for equifinality. However, AI models apply mathematical equations analysing concurrent input and output time series rather than attempt to simulate physical processes (Nourani, Hosseini Baghanam, Adamowski, & Kisi, 2014; Solomatine & Ostfeld, 2008). Applications of AI models to flood forecasting predictions have demonstrated their use for forecasting flood stage and discharge for large river basins (>204 km^2) across Asia for forecast lead times ranging from 1 hr to 1 day (Campolo, Soldati, & Andreussi, 2003; Chang et al., 2007; Nayak, Sudheer, Rangan, & Ramasastri, 2005; Nguyen & Chua, 2012).

Despite advances in technique and availability of model input data, there has been very limited focus regarding flood forecasting in small urban catchments. Streamflow in small urban catchments is flashier than in large catchments as smaller catchment areas are more responsive to storm events (Epstein, Kelso, & Baker, 2016; Walsh et al., 2005). This is in part due to the closer match between the storm scale, catchment time of concentration (Nicótina, Alessi Celegon, Rinaldo, & Marani, 2008; Wilson, Valdes, & Rodriguez-Iturbe, 1979), and catchment storage capacity (Sapriza-Azuri et al., 2015). Given the short response times of small urban catchments to extreme precipitation events, accurate estimation of peak discharge and stage with a sufficient lead time is perhaps even more critical than in large river basins. Recently, a study of Japanese watershed response to the Talas Typhoon by Yu, Nakakita, Kim, and Yamaguchi (2016) demonstrated that predictions in smaller watersheds are often more difficult than in large river basins. Prediction of extreme events in small catchments is complicated by uncertainties associated with often coarse spatial resolution precipitation and land cover data sets compared to the size of the catchment.

Across the literature, the many examples applying either adaptive neuro-fuzzy inference system (ANFIS) or physically based models to flood forecasting analysis have all focused on relatively large watersheds. We aim to provide additional benchmarking for the application of ensemble-based flood forecasting approaches for small watersheds, one of the very first studies to do so. We compare the performance of a purely data-driven model (ANFIS) alongside a conceptual model (SAC-SMA) for ensemble flood prediction at several small- to medium-sized suburban catchments (17–150 km^2) near New York City (NYC). To compare the skill of these two models for real-time flood forecasting during large- and small-scale storm events, we apply both models to reforecast the flood hydrograph of a disastrous historical extreme event, hurricane Irene, and another smaller storm that occurred a few weeks after hurricane Irene. This analysis is used to test the hypothesis that ANFIS performs as accurately as SAC-SMA for ensemble flood forecasting for forecast lead times of three to 24 hr in relatively small peri-urban catchments, which are newly developed urban catchments in close proximity of large growing cities. Further knowledge about the performance of conceptual and data-driven models for flood forecasting in small urban catchments can be valuable for local urban flood emergency management in peri-urban catchments.

2 | METHODS

2.1 | Study site description

In August 2011, hurricane Irene caused several deaths and severe property damage to the eastern coast of the United

States. Property damage was approximated at about $1.5 billion in New York (http://www.fema.gov/ar/disaster/4020) and $1 billion in New Jersey (Saleh et al., 2016). During hurricane Irene, a total of between 150 and 250 mm of accumulated precipitation occurred in a period of less than 2 days. Flood levels at most streams in proximity of NYC exceeded the mean historical annual gaged peak flow. Emergency management agencies evacuated about 1 million people from the flood-prone regions to limit loss of life (Watson, Collenburg, & Reiser, 2013). Nevertheless, several deaths occurred in flooded areas during the event. In this study, we simulated the flood hydrographs for nine peri-urban catchments near NYC that were severely impacted by hurricane Irene (Figure 1 and Table 1). Study catchment drainage areas range from small (17 km^2) to medium (150 km^2) sizes. These nine catchments are slightly to moderately developed, with impervious area ranges from 12 to 25%. The soil in the study area consists of approximately 40% silt, 10% clay, and 50% sand and has a high runoff potential (Falcone, 2011).

Subcatchment drainage areas in Table 1 were calculated using the U.S. Geological Survey (USGS) StreamStats auto delineation tool (http://water.usgs.gov/osw/streamstats/). The mean historical annual gaged peak flow and the gaged peak flow during hurricane Irene were obtained from the corresponding USGS gages.

2.2 | Model descriptions and input data sets

2.2.1 | Model input data and simulation periods

Meteorological data including hourly precipitation and temperature data were obtained from Phase 2 of the North American Land Data Assimilation System (using the HydroDesktop version 1.4 software (Ames et al., 2012). We focus on model applications for two different events: (a) hurricane Irene (160 mm, approximately 36 hr), and (b) a 35 mm storm on September 23–25, 2011. We focus primarily on these two storm events based on an extensive survey of all storms between 2004 and 2011 contained within the Global Ensemble Forecast System Reforecast (GEFS/R), one of the most reliable sources for ensemble precipitation data for reforecasting extreme events. Through this survey, we found that the U.S. National Weather Service (NWS) performed poorly for predicting the temporal distribution and the total depth for most of the extreme events in that period for the study sites. Although there is still a great uncertainty in precipitation ensembles for the two selected storm events, they are two of the best predictions of NWS among other historical extreme events in GEFS/R database. The historical

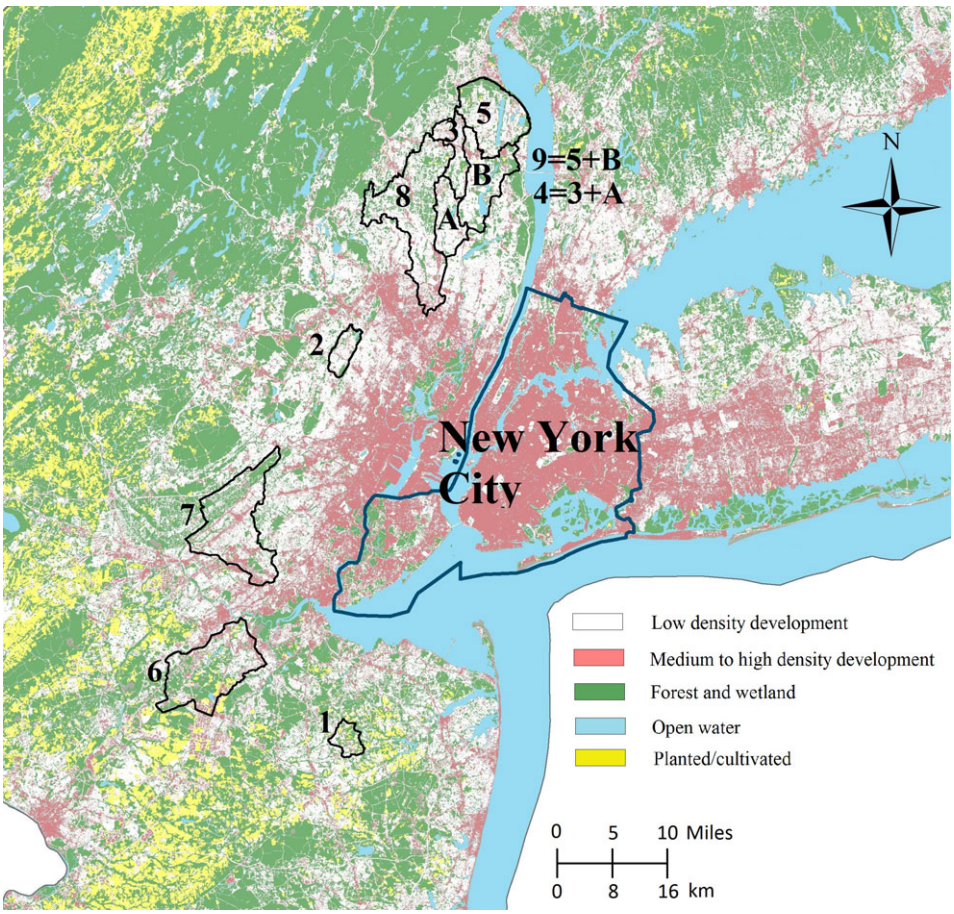

FIGURE 1 Land cover map of the study catchments. Catchment ID numbers are arranged based on drainage area while Catchments 1 and 9 are the smallest and largest study sites, respectively. Table 1 provides detailed information about the study catchments

TABLE 1 Characteristics of the study sites in proximity of NYC

Catchment ID	Subcatchment name	USGS gage number	Area (km^2)	Imperviousness (%)	Mean historical annual gaged peak flow (m^3/s)	Gaged peak flow during Hurricane Irene (m^3/s)
1	Big Brook near Marlboro, NJ	01407290	17	12	23	38
2	Peckman River at Little Falls, NJ	01389550	20	19	33	57
3	Pascack Brook at Park Ridge, NJ	01377370	35	20	66	99
4	Pascack Brook at Westwood, NJ	01377500	77	17	32	131
5	Hackensack River at West Nyack, NY	01376800	80	12	21	49
6	Lawrence Brook at Westons Mills, NJ	01405030	116	20	64	218
7	Bound Brook at Middlesex, NJ	01403900	125	25	75	166
8	Saddle River at Lodi, NJ	01391500	141	17	52	151
9	Hackensack River at Rivervale, NJ	01377000	150	13	32	132

observed streamflow discharge records from October 1, 2004 to October 1, 2014 were obtained from the corresponding USGS gages (Table 1). Observed meteorological and discharge data from October 1, 2004 to August 27, 2011 were used for model calibration for hurricane Irene. Similarly, observed data sets from October 1, 2004 to September 23, 2011 were used for model calibration for the 35 mm storm event that occurred a few weeks after hurricane Irene. The calibrated models were then validated for the following 3 years to ensure robustness. Finally, the GEFS/R precipitation data inputs and the observed temperature and discharge records for the events August 27–29, 2011 and September 23–25, 2011 were used to force the calibrated/validated models for ensemble stream flow prediction.

2.2.2 | Model selection

To find the most suitable model for this study, we tested several models, including Hydrological Simulation Program Fortran (HSPF), the Storm Water Management Model (SWMM), and SAC-SMA (unpublished results). Ultimately, we selected SAC-SMA, given its wide use in the literature (Ajami, Gupta, Wagener, & Sorooshian, 2004; Burnash, Ferral, McGuire, & McGuire, 1973; Foehn, Hernandez, Roquier, & Arquiola, 2016) and its use by the NWS for flood forecasting.

SAC-SMA model

SAC-SMA is a conceptual watershed model that distributes humidity within the soil profile to accurately simulate streamflow (Burnash et al., 1973; Foehn et al., 2016). The Hydrology Laboratory of National Oceanic and Atmospheric Administration's NWS selected the SAC-SMA lumped model as a comparison baseline for participating distributed hydrologic models in the distributed model intercomparison project, which aimed to identify the most suitable model for NWS streamflow prediction across the U.S. (Smith et al., 2004). More importantly, the NWS currently uses the lumped form of SAC-SMA for U.S. wide ensemble flood forecasting (Emerton et al., 2016). For these reasons, we chose to employ a lumped version of SAC-SMA in this study.

SAC-SMA was calibrated using the multistep automatic calibration scheme (Hogue, Sorooshian, Gupta, Holz, & Braatz, 2000). We note that, to the best of our knowledge, NWS does not use automatic calibration. We have automated the calibration process for our study sites as manual calibration relies highly on expert knowledge. As the best temporal resolution of available GEFS/R precipitation forecasts is 3 hr, all models were calibrated in three-hourly time steps. We recognise that performance would likely improve if we used a smaller time step; however, we sought to treat this as a real-world exercise, preserving the time step of the input data. In this procedure, all SAC-SMA model parameters were initially calibrated to minimise the root-mean-squared error (RMSE) of log-transformed streamflow observations and predictions. The upper zone parameters were then adjusted using the RMSE of the untransformed streamflow data while the lower zone parameter values remain fixed from the previous calibration. Finally, the lower zone parameters were readjusted using the RMSE of the log-transformed data while upper zone parameter values remained fixed from the previous step. For the validation period associated with hurricane Irene (August 27–29, 2011) and the 35 mm storm (September 23–25, 2011), we used a data-assimilation approach to account for current discharge observations. With this approach, the SAC-SMA model input parameters were allowed to vary between 10% below and above their calibrated values. This approach was shown to slightly improve the accuracy of the flood forecasting model by recalibrating the model based on the real-time discharge observations. This also allowed the model flexibility to capture current conditions.

ANFIS model

ANFIS is a data-driven model framework that combines the human logic of fuzzy inference systems (FISs) with the adaptive capability of training artificial neural networks (ANNs) (Jang & Sun, 1995). FIS is the theory of solving fuzzy processes (Zadeh, 1965) that are controlled by unclear, uncertain, or incomplete information using several if-then statements and numerical methods called membership functions. Membership functions define the degree of truth of

each fuzzy statement using a value of between 0 and 1. While, ANN's training module can be used to create appropriate membership functions and if-then rules to approximate an output data set, the FIS structure is unable to dynamically adjust with the environmental change in data sets. To overcome this shortcoming, the learning capability of ANN was added to ANFIS.

AI models are typically trained using the input variables that have the highest Pearson correlation coefficient with the outputs (Sudheer, Gosain, & Ramasastri, 2002). For hydrologic modelling, AI model input variables typically include the antecedent observed discharge and accumulated precipitation for lead times with the highest Pearson correlation coefficient. In this study, we trained ANFIS using the antecedent observed discharge N hours before present ($Q_{t\text{-}N}$; N = forecast lead time) and the 3- and 6-hr accumulated precipitation. This selection was based on the observation of greatest Pearson correlation coefficient values between the current discharge at each time step (Q_t) and the antecedent precipitation and discharge inputs ($Q_{t\text{-}N}$). An important benefit of the three-parameter ANFIS model compared to the multiparameter SAC-SMA model used in this study is the smaller number of input parameters that decreases the calibration time, the number of uncertainty sources, and the risk of equifinality (Beven, 2006) associated with model calibration. Furthermore, the training process of ANFIS can be automated and does not necessarily require expert knowledge, a key difference when calibrating a conceptual model like SAC-SMA. However, because the ANFIS model is dependent on antecedent observed discharge ($Q_{t\text{-}N}$) and forecast lead time, the model must be calibrated and validated for each lead time.

2.3 | Real-time flood forecasting system

We implemented a real-time ensemble flood forecasting approach (Figure S1, Supporting Information) to reforecast the flood discharge at nine USGS gages (Table 1) located at the outlet of the study sites for hurricane Irene and a storm event that occurred a few weeks after hurricane Irene (September 23–25, 2011). Eleven ensemble members of the GEFS/R precipitation (10 members + 1 control member) with a temporal resolution of 3 hr were used to force the calibrated models to forecast streamflow during these two events. As the available GEFS/R precipitation data are produced only once daily at 00 Universal Time Coordinated, a meteorological and discharge data updating component was added to the system to update the precipitation and streamflow discharge inputs for subdaily forecasts. Figure A2 shows an example of the precipitation updating mechanism. This updating component corrected the initial conditions of the predictor model (for SAC-SMA or ANFIS) for subdaily predictions based on the most recent meteorological and streamflow observations within the forecast system. For SAC-SMA, a data-assimilation technique was used to update model parameters based on discharge observations. In this approach (as noted in Section 2.2.2), SAC-SMA was recalibrated at each update by allowing parameters to vary between 10% below and above the original parameter values to account for uncertainty in these estimates as well as to enable real-time assimilation of observations, leading to improved agreement between modelled and observed discharge. Finally, the performance of the forecast models was assessed using the indices described in Table 2.

3 | RESULTS

3.1 | Calibration/validation

Across all catchments, both three-hourly models performed reasonably well in the calibration (2004–2011) and validation (2011–2014) periods with Nash–Sutcliffe efficiency (NSE) values ranging from 0.72 to 0.87 (Table 3). Relative Bias (*RelBIAS*) values, SAC-SMA calibration data sets, and calibration hydrographs are presented for individual watersheds in Supporting Information/Appendix (Tables A1 and A2, and Figure A3). Values presented in Table 3 represent average performance across the 11 forecast ensemble members for all study sites. Calibration and validation

TABLE 2 Statistical indices used to assess model performance

Index title	Formula	Min	Max	Best
Nash–Sutcliffe efficiency (NSE)	$1-\frac{\sum_{i=1}^{n}\left(Q_i^o-Q_i^f\right)^2}{\sum_{i=1}^{n}\left(Q_i^o-\overline{Q}_i^o\right)^2}$	$-\infty$	1	1
Percent BIAS (PBIAS)	$100\times\frac{\sum_{i=1}^{n}\left(Q_i^f-Q_i^0\right)}{\sum_{i=1}^{n}Q_i^0}$	$-\infty$ $+\infty$	$+\infty$	0
Relative bias (RelBIAS)	$\frac{\sum_{i=1}^{n}\left(\frac{Q_i^f-Q_i^o}{Q_i^o}\right)}{n}$	$-\infty$ $+\infty$	$+\infty$	0
Relative mean squared error (RelMSE)	$\frac{\sum_{i=1}^{n}\left(\frac{Q_i^f-Q_i^o}{Q_i^o}\right)^2}{n}$	0	$+\infty$	0
Average relative absolute difference (ARAD)	$\frac{\sum_{i=1}^{n}\left\lvert\frac{Q_i^f-Q_i^o}{Q_i^o}\right\rvert}{n}$	0	$+\infty$	0

Note. With and Q_i^o are the *ith* forecasted and observed discharge, respectively; and $\overline{Q}_i^o$ is the average of all observed discharge values.

TABLE 3 Averaged performance indices for the nine study sites over the calibration (October 1, 2004–August 27, 2011 for Hurricane Irene and October 1, 2004–September 23, 2011 for the small event) and validation (August 27, 2011–August 27, 2014 for Hurricane Irene and September 23, 2011–September 23, 2014 for the small event) periods

		SAC-SMA		ANFIS							
				Calibration				Validation			
Storm event	Index	Calibration	Validation	3 hr	6 hr	9 hr	24 hr	3 hr	6 hr	9 hr	24 hr
Hurricane Irene (August 27–29, 2011)	NSE	0.83	0.81	0.87	0.84	0.79	0.73	0.86	0.81	0.75	0.72
	PBIAS (%)	6.13	5.19	5.67	6.50	8.23	10.54	4.35	6.56	8.29	11.89
	RelBIAS	0.11	0.08	0.08	0.11	0.13	0.15	0.09	0.11	0.12	0.14
	RelMSE	0.15	0.11	0.19	0.22	0.27	0.34	0.18	0.25	0.37	0.45
	ARAD	0.14	0.13	0.17	0.23	0.29	0.36	0.19	0.25	0.39	0.48
Small event (September 23–25, 2011)	NSE	0.85	0.84	0.86	0.83	0.81	0.76	0.84	0.81	0.77	0.73
	PBIAS (%)	6.21	6.56	4.32	5.76	6.39	8.31	7.12	7.87	8.24	9.16
	RelBIAS	0.09	0.08	0.07	0.09	0.10	0.12	0.08	0.09	0.11	0.12
	RelMSE	0.18	0.19	0.15	0.19	0.27	0.39	0.16	0.22	0.35	0.46
	ARAD	0.12	0.15	0.13	0.26	0.38	0.45	0.15	0.19	0.24	0.36

performance indices for ANFIS decreased with forecast lead time for both events. This was consistent with the observed decrease in the statistical correlation between $Q_{t\text{-lead}}$ time and Q_t. Similarly, *RelBIAS* of the ANFIS model over the calibration period for hurricane Irene increased from 0.08 to 0.15 when forecast lead time increased from 3 to 24 hr. For SAC-SMA, the most (7) and least (1) impervious study sites had the smallest and greatest *RelBIAS* values, respectively. However, we did not find any trends between performance indices (including *RelBIAS*) and either catchment imperviousness or drainage area. For ANFIS, performance indices varied within the sites and with lead time. For example, Sites 3 and 8 for the 3-hr lead time, and Sites 4 and 1 for the 24-hr lead time, had the smallest and greatest *RelBIAS* values for the calibration period, respectively.

3.2 | Performance during extreme events

For simulated real-time flood forecasting, agreement between the observed and simulated hydrographs varied most between models for forecasts of hurricane Irene; this makes sense given this extremely rare event. Forecasts are included both ANFIS and SAC-SMA for hurricane Irene (Figure 2a) and a smaller storm (September 23–25; Figure 2b) for a single watershed (Site 7), as patterns were similar across study sites. Observed and ensemble forecasted flood hydrographs for the smallest and largest study sites (Sites 1 and 9) and minimum, average, and maximum *RelBIAS* among the 11 forecasted ensemble members for individual catchments are presented in Supporting Information (Figure A2, Tables A2 and A3).

ANFIS-simulated real-time forecasted hydrographs for hurricane Irene were best for the smallest lead times, with increasing disagreement as lead times approached 24 hr (Figure 2a). Performance, in terms of average NSE, for ANFIS forecasts of hurricane Irene declined (from 0.85 to 0.4) for increasing forecast lead times (from 3 to 24 hr; Figure 3). ANFIS largely underpredicted the hurricane Irene peak flow for forecast lead times of 24 hr (Figure 2a). Correspondingly, average *RelBIAS* values for ANFIS for 24-hr lead time for hurricane Irene ranged from −0.45 to −1.1 (Table A3). ANFIS performed well when simulating the flood hydrograph for a small storm (September 23–25) for all lead times (Figure 2b; Table A3). While the ANFIS model failed to match the peak when applied to simulate hurricane Irene at the longest lead times (Figure 2a), the ANFIS model performed reasonably well for the smaller storm event, bracketing streamflow observations.

In contrast to ANFIS simulations, SAC-SMA performed reasonably well when simulating event hydrographs for both storms. Average *RelBIAS* values ranged from −0.2 to 0.48 (Table A4), and NSE values ranged from 0.65 to 0.9 (Figure 3). As can be seen in Figure 3, SAC-SMA forecasts for ensemble members tended to bracket observations regardless of lead time. At the longest lead times, SAC-SMA overpredicted peak discharge for both hurricane Irene responses as well as discharge for the smaller event. This overprediction decreased as lead times decreased.

Finally, we sought to test whether catchment size or forecast lead time had a greater impact on model performance. While we observed that NSE was highest for short lead times for both models, we somewhat surprisingly found that for some watersheds, NSE increased as lead times changed from 9 to 24 hr (Figure 3). For example, NSE values for ANFIS for Catchment 5 increased from 0.28 to 0.36 when forecast lead time increased from 9 to 24 hr (Figure 3). Similarly, NSE value for SAC-SMA for Site 3 slightly increased between 9- and 24-hr lead times (Figure 3).

We found performance indices (*RelBIAS*, *RelMSE*, and *ARAD*) across models were insensitive to catchment size and imperviousness, but varied with forecast lead time (Figure 4; Figures A4 and A5). While performance indices for both models varied in a similar narrow range for forecast lead times of 3–9 hr, we found performance diverged between models as lead times approached 24 hr. In particular,

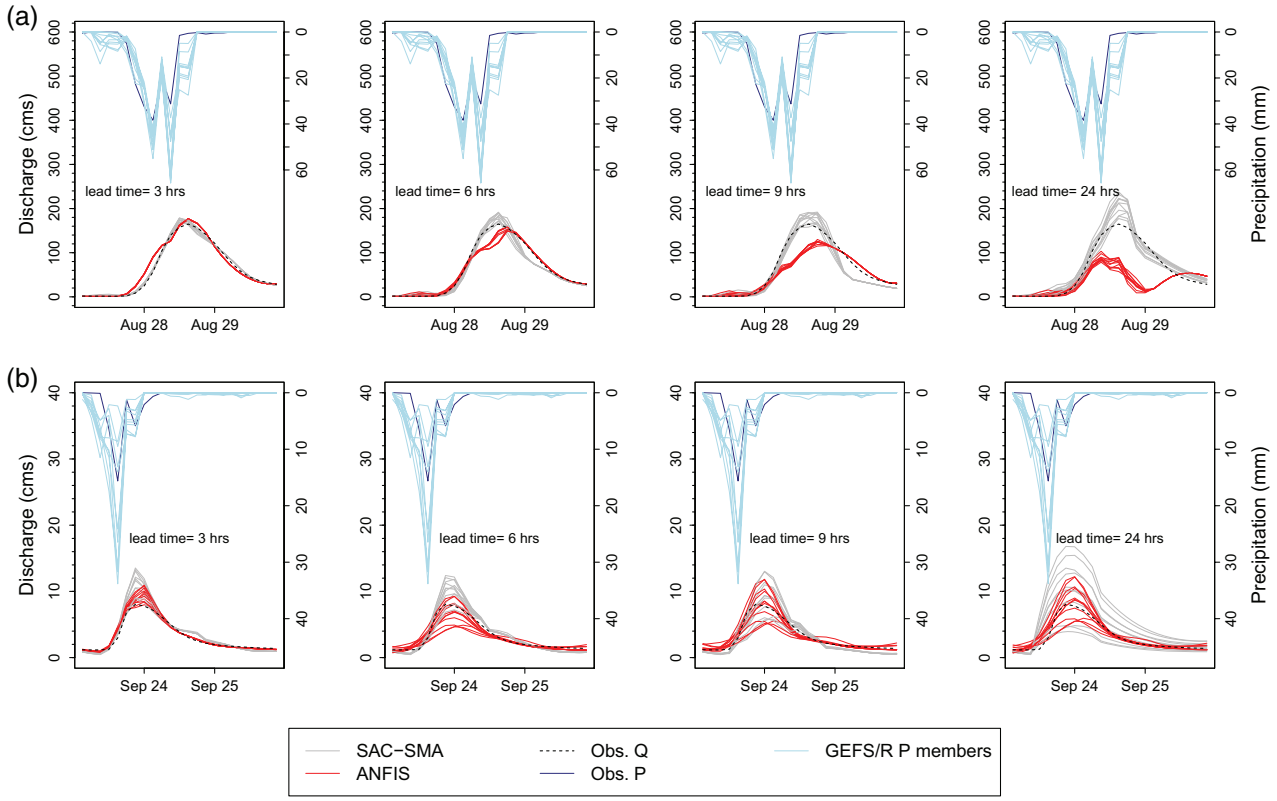

FIGURE 2 Observed and ensemble forecasted flood hydrographs of Site 7 for (a) hurricane Irene and (b) a smaller storm (September 23–25, 2011). Lead times for forecasting increase from left to right

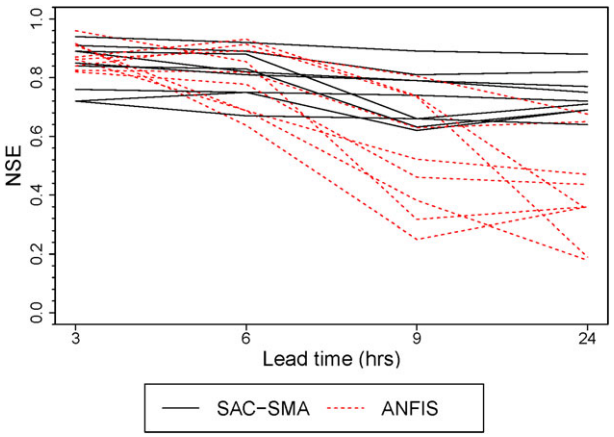

FIGURE 3 Model errors shown as *NSE* coefficient for SAC-SMA and ANFIS across the nine study catchments applied to simulate hurricane Irene flood hydrographs. Numbers on the graph represents study site IDs (Column 1 in Table 1) ordered by increasing drainage area

performance remained high for SAC-SMA, but declined for ANFIS simulations with 24-hr lead times (as compared to 3-, 6-, or 9-hr lead times) regardless of catchment size. The only trend we observed with watershed size was with respect to SAC-SMA forecasts for longer lead times. For lead times of both 9 and 24 hr, performance tended to improve with watershed size (closer to 0) for ARAD; this was not true for other performance indices. We found comparable results for ANFIS and SAC-SMA for short lead times when forecasting the smaller event (Figure A5, Supporting Information). However, for the longest forecast lead times for the smaller event, ANFIS outperformed SAC-SMA, with slightly lower values of performance indices regardless of watershed size.

4 | DISCUSSION

4.1 | Model performance and uncertainty outside of extreme event forecasts

Both SAC-SMA and ANFIS models performed reasonably well during calibration and validation periods with NSE values greater than 0.7. Deterministic flood forecasting applications of the SAC-SMA model in previous studies over a wide range of catchment scales and climate conditions have shown both similar and different performance results compared to our ensemble-based approach findings. For example, Ajami et al. (2004) and Reed, Schaake, and Zhang (2007) used a spatially distributed SAC-SMA model for streamflow forecasting in large U.S. river basins and achieved *NSE* and *RelBias* values similar to our findings (Table 3). Others have found larger biases than we observed (e.g., Khakbaz, Imam, Hsu, & Sorooshian, 2009). Taken together, these studies indicate that our model performance is comparable to other study applications of the SAC-SMA model in deterministic streamflow prediction.

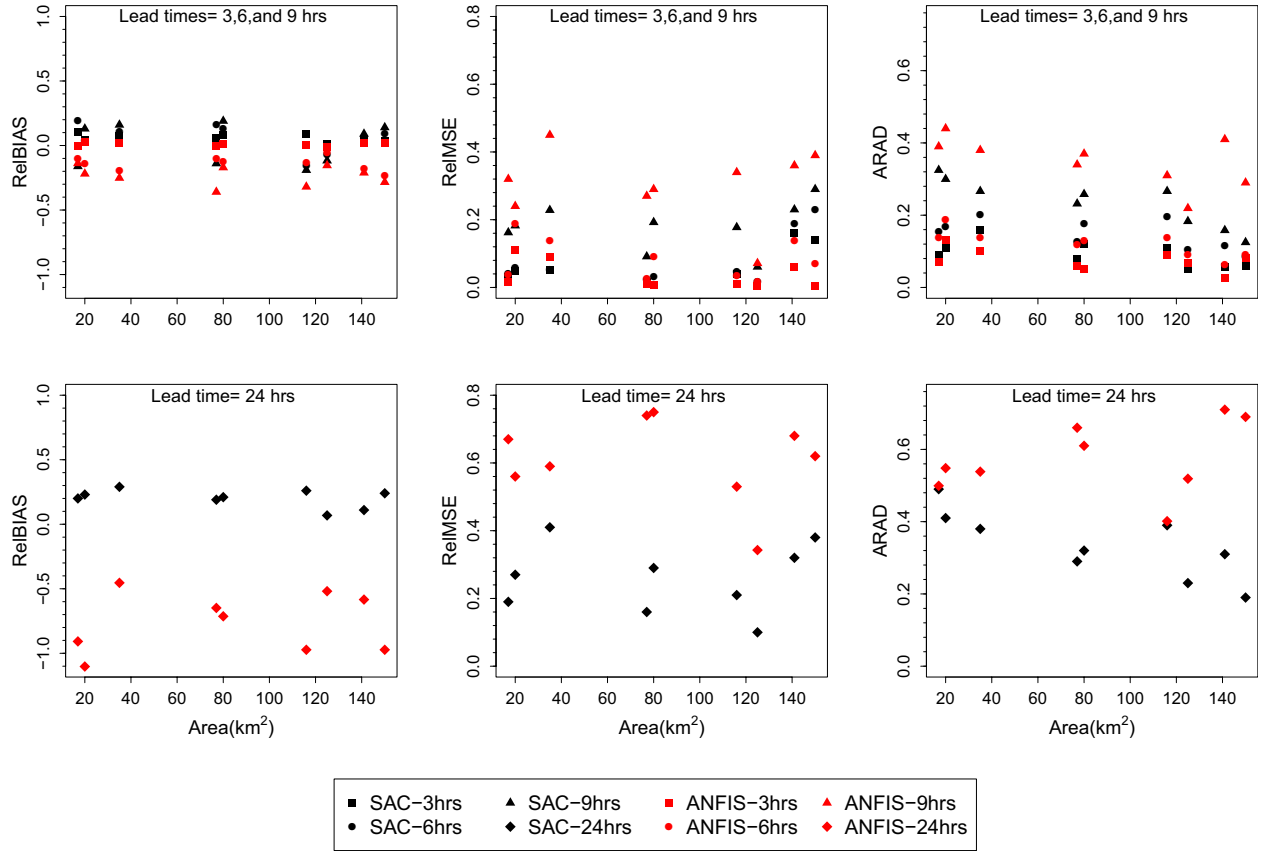

FIGURE 4 Performance indices for SAC-SMA and ANFIS averaged across the 11 ensemble members with varied lead times plotted against catchment drainage area. Results in this figure represent the hurricane Irene flood hydrograph simulation

To enable a real-world simulation of model forecasting, we did not investigate or compare the relative impacts of sources of uncertainty in this study, instead calibrating SAC-SMA following procedures used by the NWS. However, we recognise that different sources of uncertainty with respect to model parameters and input data ultimately shape results with respect to both models. We do note that the greater number of input parameters for SAC-SMA (17 parameters) as compared to ANFIS (three parameters) does increase potential sources of uncertainty and the risk of equifinality (Beven, 2006), initially a motivating factor for comparing these two models. For the ANFIS model, the main sources of uncertainty are intrinsic to the measured precipitation and discharge values used for the model calibration, uncertainty due to the length of calibration period and the presence of events similar to the validation storm event, and the uncertainties of GEFS/R precipitation ensembles for the validation period, which we discuss as further sources of discrepancy between ANFIS and SAC-SMA performance. During the real-time forecasting, we posit that the most important sources of uncertainty in streamflow forecasts for both models are associated with the uncertainties of GEFS/R precipitation ensembles. Within this study, we only focus on two events, as we found in a survey of the GEFS/R database (2004–2011) that there are few extreme events for which NWS relatively accurately predicted the precipitation intensity and total depth at least 24 hr before the event start time. We note that other studies have also found high sensitivity of real-time flood forecasting models to the predicted precipitation inputs (Amengual et al., 2015; Liechti et al., 2013; Marty, Zin, & Obled, 2013; Saleh et al., 2016), hence the need to perform this type of study, and our constraints on the events we simulate. While a shorter time step would likely yield better results, our goal was to perform an analysis as similar to real-time flood forecasting as possible, namely, the three-hourly time step corresponding to GEFS/R input data. This study highlights the need for more resolute precipitation ensembles, as floods were only forecasted well for the smallest watersheds at very close lead times. Whereas we compared only two models within this analysis, we additionally tested other models while developing this study, including the HSPF and the SWMM (results not published). Ultimately, urban hydrologic modelling would strongly benefit from an intercomparison project (Best et al., 2015; Kollet et al., 2017; Smith et al., 2004) to examine the utility of hydrologic models for average and extreme conditions, given these areas are especially challenging to simulate (e.g., Yu et al., 2016).

4.2 | How does model performance vary with lead time?

The presented study evaluates the performance of a lumped conceptual model (SAC-SMA) and an AI model (ANFIS)

through a real-time ensemble flood forecasting approach. Our results suggest that the forecast performance of both models decreases with forecast lead time, which is in agreement with results of previous findings (Campolo et al., 2003; Nayak et al., 2005; Saleh et al., 2016). For short lead times (3 and 6 hr), precipitation input data updates likely resulted in smaller errors and uncertainties with respect to GEFS/R precipitation data inputs. In contrast, forecasts at greater lead times had poorer performance, likely due to the relatively short time of concentration (1–6 hr) in the study catchments. We note that accurate flood forecasting for short lead times can still be valuable for emergency evacuation warning in small urban catchments in contrast to large river basins.

Surprisingly, NSE values of models for some catchments increased slightly as lead times increased to 9 and 24 hr (Figure 3). This may be related to the underlying processes of the updating system or uncertainties of the GEFS/R precipitation inputs for 24-hr lead time due to variability in rainfall predictions. For these long lead times, SAC-SMA generally overestimated peak flow magnitudes as the GEFS/R precipitation data for both hurricane Irene and the smaller precipitation event were slightly greater than the observed precipitation amounts (Figure 2). Note that this overprediction of peak flow magnitude is not necessarily detrimental, as it still correctly reports the major flood condition status in the catchment and may still be useful for emergency management.

4.3 | Comparing ANFIS to SAC-SMA for extreme event forecasts

While forecast performance for ANFIS and SAC-SMA was similar for shorter lead times, performance diverged as lead times increased to 9 and 24 hr (Figure 4). At lead times of 24 hr, SAC-SMA outperformed ANFIS with respect to all indices. ANFIS underestimated peak flow magnitude of hurricane Irene for lead times greater than 3 hr. Thus, we expect ANFIS is more reliable for flood forecasting with short lead times (Figure 4).

An important consideration related to the performance of both SAC-SMA and ANFIS for hurricane Irene is likely the dearth of large storm events or hurricanes in the training period (2004–2011). Due to the learning nature of the ANFIS model, these types of models can only provide accurate predictions if the training period includes storms of magnitude equal to or greater than storms in the validation period. The performance of SAC-SMA could improve if the training period includes a large hurricane due to the improvements in high flow calibration. Unfortunately, continuous streamflow discharge data for the study sites were only available for a limited period (2004–2011) during which no other storms as large as hurricane Irene occurred, and represents a real-world scenario where data in small catchments, including streamflow, may be limited. Performance might improve for future events at these sites as ANFIS would now have hurricane Irene data to train the model. The lack of floods in the training data set is a common shortcoming in hydrologic modelling of extremes; our results demonstrate that we need better methods for simulating extreme events and that starts with exploring to what extent these issues may impact our approaches and outcomes. This also highlights the importance of ongoing streamflow discharge monitoring in small urban catchments, especially for extreme events, for more accurate future flood forecasting.

Poor performance of ANFIS for long lead times was likely also due to weak statistical correlations between the antecedent discharge ($Q_{\text{t-lead time}}$) and the observed discharge at each time step (Q_t) for the relatively short times of concentration in the study catchments (1–6 hr). We inferred that antecedent discharge is not an effective input parameter for ANFIS for real-time flood forecasting in small urban catchments with lead times greater than 3 hr. In this case, we suggest using any other possible meteorological and hydrological input parameters to increase the predictability performance of the data-driven real-time flood forecasting model. In contrast to our finding, previous studies have found good predictability performance of AI models for large river basins with long times of concentration (Campolo et al., 2003; Nayak et al., 2005; Nguyen & Chua, 2012; Rezaeianzadeh, Tabari, Arabi Yazdi, Isik, & Kalin, 2014). As there has been very limited focus on applying data-driven models for real-time flood forecasting in relatively small urban catchments in the previous literature, our study is one of the first to show potential trade-offs in model frameworks for real-time flood forecasting.

We note that forecast performance was similar for ANFIS and SAC-SMA for the smaller storm (Figure A3). In this case, we found ANFIS outperformed SAC-SMA for long lead times. This suggests that both models can be reliable options for real-time flood forecasting in small urban catchments for predicting small storm events, and that ANFIS should have improved performance as more training data from large precipitation events becomes available for model training.

The model performance indices for the nine study catchments with drainage areas ranging from 17 to 150 km^2 and fractional impervious areas ranging from 12 to 25% lead us to conclude that the accuracy of both SAC-SMA and ANFIS models for ensemble flood prediction may not change with catchment size and imperviousness (Figure 4; Figures A4 and A5). We did not find a strong statistical correlation between model performance indices and catchment drainage area or fractional impervious area (Figure 4; Figures A4 and A5). However, the scope of our study has a limited climatic and spatial extent, and we caution that relationships between catchment size and imperviousness may differ for other areas. While we primarily contextualise our results with

drainage area and imperviousness, we recognise that several other physical properties of urban catchments affect flood response, including soil types, vegetation, sewer systems type and location, road network geometry, and catchment shape. In particular, there is growing evidence that urban soil is incredibly heterogeneous and poorly characterised (Herrmann, Shuster, & Garmestani, 2017). Urban sewer and storm water collection systems can also affect the flood response of urban watersheds by subsurface flow routing (Miller et al., 2014; Roodsari & Chandler, 2017). Due to the limited number of catchments in this study, future work would benefit from the application of SAC-SMA and ANFIS models for real-time flood forecasting in a greater number of small suburban catchments with a wide range of fractional impervious areas, drainage patterns, and across climatic regions to assess the sensitivity of model performance indices to different catchment characteristics.

5 | CONCLUSION

In this study, we applied a lumped conceptual model, SAC-SMA, and one of the most widely used data-driven models in hydrologic forecasting, ANFIS, to reforecast streamflow discharge at several small to medium size peri-urban catchments of NYC during hurricane Irene and another storm event. Comparison of various statistical performance indices for SAC-SMA and ANFIS indicated that SAC-SMA can perform reasonably well for flood prediction in relatively small urban catchments (drainage area < 150 km^2) with NSE values mostly greater than 0.75. In contrast, ANFIS largely underpredicted the rising limb and the peak flow of hurricane Irene flood hydrographs, especially for lead times greater than 3 hr, but performed well when forecasting a smaller storm event. We infer that the poor performance of ANFIS for hurricane Irene is likely due to the absence of similarly large storms included in the training period. Poor quality precipitation data remain one of the challenges in real-time flood forecasting, necessitating advances in higher spatial and temporal NWPs to improve forecasting in small urban catchments where emergency notification is needed.

Our work also suggests that the flood forecasting performance of the lumped SAC-SMA and ANFIS models may not depend on the catchment scale and impervious area for relatively small urban catchments. Quantitative performance parameters (*RelBIAS*, *RelMSE*, and *ARAD*) for both models varied in a relatively similar range for the nine study sites with drainage areas ranging from 17 to 150 km^2 and fraction of impervious areas ranging from 12 to 25%. However, we suggest examining these models for real-time flood prediction systems in a greater number of small to medium-sized catchments with a wide range of imperviousness, drainage patterns, and climate to study the model's sensitivity to different characteristics of the catchments and their performance under varying conditions. We posit that model selection can significantly impact the accuracy of flood predictions due to the complexity of land cover and hydrology in these areas. For this study, we initially tested HSPF, SWMM, SAC-SMA, and finally selected SAC-SMA for analysis due to a better accuracy and its wider historical application in practical flood forecasting systems.

Despite better performance of SAC-SMA compared to ANFIS for predicting the flood hydrograph of hurricane Irene in the nine study catchments, the use of AI models shows some promise as an alternative to physical or conceptual models in local urban flood forecasting systems if a long training period with a wide range of storm scales are available for the site. Indeed, we demonstrate for short forecast lead times that performance of ANFIS forecasts was comparable to SAC-SMA forecasts, despite the large increase in degrees of freedom associated with the large number of model parameters associated with SAC-SMA. However, we also emphasise the importance of applying physical or conceptual models for the real-time flood forecasting systems due to uncertain future climatic conditions and potential changing physical characteristics of a watershed. The streamflow hydrograph for the future extreme events may not be accurately predicted by AI models as AI models are learning algorithms that are highly dependent on the past memory. One solution for improving flood forecasting performance may be to apply stochastic hydrology approaches (e.g., Vogel, 2017). Although stochastic approaches are typically applied to deterministic hydrologic simulations (Vogel, 2017) they may be applied for real-time flood forecasting by combining precipitation with stochastically generated precipitation ensembles from the historical data with actual NWPs for the extreme event. Overall, our study demonstrates accurate flood forecasting in small watersheds requires long continuous periods of streamflow discharge monitoring and higher temporal resolution of predicted precipitation inputs. More importantly, increased data density and flood hydrographs of extreme events in small catchments are needed to benchmark and improve the predictability of real-time flood forecasting models.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation under Grant Numbers SBE-1444755 and 1417542. The data used for this research including observed meteorological data, observed discharge, and the GEFS/R precipitation ensemble members for the NYC study catchments are available online in references provided throughout the text.

ORCID

Babak K. Roodsari http://orcid.org/0000-0001-6155-4638

REFERENCES

Ajami, N. K., Gupta, H., Wagener, T., & Sorooshian, S. (2004). Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *Journal of Hydrology*, *298*, 112–135. https://doi.org/10.1016/j.jhydrol.2004.03.033 The Distributed Model Intercomparison Project (DMIP).

Amengual, A., Homar, V., & Jaume, O. (2015). Potential of a probabilistic hydrometeorological forecasting approach for the 28 September 2012 extreme flash flood in Murcia, Spain. *Atmospheric Research*, *166*, 10–23. https://doi.org/10.1016/j.atmosres.2015.06.012

Ames, D. P., Horsburgh, J. S., Cao, Y., Kadlec, J., Whiteaker, T., & Valentine, D. (2012). HydroDesktop: Web services-based software for hydrologic data discovery, download, visualization, and analysis. *Environmental Modelling & Software*, *37*, 146–156. https://doi.org/10.1016/j.envsoft.2012.03.013

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., … Vuichard, N. (2015). The plumbing of land surface models: Benchmarking model performance. *Journal of Hydrometeorology*, *16*, 1425–1442. https://doi.org/10.1175/JHM-D-14-0158.1

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*, 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Burnash, R. J. C., Ferral, R. L., McGuire, R. A., & McGuire, R. A. (1973). *A generalized streamflow simulation system: Conceptual modeling for digital computers*. U.S. Department of Commerce, National Weather Service, and State of California, Department of Water Resources, *1973*, 204.

Campolo, M., Soldati, A., & Andreussi, P. (2003). Artificial neural network approach to flood forecasting in the river Arno. *Hydrological Sciences Journal*, *48*, 381–398. https://doi.org/10.1623/hysj.48.3.381.45286

Chang, F.-J., Chiang, Y.-M., & Chang, L.-C. (2007). Multi-step-ahead neural networks for flood forecasting. *Hydrological Sciences Journal*, *52*, 114–130. https://doi.org/10.1623/hysj.52.1.114

Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. *Journal of Hydrology*, *375*, 613–626. https://doi.org/10.1016/j.jhydrol.2009.06.005

Cloke, H. L., Wetterhall, F., He, Y., Freer, J. E., & Pappenberger, F. (2013). Modelling climate impact on floods with ensemble climate projections. *Quarterly Journal of the Royal Meteorological Society*, *139*, 282–297. https://doi.org/10.1002/qj.1998

Day, G. N. (1985). Extended streamflow forecasting using NWSRFS. *Journal of Water Resources Planning and Management*, *111*, 157–170. https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157)

Du, J., Qian, L., Rui, H., Zuo, T., Zheng, D., Xu, Y., & Xu, C.-Y. (2012). Assessing the effects of urbanization on annual runoff and flood events using an integrated hydrological modeling system for Qinhuai River basin, China. *Journal of Hydrology*, *464–465*, 127–139. https://doi.org/10.1016/j.jhydrol.2012.06.057

Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., … Cloke, H. L. (2016). Continental and global scale flood forecasting systems. *Wiley Interdisciplinary Reviews Water*, *3*, 391–418. https://doi.org/10.1002/wat2.1137

Epstein, D. M., Kelso, J. E., & Baker, M. A. (2016). Beyond the urban stream syndrome: Organic matter budget for diagnostics and restoration of an impaired urban river. *Urban Ecosystem*, *19*, 1041–1061. https://doi.org/10.1007/s11252-016-0557-x

Falcone, J. (2011) *GAGES-II: Geospatial attributes of gages for evaluating streamflow*. Reston, VA: U.S. Geological Survey.

Foehn, A., Hernandez, J.G., Roquier, B., & Arquiola, J.P. (2016). RS MINERVE—User's Manual v2.4.

Gouweleeuw, B., Thielen Del Pozo, J., Franchello, G., De Roo, A., & Buizza, R. (2005). Flood forecasting using medium-range probabilistic weather prediction. *Hydrology and Earth System Sciences*, *9*, 365–380.

Hally, A., Caumont, O., Garrote, L., Richard, E., Weerts, A., Delogu, F., … Clematis, A. (2015). Hydrometeorological multi-model ensemble simulations of the 4 November 2011 flash flood event in Genoa, Italy, in the framework of the DRIHM project. *Natural Hazards and Earth System Sciences*, *15*, 537–555. https://doi.org/10.5194/nhess-15-537-2015

Herrmann, D. L., Shuster, W. D., & Garmestani, A. S. (2017). Vacant urban lot soils and their potential to support ecosystem services. *Plant and Soil*, *413*, 45–57. https://doi.org/10.1007/s11104-016-2874-5

Hogue, T. S., Sorooshian, S., Gupta, H., Holz, A., & Braatz, D. (2000). A multistep automatic calibration scheme for river forecasting models. *Journal of Hydrometeorology*, *1*, 524–542. https://doi.org/10.1175/1525-7541(2000)001<0524:AMACSF>2.0.CO;2

Jang, J.-S. R., & Sun, C.-T. (1995). Neuro-fuzzy modeling and control. *Proceedings of the IEEE*, *83*, 378–406. https://doi.org/10.1109/5.364486

Jha, A. K., Bloch, R., & Lamond, J. (2012). *Cities and flooding: A guide to integrated urban flood risk management for the 21st century*. Washington, DC: The World Bank.

Nguyen, P. K.-T., & Chua, L. H.-C. (2012). The data-driven approach as an operational real-time flood forecasting model. *Hydrological Processes*, *26*, 2878–2893. https://doi.org/10.1002/hyp.8347

Khakbaz, B., Imam, B., Hsu, K., & Sorooshian, S. (2009). From lumped to distributed via semi-distributed: Calibration strategies for semi-distributed hydrologic models. *Journal of Hydrology*, The Distributed Model Intercomparison Project (DMIP)—Phase 2 Experiments in the Oklahoma Region, USA, *418–419*, 61–77. https://doi.org/10.1016/j.jhydrol.2009.02.021

Kollet, S., Sulis, M., Maxwell, R. M., Paniconi, C., Putti, M., Bertoldi, G., … Sudicky, E. (2017). The integrated hydrologic model intercomparison project, IH-MIP2: A second set of benchmark results to diagnose integrated hydrology and feedbacks. *Water Resources Research*, *53*, 867–890. https://doi.org/10.1002/2016WR019191

Liechti, K., Zappa, M., Fundel, F., & Germann, U. (2013). Probabilistic evaluation of ensemble discharge nowcasts in two nested alpine basins prone to flash floods. *Hydrological Processes*, *27*, 5–17. https://doi.org/10.1002/hyp.9458

Liu, Y. B., De Smedt, F., Hoffmann, L., & Pfister, L. (2005). Assessing land use impacts on flood processes in complex terrain by using GIS and modeling approach. *Environmental Modeling and Assessment*, *9*, 227–235.

Marty, R., Zin, I., & Obled, C. (2013). Sensitivity of hydrological ensemble forecasts to different sources and temporal resolutions of probabilistic quantitative precipitation forecasts: Flash flood case studies in the Cévennes-Vivarais region (Southern France). *Hydrological Processes*, *27*, 33–44. https://doi.org/10.1002/hyp.9543

Miller, J. D., Kim, H., Kjeldsen, T. R., Packman, J., Grebby, S., & Dearden, R. (2014). Assessing the impact of urbanization on storm runoff in a peri-urban catchment using historical change in impervious cover. *Journal of Hydrology*, *515*, 59–70. https://doi.org/10.1016/j.jhydrol.2014.04.011

Napolitano, G., See, L., Calvo, B., Savi, F., & Heppenstall, A. (2010). A conceptual and neural network model for real-time flood forecasting of the Tiber River in Rome. *Physics and Chemistry of the Earth Parts ABC*, Modelling and simulation of dangerous phenomena, and innovative techniques for hazard mapping and mitigation, *35*, 187–194. https://doi.org/10.1016/j.pce.2009.12.004

Nayak, P. C., Sudheer, K. P., Rangan, D. M., & Ramasastri, K. S. (2005). Short-term flood forecasting with a neurofuzzy model. *Water Resource Research*, *41*. https://doi.org/10.1029/2004WR003562

Nicótina, L., Alessi Celegon, E., Rinaldo, A., & Marani, M. (2008). On the impact of rainfall patterns on the hydrologic response. *Water Resources Research*, *44*, W12401. https://doi.org/10.1029/2007WR006654

Nirupama, N., & Simonovic, S. P. (2007). Increase of flood risk due to urbanisation: A Canadian example. *Natural Hazards*, *40*, 25–41. https://doi.org/10.1007/s11069-006-0003-0

Nourani, V., Hosseini Baghanam, A., Adamowski, J., & Kisi, O. (2014). Applications of hybrid wavelet–artificial intelligence models in hydrology: A review. *Journal of Hydrology*, *514*, 358–377. https://doi.org/10.1016/j.jhydrol.2014.03.057

Qaiser, K., Yuan, Y., & Lopez, R. D. (2012). Urbanization impacts on flooding in the Kansas River basin and evaluation of wetlands as a mitigation measure. *Transactions of the ASABE*, *55*, 849–859.

Reed, S., Schaake, J., & Zhang, Z. (2007). A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. *Journal of Hydrology*, *337*, 402–420. https://doi.org/10.1016/j.jhydrol.2007.02.015

Rezaeianzadeh, M., Tabari, H., Arabi Yazdi, A., Isik, S., & Kalin, L. (2014). Flood flow forecasting using ANN, ANFIS and regression models. *Neural Computing & Applications*, *25*, 25–37. https://doi.org/10.1007/s00521-013-1443-6

Roodsari, B. K., & Chandler, D. G. (2017). Distribution of surface imperviousness in small urban catchments predicts runoff peak flows and stream

flashiness. *Hydrological Processes*, *31*, 2990–3002. https://doi.org/10.1002/hyp.11230

Saleh, F., Ramaswamy, V., Georgas, N., Blumberg, A. F., & Pullen, J. (2016). A retrospective streamflow ensemble forecast for an extreme hydrologic event: A case study of hurricane Irene and on the Hudson River basin. *Hydrology and Earth System Sciences*, *20*, 2649–2667. https://doi.org/10.5194/hess-20-2649-2016

Sapriza-Azuri, G., Jódar, J., Navarro, V., Slooten, L. J., Carrera, J., & Gupta, H. V. (2015). Impacts of rainfall spatial variability on hydrogeological response. *Water Resources Research*, *51*, 1300–1314. https://doi.org/10.1002/2014WR016168

Smith, M. B., Seo, D.-J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., … Cong, S. (2004). The distributed model intercomparison project (DMIP): Motivation and experiment design. *Journal of Hydrology*, The Distributed Model Intercomparison Project (DMIP), *298*, 4–26. https://doi.org/10.1016/j.jhydrol.2004.03.040

Solomatine, D. P., & Ostfeld, A. (2008). Data-driven modelling: Some past experiences and new approaches. *Journal of Hydroinformatics*, *10*, 3–22. https://doi.org/10.2166/hydro.2008.015

Sudheer, K. P., Gosain, A. K., & Ramasastri, K. S. (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrological Processes*, *16*, 1325–1330. https://doi.org/10.1002/hyp.554

Suriya, S., & Mudgal, B. V. (2012). Impact of urbanization on flooding: The Thirusoolam sub watershed—A case study. *Journal of Hydrology*, *412–413*, 210–219. https://doi.org/10.1016/j.jhydrol.2011.05.008

Vogel, R. M. (2017). Stochastic watershed models for hydrologic risk management. *Water Security*, *1*, 28–35. https://doi.org/10.1016/j.wasec.2017.06.001

Walsh, C. J., Roy, A. H., Feminella, J. W., Cottingham, P. D., Groffman, P. M., & Morgan, R. P. (2005). The urban stream syndrome: Current knowledge and the search for a cure. *Journal of the North American Benthological Society*, *24*, 706–723. https://doi.org/10.1899/04-028.1

Wang, Y., & Yang, X. (2013). Land use/cover change effects on floods with different return periods: A case study of Beijing, China. *Frontiers of Environmental Science & Engineering*, *7*, 769–776. https://doi.org/10.1007/s11783-013-0542-z

Watson, K. M., Collenburg, J. V., & Reiser, R.G. (2013). Hurricane Irene and associated floods of August 27-30, 2011, in New Jersey (USGS Numbered Series No. 2013-5234). Scientific Investigations Report. U.S. Geological Survey, Reston, VA.

Wilson, C. B., Valdes, J. B., & Rodriguez-Iturbe, I. (1979). On the influence of the spatial distribution of rainfall on storm runoff. *Water Resources Research*, *15*, 321–328. https://doi.org/10.1029/WR015i002p00321

World Bank. (2013). *World development report 2014: Risk and opportunity—Managing risk for development*. Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/16092 License: CC BY 3.0 IGO

Yu, W., Nakakita, E., Kim, S., & Yamaguchi, K. (2016). Impact assessment of uncertainty propagation of ensemble NWP rainfall to flood Forecasting with catchment scale. *Advances in Meteorology*. Research article, *2016*. https://doi.org/10.1155/2016/1384302

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*, 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zope, P. E., Eldho, T. I., & Jothiprakash, V. (2015). Impacts of urbanization on flooding of a coastal urban catchment: A case study of Mumbai City, India. *Natural Hazards*, *75*, 887–908. https://doi.org/10.1007/s11069-014-1356-4

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Roodsari BK, Chandler DG, Kelleher C, Kroll CN. A comparison of SAC-SMA and Adaptive Neuro-fuzzy Inference System for real-time flood forecasting in small urban catchments. *J Flood Risk Management*. 2018;e12492. https://doi.org/10.1111/jfr3.12492