High-resolution hydrologic forecasting for very large urban areas

Hamideh Habibi, Ishita Dasgupta, Seongjin Noh, Sunghee Kim, Michael Zink, Dong-Jun Seo, Matthew Bartos and Branko Kerkez

ABSTRACT

With continuing growth of urban populations worldwide, high-resolution hydrologic forecasting is an increasingly important hydroinformatics service for large urban areas. In the Dallas-Fort Worth (DFW) area, the Collaborative Adapting Sensing of Atmosphere (CASA) WX program has been providing real-time hydrologic products, such as rainfall and streamflow, at 1 min–500 m resolution using the NWS Research Hydrologic Distributed Model forced by the Quantitative Precipitation Estimate from a network of X-band weather radars. There is an increasing demand, however, for even higher-spatial resolution hydrologic products. In this paper, we assess the ability of the current streamflow product to capture the hydrologic response of urban catchments in the DFW area, the utility of ultrasonic distance sensors for real-time sensing of water level in urban streams, and the feasibility of higher-resolution operation using parallel processing and cloud computing. We show that the CASA WX streamflow product skillfully captures the stage and streamflow response from rainfall for the majority of the nine catchments studied, but that timing errors significantly deteriorate the quality of streamflow prediction for certain basins. Comparative evaluation of different computing models shows that a reduction in runtime of up to 34% is possible with parallel processing at 1 min–250 m resolution.

Key words | distributed computing, urban hydrologic forecasting, water-level sensing

Hamideh Habibi (corresponding author)
Seongjin Noh
Sunghee Kim
Dong-Jun Seo
Department of Civil Engineering,
The University of Texas at Arlington,
Arlington, TX,
USA
E-mail: *hamideh.habibi@mavs.uta.edu*

Ishita Dasgupta
College of Information and Computer Sciences,
The University of Massachusetts Amherst,
Amherst, MA,
USA

Michael Zink
Department of Electrical and Computer Engineering,
The University of Massachusetts Amherst,
Amherst, MA,
USA

Matthew Bartos
Branko Kerkez
Department of Civil and Environmental Engineering,
The University of Michigan,
Ann Arbor, MI,
USA

Hamideh Habibi
Present address:
Prairie View A&M University,
Prairie View, TX,
USA

INTRODUCTION

High-resolution hydrologic forecasting is an increasingly important hydroinformatics service for large urban areas. It provides time- and location-specific predictive information necessary for life or property-saving actions and minimizes disruptions to daily lives. Flooding, in particular, poses one of the most significant natural hazards in urban areas. With increasing occurrences of heavy-to-extreme precipitation in many parts of the world (Stocker *et al.* 2013), large population centers with large impervious surfaces are particularly vulnerable where even a small but intense rainfall event can cause deadly flash floods and extensive damage (Grumm & Junker 2016). For location- and time-specific warnings, high-resolution hydrologic forecasting is a natural progression that utilizes weather radar and distributed hydrologic modeling (Seo *et al.* 2015; Rafieeinasab *et al.* 2015a). In the Dallas-Fort Worth (DFW) area, the Collaborative Adapting Sensing of Atmosphere (CASA) WX program has been operating a network of X-band radars (see Figure 1, upper-left

doi: 10.2166/hydro.2019.100

Sensor location	Contributing area (km^2)	Impervious fraction
DFW002	7.7	0.42
DFW005	24.9	0.22
DFW006	17.6	0.24
DFW030	43.7	0.39
GP6033	42.1	0.48
GP6103	32.0	0.37
GP6123	11.1	0.37
GP6133	54.3	0.31
GP6143	31.2	0.31

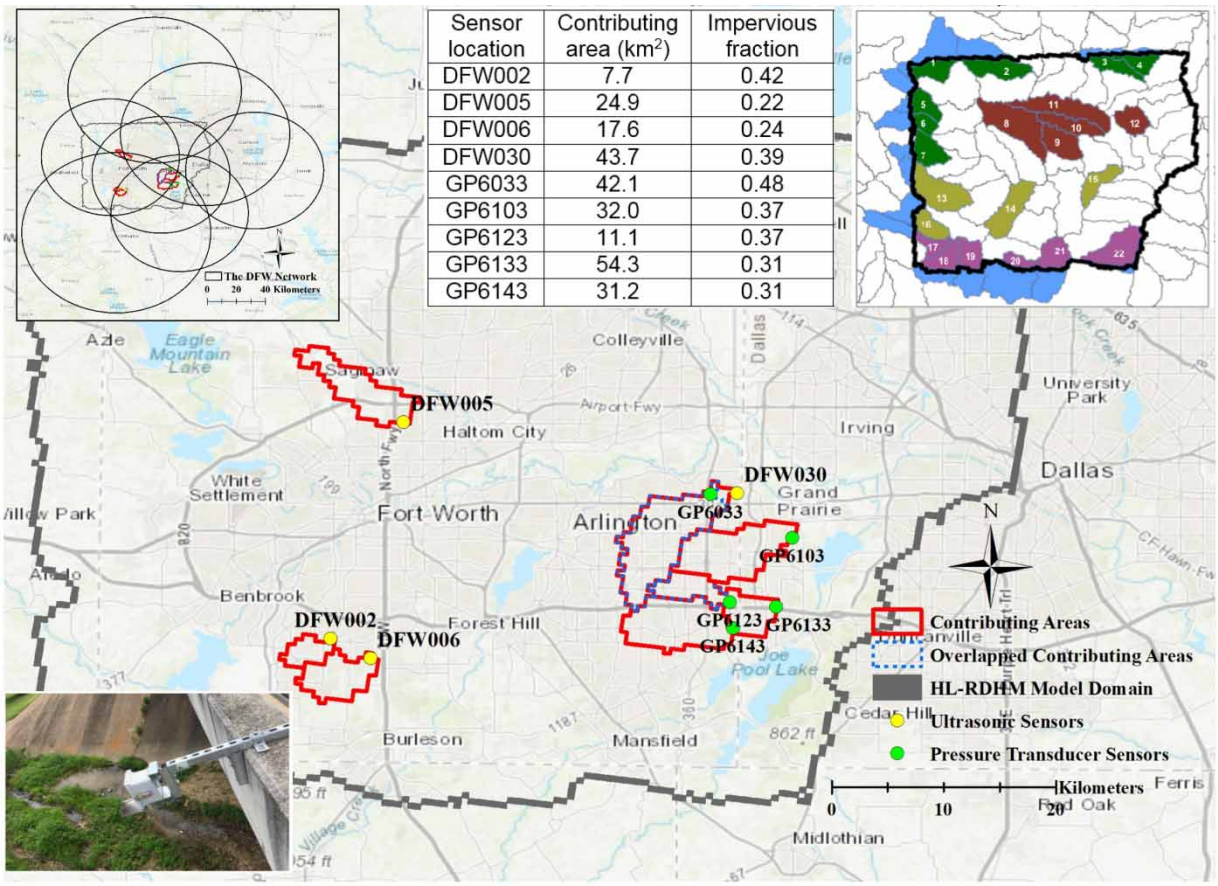

Figure 1 | The CASA WX hydrologic model domain and the location of ultrasonic and pressure transducer water-level sensors used, (upper-left inset) the DFW demonstration network, (middle inset) list of water-level sensor locations and associated attributes, (right inset) HUC-12 delineation of the model domain in which the colored and uncolored units are the HWB and DSBs, respectively, and (lower-left inset) photograph of a deployed ultrasonic sensor node. Please refer to the online version of this paper to see this figure in color: http://dx.doi:10.2166/hydro.2019.100.

inset) to provide a suite of meteorological and hydrologic products in support of severe weather and flash flood monitoring and prediction for several years now (Chen & Chandrasekar 2015; Habibi 2017). A salient aspect of the above operation is that the radar rainfall data are available at a very high resolution of 500 m and 1 min (see Chen & Chandrasekar (2015) for radar locations). The above rainfall data are input to the Research Hydrologic Distributed Model (RDHM; Koren *et al.* 2004) to produce runoff, streamflow, and return period products in real time. The characteristic spatial scale of both natural and man-made physiographic features in the area suggests that further increase in the resolution of observed rainfall, rainfall-runoff modeling, and routing would significantly improve the information content of the model output for flash flood warning, and monitoring and prediction of the urban water cycle. Toward that end, we assess, in this paper, the ability of the current streamflow product to capture hydrologic response of urban catchments in the DFW area, the utility of low-cost ultrasonic distance sensing for water-level observation for urban streams, and the feasibility of increasing the resolution of the CASA WX hydrologic products via parallel processing and cloud computing. The new and significant contributions of this paper are the evaluation of the CASA WX streamflow product for high-resolution urban water forecasting, high temporal-resolution observation of water level in urban streams via ultrasonic distance sensing, and the evaluation of different parallelization approaches to reduce

runtime for higher-resolution operation. This paper is organized as follows. In the section 'Sensing and hydrologic modeling', we describe sensing and hydrologic modeling. Section 'Computing and evaluation' describes computing and evaluation. Section 'Results and discussion' presents the results and discussion. Section 'Conclusions and future research recommendations' provides the conclusions and future research recommendations.

SENSING AND HYDROLOGIC MODELING

This section describes rainfall sensing, hydrologic modeling, and water-level sensing.

Rainfall sensing

The CASA WX radar network currently consists of seven X-band polarimetric radars located at Addison, Arlington, Cleburne, Denton, Fort Worth, Mesquite, and Midlothian, TX. For details of the radar systems and real-time aspects, the reader is referred to Chen & Chandrasekar (2015). The spatiotemporal resolution of the current CASA WX Quantitative Precipitation Estimate (QPE) product is 500 m and 1 min, which have been evaluated extensively (Chandrasekar *et al.* 2012; Chen *et al.* 2016; Cifelli *et al.* 2018). Comparative evaluation of different radar-based QPE products (Rafieeinasab *et al.* 2015b) showed that in the study area, the CASA WX QPE is generally more accurate for larger precipitation amounts, whereas the Multisensor Precipitation Estimator (MPE; Seo *et al.* 2010) estimates are more accurate for smaller amounts. Recently, the CASA WX QPE operation began fusing the QPE from the X-band radar network with that from the WSR-88D located in Burleson, TX (Chen & Chandrasekar 2015). The rainfall estimates used in this study are the resulting fused QPE product.

Hydrologic modeling

Hydrologic modeling using the RDHM for the study area has been reported extensively in Rafieeinasab *et al.* (2015a), Habibi *et al.* (2016), and Habibi & Seo (2018). The RDHM (Koren *et al.* 2004) is a grid-based distributed model and includes the Sacramento soil moisture accounting model (SAC; Burnash *et al.* 1973) and kinematic-wave models (Chow *et al.* 1988) for hillslope and channel routing. The RDHM has been used at various River Forecast Centers for flash flood and river flood forecasting and at the National Centers for Environmental Prediction for high-resolution soil moisture analysis at the continental scale (Smith *et al.* 2004). The CASA WX hydrologic model domain includes the Cities of Fort Worth, Arlington, Grand Prairie, and Dallas with a combined area of almost 2,400 km^2 (see Figure 1). The spatial resolution of the model currently in operation for CASA WX is approximately $500 \times 500\ m^2$ or 1/8 Hydrologic Rainfall Analysis Projection (HRAP). Estimation of the a priori parameters for the SAC and kinematic-wave routing models used in RDHM, limited calibration, and validation for the DFW area is described in Rafieeinasab *et al.* (2015a), Norouzi (2016), Habibi *et al.* (2016), Habibi (2017), and Habibi & Seo (2018).

Rafieeinasab *et al.* (2015a) has shown that the RDHM produces skillful simulations in the study area with the a priori parameters (Koren *et al.* 2004) and that limited calibration improves accuracy. In reality, calibration is possible only for a relatively small number of catchments where streamflow observations are available. Even though a number of cities in the DFW area operate water-level sensors, they do not, in general, develop rating curves. To estimate observed flow, it is therefore necessary to derive and update rating curves via *in situ* sensing (Bowden & Clayton 2010) or numerical modeling (Norouzi 2016). Developing rating curves for a large number of locations across multiple municipalities is a large effort and is not expected to be undertaken within the foreseeable future. Given this reality, we derive in this work pseudo rating curves by probability matching the model-simulated flow with high temporal-resolution observations of water level (see the section 'Hydrologic evaluation').

Water-level sensing

In the DFW Metroplex, a number of cities operate networks of rainfall and pressure transducer-based water-level sensors. The primary purpose of water-level sensing by the cities is to detect localized flooding that warrants immediate action such as flashing warning signs or closing roadways.

As such, water-level observations from many of their sensors are not very well suited for hydrologic evaluation. To aid hydrologic modeling and to help understand the hydrologic response of urban catchments, a network of ultrasonic distance sensors has been newly deployed by the authors. The ultrasonic stream stage-monitoring network is built using the *open storm* stack – an open-source, end-to-end wireless sensing framework which includes a low-power cellular-enabled data logger along with integrated cloud services for rapid ingest, processing, and communication of sensor readings (Bartos *et al.* 2018). To date, a total of 32 sensor nodes have been deployed throughout the DFW area of which 15 are currently active. Figure 1 shows the map of the locations of the sensors used in this work along with the photograph of a deployed ultrasonic sensor node (lower-left inset). Ultrasonic sensor sites are selected in consultation with the cities to fill gaps in the existing sensor network and to capture runoff response at varying spatial scales. The contributing areas at the sensor locations range from 2 to 80 km^2 (see Figure 1, middle inset), enabling the characterization of both small-scale pluvial floods and larger-scale riverine floods. The stage-monitoring network has been in operation since May 2016 with several sites possessing continuous data records of over 2 years.

Water level is measured using ultrasonic depth sensors mounted to bridge decks or railings directly overlooking the waterways of interest (see Figure 1, lower-left inset). These sensors measure the distance to the water level by emitting ultrasonic pulses and measuring the time it takes for a reflected pulse to reach the sensor's receiver. The ultrasonic sensor has a maximum range of about 10 m with a typical reading-to-reading measurement error (standard deviation) of less than 1 mm assuming no obstructions. The material cost for an ultrasonic sensor unit is about US$500. Data are collected at temporal resolutions ranging from 1 min to 1 h using a custom-adaptive sampling routine. This adaptive sampling routine uses web services to automatically adjust the sampling frequency of the sensor node in response to local weather forecasts – increasing the frequency when rain is imminent and decreasing it when no rain is anticipated. This functionality allows for hydrologically significant events to be captured at a high temporal resolution of 1–5 min while saving power during dry-weather conditions. Sensor data are transmitted in real time from field-deployed sensor nodes to a remote server using cellular telemetry. Quality control of sensor data is performed using basic automated filtering, including range checks, as well as manual classification. Spurious readings such as large spikes resulting from sensor obstruction are flagged for removal from subsequent analyses.

COMPUTING AND EVALUATION

This section describes parallelization modeling and hydrologic and computational evaluation.

Computing

There is a strong need for operation at a higher-spatial resolution. The increase in resolution, however, poses a large increase in computing time which currently prevents the execution of the model in real time. Although high-performance computing is a possibility, dependence on such a capability poses a possible resource bottleneck for real-time forecasting if unavailable or unaffordable. Generic cloud virtual machines (VM), on the other hand, are now readily available. The above situation motivated parallelization approaches for running the existing hydrologic model in the cloud. We use single-host python multithreading to implement our parallelization approaches. If parallelization makes predictions scalable and robust in real or near real time for cloud computing of high-resolution distributed hydrologic models, dependence on high-performance compute clusters would be reduced. This is particularly important for real-time forecasting for large urban areas where a large-scale heavy-to-extreme event would require operating high-resolution hydrologic models over a large model domain. Currently, the RDHM runs in real time on an Intel® Xeon® CPU E5620 @2.40 GHz 4 CPU-core Linux computer over the DFW domain at 500 m^{-1} min resolution (see Figure 1). In this work, we assess the reduction in runtime attainable from different parallel computing approaches toward real-time operation at 250 m^{-1} min resolution. For this purpose, we divide a model run into the headwater basin (HWB) and downstream basin (DSB) processes instead of running the model over the entire domain sequentially as is currently the case. The right inset in Figure 1 shows the basins (thin solid

black line) within and around the model domain (thick solid black line) as delineated by the Hydrologic Unit Code Level 12 (Seaber *et al.* 1987). The HWBs are shaded in color, whereas the DSBs are in white. Two computational approaches are considered in place of the current sequential processing: parallel processing of HWBs and interleaved processing of HWBs and DSBs. Below we describe the processing approaches compared in this work.

Sequential processing (Model A)

In sequential processing, we run the RDHM over the entire DFW domain with no division of the HWBs and DSBs processes. This baseline model, referred to as Model A, is used currently for the generation of the CASA WX hydrology products in real time.

Parallel processing of headwater catchments (Model B)

In this approach, referred to as Model B, the RDHM is first executed for only the HWBs in parallel. This is followed by a serial single execution of the DSBs based on the HWB results for the run period requested. For the HWBs, two cases of parallelism are considered. In the first, the HWBs are divided into subgroups of four HWBs (see Figure 1 for different subgroups in different colors) and the subgroups are processed in parallel. In the second, all HWBs are processed in parallel.

Interleaved processing of headwater and downstream catchments (Models C and D)

In this approach, the RDHM is executed for the HWBs and DSBs for each timestep in parallel. Within each timestep, parallel HWB processing is followed by execution for the entire DSBs for each timestep, hence the name 'interleaved'. In typical hydrologic model execution, the DSBs are processed after the HWBs are run to prescribe the upstream boundary conditions for channel flow for the DSBs. In our operation, the inflows from the HWBs into the DSBs differ little over two successive timesteps owing to the very small timestep of 1 min. As such, one may run the model for both the HWBs and DSBs in parallel, but with the inflow into the DSBs prescribed by the channel flow from the HWBs calculated from the previous timestep. In this way, the RDHM may effectively be executed in parallel for all basins with the upstream boundary conditions for the DSBs staggered by a timestep. Models C and D fall under this category as they both use the same design concept of parallelism. They only differ in their implementation method as described below.

The parallel processing of the RDHM uses the same resource at all timesteps. To deal with this resource contention, i.e., a conflict over access to a shared resource such as memory, CPU, network, or storage, Models C and D use thread locking and resource copying, respectively, as explained below. Parallel processing of each HWB requires access to the common resource that defines the run instructions for the RDHM for each timestep. The model run for each timestep is then processed by a single thread for the given range of timesteps. If all threads attempt to access and modify this common resource, they may be deadlocked. To address the situation, Model C uses a thread-locking mechanism which allows only one thread to modify the common resource while other threads wait. Alternatively, as in Model D, we may provide each thread with a unique copy of the resource such that all threads can access and modify its own copy at the same time. Thread locking diminishes the benefits of the HWB parallelization somewhat, resulting in increased processing time. Resource copying eliminates this drawback but potentially at the expense of increased memory usage due to running multiple copies per timestep. In our experiments, however, the increase in memory usage is negligible.

Evaluation

To evaluate the accuracy of streamflow simulations from the RDHM forced by the operationally produced CASA WX QPE, we compared the model-predicted stage with water-level observations and model-simulated flow with pseudo observations of flow. To assess the potential gains from the parallel processing models described earlier, we designed and carried out a set of computing experiments using three flooding events. Hydrologic evaluation is based on the real-time products, whereas computational evaluation is not.

Hydrologic evaluation

Water-level observations from four ultrasonic sensors deployed by the authors and five pressure transducer sensors operated by the City of Grand Prairie, TX, are used. Figure 1 shows the contributing areas associated with the sensor locations. The period of record is June 2016 to November 2017 for DFW002 and DFW005, March 2016 to May 2017 for DFW030, and January 2016 to March 2018 for all GP locations. The sampling interval is 10 and 15 min for all ultrasonic and pressure transducer locations, respectively. The period includes a few to several relative large events whose magnitudes approach return periods of 5 years for 12- and 24-h durations and 10 years for 6-h duration. The fast-moving convective front on 16 January 2017, for example, produced close to 100 mm of rain in about 6 h in parts of the Johnson Creek Catchment (GP6033, DFW030). The accuracy of a pressure transducer varies with the manufacturer, measurement range, and water depth. The measurement error for most situations is about 3 mm (USGS 2010). Rating curves currently do not exist at these locations. In this work, we use a data-driven approach to obtain flow from the sensor observations instead of developing rating curves using *in situ* observations or hydraulic models (Norouzi 2016). The resulting relationships are referred to herein as pseudo rating curves. To derive pseudo rating curves, we used probability matching (Hashino *et al.* 2002) and related the model-simulated flow with observed stage via quantile–quantile mapping. One could also use parametric models such as the power law model used in Norouzi (2016) to obtain the relationship. Our experience, however, is that parametric models lack the degrees of freedom necessary to capture the complex channel geometry and flow conditions frequently observed in diverse urban streams in the DFW area. Also, unlike probability matching, parametric modeling necessarily assumes negligible timing errors in simulated flow, which is often not met in reality particularly in high-resolution modeling. Due to the empirical nature, probability matching can only handle up to the largest historically simulated flow. To convert flow to stage when the model predicts record-breaking flow in real-time operation, it is necessary to model the extreme right tail. In this work, we used the hyperbolic model (Deutsch & Journel 1998), which parsimoniously extends the pseudo rating curve beyond the largest simulated flow.

The accuracy of pseudo rating curves for all-important large flows depends greatly on the sample size and the quality of the upper-tail modeling. Because the period of record is rather short in this work, it was not possible to perform true validation of the model-predicted stage. Instead, we assume that the pseudo rating curves are updated daily and carried out leave-one-day-out cross validation. In this way, the sample size is maximized while avoiding dependent validation. In the context of real-time operation, the above strategy amounts to updating the pseudo rating curves daily thereby utilizing the most recent data. To assess the goodness of the resulting stage and flow predictions, we used the correlation coefficient (CORR), the Nash–Sutcliffe Efficiency (NSE; Nash & Sutcliffe 1970), the probability of detection (POD) or hit rate, and the probability of false detection (POFD) or false alarm rate (Wilks 2006). The use of the normalized measures and skill scores is to reduce the influence of catchment size and the magnitude of flow or stage in our assessment. Their widespread use also helps communicate the results to diverse stakeholders. The CORR and NSE are calculated unconditionally, i.e., using all available observations, and conditionally on the observed stage exceeding the 90th percentile. The 90th percentile threshold isolates almost all periods of increased water level due to rainfall. The conditional statistics therefore measure the skill in model-predicted stage during rainfall events. The POD and POFD are calculated for the 90th and 97.5th percentiles of observed stage. For each of the two percentiles, an event is defined as instantaneous water level exceeding the respective threshold.

Ideally, we would like to produce inundation maps based on water-level predictions. Without hydraulic modeling, however, water level may be predicted only at the sensor locations. It is hence necessary in the current operation to translate the magnitude of model-simulated flow into a spatially continuous depiction of flooding potential within the model domain. For the above, the CASA WX uses the threshold frequency (TF) technique developed for the RDHM or DHM-TF (Reed *et al.* 2007), which expresses the magnitude of flow in terms of the return period. For CASA WX, the frequency analysis necessary for the technique was carried out using the historical MPE data due

to the shortness of the historical CASA WX QPE archive. Because the accuracy of translating flow into return period depends on the absolute accuracy of the model-predicted flow, we are interested in evaluating the accuracy of model-simulated flow in addition to that of stage. Because streamflow observations are not available, we convert observed water level into pseudo streamflow observations via probability matching. This process is completely analogous to comparing model-simulated stage with observed stage described earlier but between the model-simulated flow and the pseudo-observed flow. Model-simulated flow is subject not only to amplitude errors but also to phase errors (Liu *et al.* 2011). To assess timing errors, we correlate the time series of simulated flow with that of observed stage. Necessarily, such timing error analysis is possible only at the sensor locations. Although the resulting diagnostic information may be used to improve or calibrate the models, it is not readily possible to use such information in real-time operation. For this reason, we made no attempt in our evaluation at correcting timing errors to emulate the real-world conditions.

Computational evaluation

To test and compare the computing models described earlier, a set of computational experiments were designed and carried out. In these experiments, we considered that 1-min CASA WX rainfall data are available in 5-min batches as in the current real-time operation. The resolution of the RDHM and the radar rainfall data used is 1/16 HRAP (250 m × 250 m) to emulate the targeted higher-resolution operation. For computing, we used the Cloudlab linux x86_64 VMs with 56 processors (Intel® Xeon® CPU E5-2683 v3 @ 2.00 GHz). To evaluate the performance of our approach under the typical conditions in the Cloud, we use the Cloudlab resources that are readily available for any user at any time. The set-up for the RDHM model is available online (NWS 2009) serves as the baseline for the hydrologic model. Some modifications were necessary to the existing RDHM model so that the HWBs and DSBs may be processed separately for Models B, C, and D. To run the models, one only has to transfer the data and model implementation to any such Cloudlab resource. Three experiments were carried out as described below.

In Experiment 1, we compare the performance of the different computing models when all 5-min data are received at once vs. when only 1 min's worth of data is received at an interval of 1 or 2 min. For each timestep, the RDHM needs start and end times for processing the rainfall data. If the time interval between receiving rainfall data is 1 min, two radar QPE files are necessary to process a unit timestep range for a total wait time of 2 min. In Model B, all HWBs are processed for all timesteps before the DSBs. In Model D, with multi-timestep run periods, the DSB processing for t_i at the ith timestep starts right after all HWBs are processed for t_i rather than waiting for their processing for t_{i+1} through t_n to be completed where n is the total number of timesteps in the run period. In Experiment 2, we compare the performance of lower-degree parallelization vs. full parallelization of HWB processing and assess the marginal value of full parallelization vs. additional computational cost. In Experiment 3, we compare the performance of the different computing models for three different types of rainfall events to assess robustness: Case (1) squall line with small tornadoes, 0700–0800 29 March 2017, Case (2) non-tornadic localized convective rainfall, 2200–2300 UTC 5 July 2017, and Case (3) widespread stratiform rainfall, 1,400–1,500 UTC, 8 November 2017.

RESULTS AND DISCUSSION

This section presents the results from hydrologic and computational evaluation.

Hydrologic evaluation

To provide a sense of the quality of the simulation, we first show an example of observed vs. model-simulated stage. Figure 2 shows the observed (black) vs. predicted (red) stage hydrographs at DFW006 and the associated hyetograph (blue) for a series of significant rainfall events that occurred from 26 May to 3 June 2016. It shows that the probability-matched RDHM stage simulation agrees very well with the observed stage, but that significant differences exist in the last two peaks for this relatively small and fast-responding urban catchment (17.6 km^2). It also shows the utility of probability matching for relating

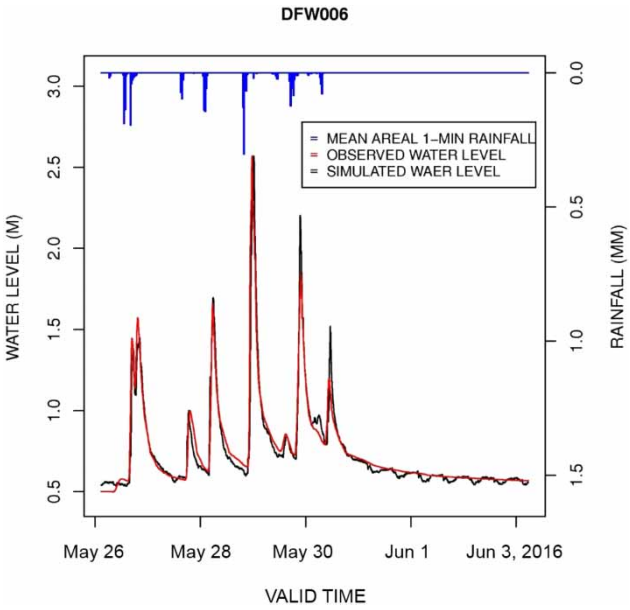

Figure 2 | Example of the observed (black) vs. predicted (red) stage hydrographs and the associated hyetograph (blue) for a series of significant rainfall events in 26 May through 3 June 2016, at DFW006. Please refer to the online version of this paper to see this figure in color: http://dx.doi:10.2166/hydro.2019.100.

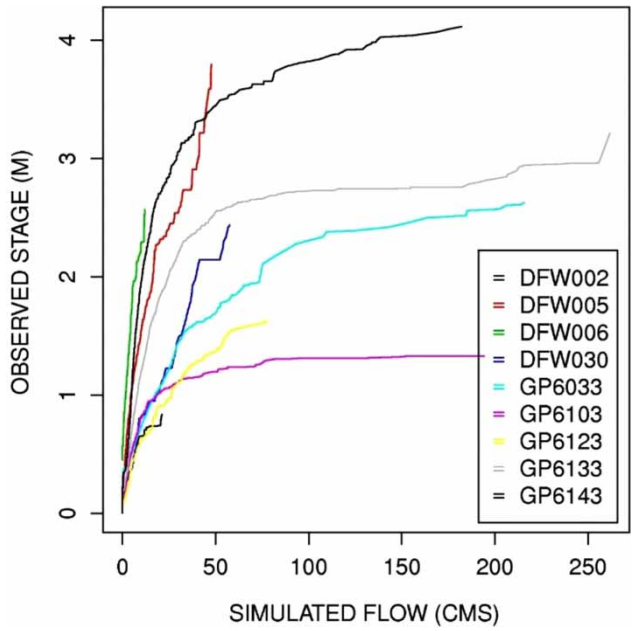

Figure 3 | Observed stage vs. simulated flow relationships at all nine locations.

observed stage and simulated flow in the absence of rating curves; the pseudo rating curves can easily be updated in real time so that, as the period of operation becomes longer, the more statistically reliable they become while reflecting possible changes in the stage–discharge relationship. In Figure 2, apparent diurnal variations are also seen in the recession limb of the observed water level. Such variations are observed only at the ultrasonic sensor locations and are more pronounced in the inter-storm periods. The most likely explanation for these variations is related to minor temperature dependence – the ultrasonic sensor assumes a constant temperature to estimate the speed of sound. However, it is possible that the diurnal cycle of evapotranspiration (Cuevas *et al.* 2010; Mutzner *et al.* 2015) and the urban water cycle (e.g., the household use of water in urban areas) may also be the contributing factors. We are currently testing a new module that accounts for temperature variations in distance sensing, and the results will be reported in the near future.

Figure 3 shows the observed stage vs. simulated flow relationships at all nine locations. The varying degree of nonlinearities in the pseudo rating curves reflects the large variations in the channel cross-section at the outlet of the catchments and indicates the difficulty of modeling rating curves using power law functions alone. In high-resolution forecasting, even a relatively small timing error may greatly deteriorate the predictive skill, particularly for fast-responding urban streams with very small time-to-peak. To assess possible timing errors, we examine the cross-correlograms between the simulated flow and the observed stage at all nine locations in Figure 4. They are conditioned on the observed stage exceeding the 90th percentile. The noisy scatters around the ultrasonic sensor results are associated with very small sample size and may safely be ignored. Cross correlation peaking at positive and negative lags indicates that the model-simulated flow trails and precedes the observed stage, respectively. To assess the possible dependence of the speed of flood wave on the magnitude of flow, we carried out correlogram analysis using different thresholds. It was found that above the 90th percentile, the variations are negligible. Figure 4 indicates that, at DFW030 and GP6033, the RDHM simulation rises a little too late whereas at DFW006 and GP6103, the model simulation rises too early. At all other locations, no clearly discernible timing errors are observed.

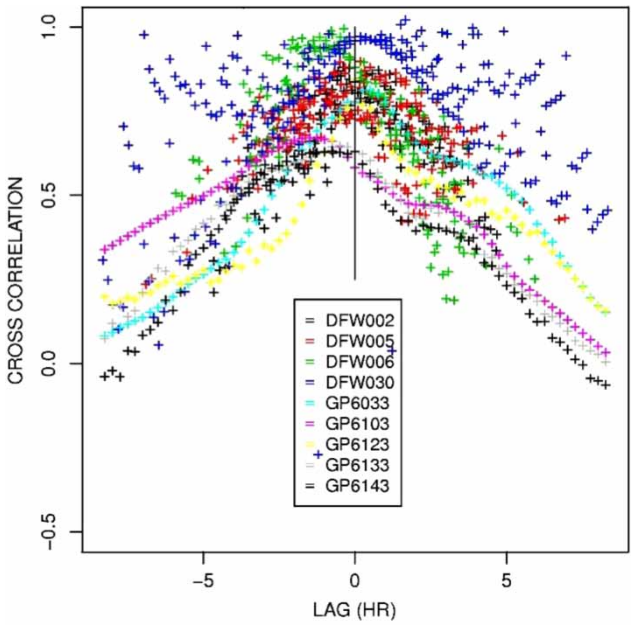

Figure 4 | Cross-correlograms between the simulated flow and observed stage at all nine locations.

Figure 5 shows the CORR and NSE of model-predicted stage for all nine locations. The unconditional and conditional statistics are in black and red, respectively. The conditional statistics are based on the verifying observed water level exceeding the 90th percentile. Due to the shortness of the period of record, conditioning on higher percentiles was not possible. The 90th percentile is most useful in assessing the skill of the CASA WX streamflow product in predicting how the flow and stage may respond during rainfall events. In Figure 5, the closer to the upper-right and lower-left corners of the panel, the more and less skillful the model prediction is, respectively. For brevity, the names of the sensor locations have been shortened in the figure. The prefixes, 'd' and 'g', signify DFW and GP, respectively, and the first digit '6' has been dropped in the names of the Grand Prairie locations. For all cases, CORR is about 0.80 or larger and NSE is about 0.60 or larger for stage prediction at all sensor locations. Expectedly, the skill levels are lower for prediction of water levels exceeding the 90th percentile.

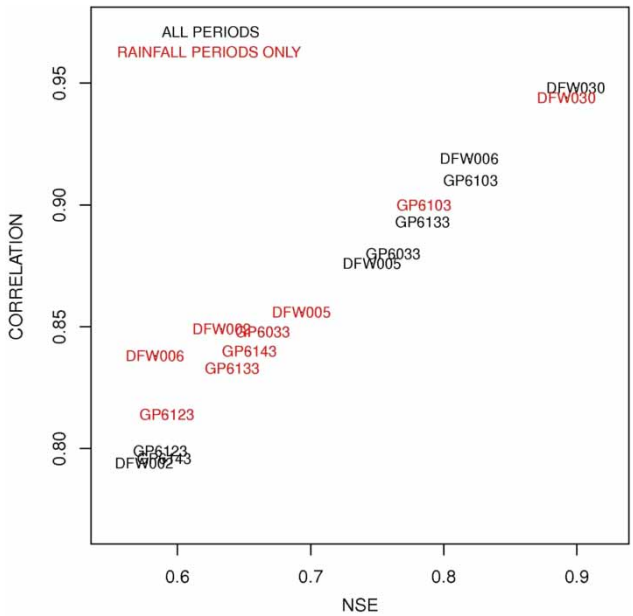

Figure 5 | CORR and NSE of model-predicted stage for all nine locations. Please refer to the online version of this paper to see this figure in color: http://dx.doi:10.2166/hydro.2019.100.

Examination of model-simulated flow vs. pseudo-observed flow indicates that the skill scores for flow are much more widespread than those for stage. Whereas the highest scores for flow prediction remain close to those for stage prediction, the lower scores for flow prediction are significantly deteriorated compared with those for stage prediction. These changes in skill scores reflect the site-specific nonlinear relationships of stage vs. flow shown in Figure 3 and point out the importance of validating both stage and flow for hydrologic evaluation. Probability matching used for pseudo rating curves depends solely on the marginal distributions of observed stage and simulated flow and, hence, does not reflect timing errors in model simulations. In flow predictions, timing errors are greatly amplified due to the large magnitude particularly in high flow conditions. Specifically, the scores for GP6103 for flow prediction are significantly lower than those for stage prediction, a reflection that the model simulation has significant timing errors at this location (see Figure 4). Excluding GP6103, the CORR and NSE for streamflow prediction are about 0.70 or larger and 0.40 or larger, respectively, for all cases. The above results suggest that, overall, the CASA WX streamflow product has good skill in capturing the streamflow response of urban catchments to rainfall, and that, beyond flash flood forecasting, the product may also be used for routine high-resolution monitoring and prediction of streamflow transiting through the urban water cycle.

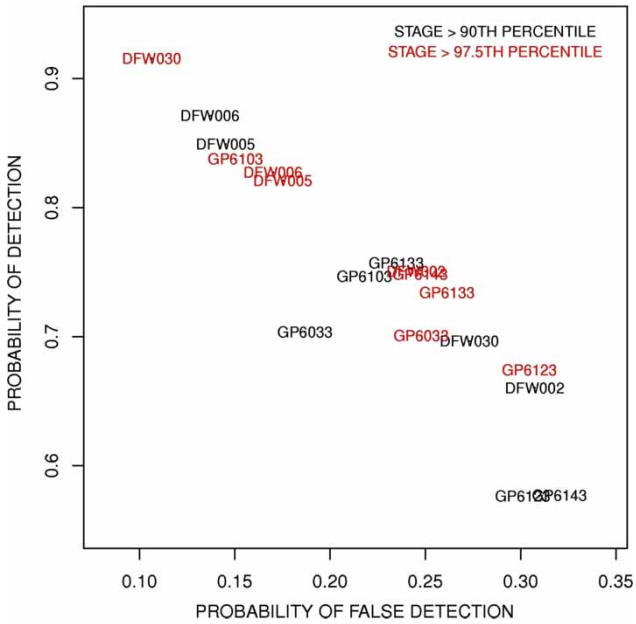

Figure 6 | Same as Figure 5 but for (POD, POFD).

We examine the (POD, POFD) results for stage prediction at all locations in Figure 6. The black and red results represent the skill in predicting events in which water level exceeds the 90th and 97.5th percentiles of observed stage, respectively. The 97.5th percentile results assess the ability of the CASAWX streamflow product to identify the time and location of potentially flood-producing streamflow conditions. The closer to the upper-left and lower-right corners, the larger and smaller the skill is, respectively, in discriminating an event vs. a non-event as defined above. An imaginary line connecting (0,0) and (1,1) would represent the no-skill line as in the Relative Operating Characteristic curve used widely in ensemble verification (Wilks 2006). We also examined the (POD, POFD) results for streamflow prediction. Comparison with the stage prediction results indicates that stage prediction generally has larger discriminatory skill than flow prediction owing to the reduced variability in stage. In particular, the POFD is significantly smaller for stage prediction. In Figure 6, the (POD, POFD) results loosely form three different groups of high, moderate, and low discriminatory skill. GP6123 and GP6143 show, by far, the smallest discriminatory skill. It may be explained by the fact that GP6123 is the second smallest basin (11.1 km^2) and hence has very short memory, and GP6143 has the largest dynamic range in stage (see Figure 3). The reason for the lower discriminatory skill for the 97.5th-percentile threshold for GP6033 is less clear. This location is at the outlet of a highly urbanized, extremely flashy catchment. Although relatively large (42.1 km^2), the catchment has a time-to-peak of only about 45 min with multimodal empirical unit hydrograph due to the elongated shape (see Figure 12 in Rafieeinasab *et al.* (2015a)). It is very likely that, albeit relatively small, the timing error of about 30 min (see Figure 4) significantly reduces the discriminatory skill in flow prediction. Figure 6 and the streamflow results indicate that, for events that produce significant hydrologic response, POD is about 0.65 or larger and POFD is about 0.40 or smaller for all cases. For the most skillful subgroup, POD and POFD are greater and less than 0.80 and 0.20, respectively. The above results are only for significant events rather than flood events. As such, the POD and POFD numbers above are significantly larger and smaller, respectively, than those for flash flood forecasts even though the spatiotemporal scale of the events dealt with in this work is much smaller (see, e.g., Clark *et al.* 2014). We are continuing searching the archives of the data and model output to increase the sample size, and the verification results for larger thresholds will be reported in the near future.

Computational evaluation

This subsection presents the results from the computational experiments described in the section 'Computational evaluation'.

Comparison under different data availability

We compare the runtime among the different computing models when 5 min's worth of rainfall data is available all at once vs. when each min's worth is received in 1- or 2-min intervals. This comparison is carried out on a 56-core cloud machine. Figure 7 shows the results from full parallelization. When all rainfall data are available in 5-min batches, Models B, C, and D achieve 23, 28, and 34% reduction in runtime over the baseline in full parallelization. As the time interval between the data ingest

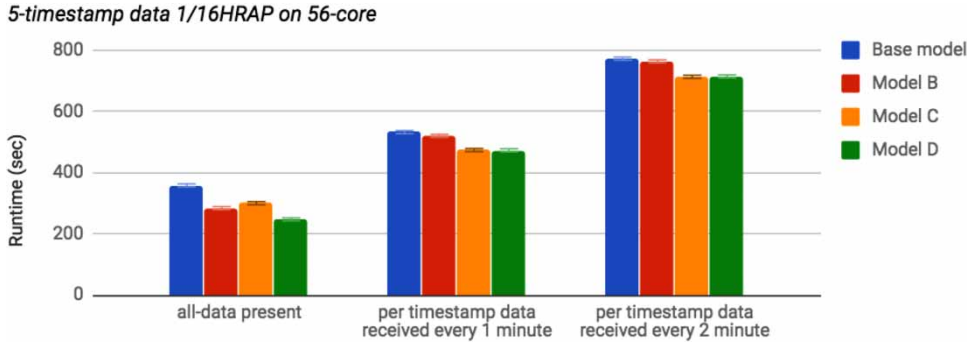

Figure 7 | Runtime comparison among the different computing models with full parallelization when 5 min's worth of rainfall data is available all at once vs. when each min's worth is received in 1- or 2-min intervals.

increases, however, the reduction in runtime over the baseline decreases regardless of the model. Figure 7 also shows that, for all three cases of data arrival, the best performance is achieved by Model D. Figure 7 shows the large impact of reducing data latency on runtime reduction relative to the gains from the different computing models considered. Currently, the CASA radar network provides 5 min's worth of rainfall data every 5 min. Figure 7 indicates that Model D is within the feasible bound for 1-min 250-m operation but only with no data latency. Because the CASA data run about 75 s latent in reality, the real-world situation would be somewhere between the middle and rightmost bar graphs in Figure 7. Accordingly, for real-time implementation of Model D at 1-min 250-m resolution over the entire model domain, it is necessary to reduce runtime by an additional min or so.

Figure 7 indicates that the interleaved processing of Model D performed approximately 15% better compared to the parallel processing of Model B. When all rainfall data are available at once, the bottleneck in Model B occurs at the timestep where the HWB processing takes the longest. In Model D, on the other hand, the bottleneck occurs at the timestep where the joint processing of HWBs and DSBs takes the longest. Because the delay due to the bottleneck in Model D is always smaller than that in Model B, Model D always performs better than Model B. Both Models C and D use interleaved processing with an equal degree of HWB parallelization and differ only in that the Model C implements thread locking, whereas Model D uses resource copying. When all data are present, the resource copying mechanism in Model D improved the runtime by 8% compared to Model C.

Effect of degree of parallelization

In full parallelization, each HWB is executed with a single thread. Full parallelization, however, can be very resource intensive as the run period increases. In four-fold parallelization, all HWBs are divided into subgroups of four, and each headwater subgroup is processed with a single thread. When computing resources are not available, lower-degree parallelization can be an alternative to full parallelization. In our experiments, a total of 66 threads were active in full parallelization, whereas at most 12 threads were active at any time in four-fold parallelization. Under four-fold parallelization, Model D improved its runtime by 31% compared to the baseline, whereas under full parallelization, the improvement was 34%. We therefore tested if a reduced degree of parallelization may yield results that are comparable to or competitive with full parallelization. The runtime differences between the four-fold and full parallelization approaches in Models B, C, and D indicate that Model D reduces runtime by 6.5% due to full parallelization of the HWB processes when compared to the four-fold approach. The largest improvement by full parallelization is in Model C due to the significantly reduced overhead associated with the thread locking which is absent in other models. The large overhead above arises from the fact that, in Model C, the wait for a group of HWBs in four-fold parallelization is significantly longer than the wait for a single HWB in full parallelization.

Performance for different types of rainfall events

Finally, we assess Model D's ability to reduce runtime consistently for different types of rainfall events. The 5-min test periods each consisting of five timesteps were used for Cases 1, 2, and 3 described in the section 'Computational evaluation', respectively. Examination of the runtime of the baseline vs. Model D shows that Model D performs better than the baseline for all the three cases. The reduction in runtime ranges from 33 to 35% which is consistent with the margin of improvement seen in the section 'Hydrologic evaluation'. It is also seen that the baseline run takes slightly longer for Cases 1 and 3 compared to Case 2 due to the larger spatial extent of rainfall. The runtime for Model D, on the other hand, remains approximately the same for all cases, but always better than the baseline. These results indicate that Model D is expected to reduce runtime by 30% or more consistently regardless of the type of the rainfall event with four-fold or full parallelization of the HWB processing.

CONCLUSIONS AND FUTURE RESEARCH RECOMMENDATIONS

The main conclusions from this work are as follows. The CASA WX streamflow product skillfully captures the streamflow response to rainfall for the majority of the catchments. At these locations, the CORR and the NSE exceed 0.80 and 0.60, respectively. For basins with relatively large timing errors, however, the skill is significantly lower. Collectively, the skill is larger and more uniform across the basins for prediction of water-level response than of streamflow response due to the much smaller variability of the former. Similarly, in discriminating significant hydrologic response due to rainfall, the CASA WX streamflow product has larger skill for stage prediction than flow prediction. The probabilities of detection and false detection for capturing significant stage response are 0.65 or higher and 0.30 or lower, respectively. Overall, the results indicate that the CASA WX streamflow product may be used not only for flash flood forecasting but also for routine monitoring and prediction of streamflow transiting through the urban water cycle.

Comparative evaluation of different computing models for parallel processing and cloud computing shows that up to 34% reduction in runtime is possible for model execution at 1-min 250-m resolution relative to the current sequential computing approach. The best performing model employs interleaved processing of the HSBs and DSBs, full parallelization of the HSB processing, resource copying for parallel runs, and no latency in arrival of rainfall data. Any multi-thread-supporting computing environment with 16 or more core CPUs can be adopted for the real-time high-resolution hydrologic forecasting. Performance evaluation indicates that the above computing model reduces runtime equally well for different types of rainfall events. Additional research is needed, however, to test the scalability of the above computing model when the spatiotemporal coverage of the radar rainfall data is increased. The approaches described in this work are applicable to operation of other distributed hydrologic models for large urban areas. We are currently implementing WRF-Hydro (Gochis *et al.* 2014), the core model for the National Water Model (Graziano *et al.* 2017), for parallel operation with the RDHM for CASA WX, and the experience and evaluation results will be reported in the near future. A significant gap in high-resolution hydrologic forecasting is the very limited availability of high-quality observations of flow or water level in urban streams and of soil moisture in urban catchments. Much greater observational efforts, such as the ultrasonic water-level sensing described in this work, are necessary for modeling, parameter estimation, validation, calibration, data assimilation, and data-driven discovery in support of high-resolution hydrologic forecasting for large urban areas.

ACKNOWLEDGEMENTS

This material is based upon work supported in part by the NSF under Grants CyberSEES-1442735 and CNS-1350752, and by the NOAA's JTTI Program under Grant NA17OAR4590184. These supports are gratefully acknowledged. We would like to thank Stephanie Griffin and Barry Fulfer of the City of Grand Prairie for providing the water-level data and Eric Lyons of the University of Massachusetts Amherst for help in various phases of this work.

REFERENCES

Bartos, M., Wong, B. & Kerkez, B. 2018 Open storm: A complete framework for sensing and control of urban watersheds. *Environmental Science: Water Research & Technology* **4** (3), 346–358.

Bowden, W. B. & Clayton, M. 2010 *Vermont stormwater flow monitoring project final report 2006–2008*. Prepared for the Vermont Agency of Natural Resources, Department of Environmental Conservation, February 19.

Burnash, R. J. C., Ferral, R. L. & McGuire, R. A. 1973 *A Generalized Streamflow Simulation System – Conceptual Modeling for Digital Computers*. National Weather Service, NOAA, and the State of California Department of Water Resources Tech. Rep., Joint Federal-State River Forecast Center, Sacramento, CA, 68 pp.

Chandrasekar, V., Wang, Y. & Chen, H. 2012 The CASA quantitative precipitation estimation system: a five year validation study. *Nat. Hazards Earth Syst. Sci.* **12** (9), 2811–2820.

Chen, H. & Chandrasekar, V. 2015 The quantitative precipitation estimation system for Dallas–Fort Worth (DFW) urban remote sensing network. *J. Hydrol.* **531**, 259–271.

Chen, H., Lim, S., Chandrasekar, V. & Jang, B. J. 2016 Urban hydrological applications of dual-polarization X-band radar: case study in Korea. *J. Hydrol. Eng.* **22** (5), E5016001.

Chow, V. T., Maidment, D. R. & Mays, L. W. 1988 *Applied Hydrology*. McGraw-Hill, New York, pp. 530–537.

Cifelli, R., Chandrasekar, V., Chen, H. & Johnson, L. E. 2018 High resolution radar quantitative precipitation estimation in the San Francisco Bay area: rainfall monitoring for the urban environment. *J. Meteorol. Soc. Japan Ser. II* **96**, 141–155.

Clark, R. A., Gourley, J. J., Flaming, Z. L., Hong, Y. & Clark, E. 2014 CONUS-wide evaluation of National Weather Service flash flood guidance products. *Weather Forecast.* **29** (2), 377–392.

Cuevas, J., Calvo, M., Little, C., Pino, M. & Dassori, P. 2010 Are diurnal fluctuations in streamflow real? *J. Hydrol. Hydromech.* **58** (3), 149–162.

Deutsch, C. V. & Journel, A. G. 1998 *Geostatistical Software Library and User's Guide*. Oxford University Press, New York.

Gochis, D. J., Yu, W. & Yates, D. N. 2014 *The WRF-Hydro Model Technical Description and User's Guide, Version 2.0*. NCAR Technical Document.

Graziano, T., Clark, E., Cosgrove, B. & Gochis, D. 2017 *Transforming NOAA Water Resources Prediction*. AMS Meeting, Seattle, WA.

Grumm, R. H. & Junker, N.W. 2016 *Historic Ellicott City Flood of 30 July 2016*. National Weather Service State College, PA, 23 pp. Available from: http://cms.met.psu.edu/sref/severe/2016/30Jul2016.pdf.

Habibi, H. 2017 *Integrated Modeling of Storm Drain and Natural Channel Networks for Real-Time Flash Flood Forecasting and Stormwater Planning and Management in Large Urban Areas*. Doctoral Dissertation, Department of Civil Engineering, The University of Texas at Arlington, Arlington, TX.

Habibi, H., Arezoo, R. N., Amir, N., Behzad, N., Seo, D.-J., Ranjan, M. & Davis, C. 2016 High resolution flash flood forecasting for the Dallas-Fort Worth metroplex. *Journal of Water Management Modeling* **24**, C401. doi:10.14796/JWMM.C401.

Habibi, H. & Seo, D. J. 2018 Simple and modular integrated modeling of storm drain network with gridded distributed hydrologic model via grid-rendering of storm drains for large urban areas. *J. Hydrol.* **567**, 637–653.

Hashino, T., Bradley, A. A. & Schwartz, S. S. 2002 *Verification of Probabilistic Streamflow Forecasts*. IIHR Report No. 427, The University of Iowa, Iowa City, IA, 125 pp.

Koren, V., Reed, S., Smith, M., Zhang, Z. & Seo, D.-J. 2004 Hydrology laboratory research modeling system (HL-RMS) of the US national weather service. *J. Hydrol.* **291** (3–4), 297–318.

Liu, Y., Brown, J., Demargne, J. & Seo, D. J. 2011 A wavelet-based approach to assessing timing errors in hydrologic predictions. *J. Hydrol.* **397** (3–4), 210–224.

Mutzner, R., Weijs, S. V., Tarolli, P., Calaf, M., Oldroyd, H. J. & Parlange, M. B. 2015 Controls on the diurnal streamflow cycles in two subbasins of an alpine headwater catchment. *Water Resour. Res.* **51** (5), 3403–3418.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I – a discussion of principles. *J. Hydrol.* **10** (3), 282–290.

National Weather Service 2009 *Hydrology Laboratory-Research Distributed Hydrologic Model User Manual v. 3.0.0*.

Norouzi, A. 2016 *Improving Hydrologic Prediction for Large Urban Areas Through Advanced Sensing, High-resolution Modeling and Probabilistic Analysis of Scale-dependent Runoff Response*. Doctoral Dissertation, Department of Civil Engineering, The University of Texas at Arlington, Arlington, TX.

Rafieeinasab, A., Norouzi, A., Kim, S., Habibi, H., Nazari, B., Seo, D. J., Lee, H., Cosgrove, B. & Cui, Z. 2015a Toward high-resolution flash flood prediction in large urban areas – analysis of sensitivity to spatiotemporal resolution of rainfall input and hydrologic modeling. *J. Hydrol.* **531**, 370–388.

Rafieeinasab, A., Norouzi, A., Seo, D.-J. & Nelson, B. 2015b Improving high-resolution quantitative precipitation estimation via fusion of multiple radar-based precipitation products. *J. Hydrol.* **531** (Part 2), 320–336.

Reed, S., Schaake, J. & Zhang, Z. 2007 A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. *J. Hydrol.* **337** (3–4), 402–420.

Seaber, P. R., Kapinos, F. P. & Knapp, G. L. 1987 *Hydrologic Unit Maps: U.S. Geological Survey Water-Supply Paper 2294*, 63 pp.

Seo, D.-J., Seed, A. & Delrieu, G. 2010 Radar-based rainfall estimation. In: *Rainfall: State of Science* (F. Y. Testik & M. Gebremichael, eds). American Geophysical Union, Washington, DC, pp. 79–104.

Seo, D.-J., Kerkez, B., Zink, M., Fang, N., Gao, J. & Yu, X. 2015 iSPUW: A vision for integrated sensing and prediction of urban water for sustainable cities (S. Ravela & A. Sandu, eds). *Dynamic*

Data-Driven Environmental Systems Science, DyDESS 2014. Lecture Notes in Computer Science, Vol. 8964. Springer, Cham, Switzerland. doi:10.1007/978-3-319-25138-7_7.

Smith, M. B., Seo, D.-J., Koren, V. I., Reed, S., Zhang, Z., Duan, Q.-Y., Moreda, F. & Cong, S. 2004 The Distributed Model Intercomparison Project (DMIP): motivation and experiment design. *J. Hydrol.* **298** (1–4), 4–26.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M. M. B., Allen, S. K., Boschung, J., Nauels, A., Xia, Y. & Bex, V. 2013 *Climate Change 2013 – The Physical Science Basis, Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge.

USGS 2010 GWPD 16 – measuring water levels in wells and piezometers by use of a submersible pressure transducer, 6 pp. Available from: https://pubs.usgs.gov/tm/1a1/pdf/GWPD16.pdf.

Wilks, D. S. 2006 *Statistical Methods in the Atmospheric Sciences*. Elsevier Academic Press, San Diego, CA, 648 pp.

First received 12 September 2018; accepted in revised form 7 February 2019. Available online 27 February 2019