Expanding and Enhancing Streamflow Prediction Capability of the National Water Model Using Real-Time Low-Cost Stage Measurements

BONG-CHUL SEO,[a,b] MARCELA ROJAS,[a,b] FELIPE QUINTERO,[a,b] WITOLD F. KRAJEWSKI,[a,b] AND DONG HA KIM[c]

[a] *Iowa Flood Center, The University of Iowa, Iowa City, Iowa*
[b] *IIHR—Hydroscience and Engineering, The University of Iowa, Iowa City, Iowa*
[c] *NOAA/National Water Center, Tuscaloosa, Alabama*

(Manuscript received 16 March 2022, in final form 13 June 2022)

ABSTRACT: This study demonstrates an approach to expand and improve the current prediction capability of the National Water Model (NWM). The primary objective is to examine the potential benefit of real-time local stage measurements in streamflow prediction, particularly for local communities that do not benefit from the improved streamflow forecasts due to the current data assimilation (DA) scheme. The proposed approach incorporates real-time local stage measurements into the NWM streamflow DA procedure by using synthetic rating curves (SRC) developed based on an established open-channel flow model. For streamflow DA and its evaluation, we used 6-yr (2016–21) data collected from 140 U.S. Geological Survey (USGS) stations, where quality-assured rating curves are consistently maintained (verification stations), and 310 stage-only stations operated by the Iowa Flood Center and the USGS in Iowa. The evaluation result from NWM's current DA configuration based on the USGS verification stations indicated that DA improves streamflow prediction skills significantly downstream from the station locations. This improvement tends to increase as the drainage scale becomes larger. The result from the new DA configuration including all stage-only sensors showed an expanded domain of improved predictions, compared to those from the open-loop simulation. This reveals that the real-time low-cost stage sensors are beneficial for streamflow prediction, particularly at small basins, while their utility appears to be limited at large drainage areas because of the inherent limitations of lidar-based channel geometry used for the SRC development. The framework presented in this study can be readily applied to include numerous stage-only stream gauges nationwide in the NWM modeling and forecasting procedures.

KEYWORDS: Gauges; Measurements; Data assimilation; Hydrologic models

1. Introduction

The strategic goal of the National Weather Service (NWS) to build a "Weather Ready Nation" requires not only development of widespread community resilience, but also a better scientific understanding of and preparation for the increasing risks and vulnerability to extreme weather and water events (Uccellini and Ten Hoeve 2019). Accurate hydrologic prediction is a critical component of efforts to accomplish this goal and manage potential threats and disasters from high impact weather, water, and climate events. As part of these endeavors, the National Water Center (NWC) implemented the National Water Model (NWM) into operations in 2016 and continues to provide the upgrades since then. The NWM is a high-resolution distributed hydrologic modeling system built on WRF-Hydro (Gochis et al. 2018) that simulates and forecasts streamflow over the continental United States (e.g., Maidment 2016; Cohen et al. 2018). The operational implementation of the NWM demonstrates an increasing demand for high-resolution hydrologic modeling and forecasting system. The NWM forecasting system enables researchers and forecasters to obtain useful insights regarding detailed aspects of the interactions among modeling elements (e.g., atmosphere, surface, and subsurface) that have not been explored by conventional approaches based on lumped and mesoscale models (e.g., Sorooshian et al. 1993; Koren et al. 2014). This effort on distributed modeling also complements hydrologic guidance at current NWS forecast points and expands forecast capabilities and coverage to ungauged locations (Cosgrove et al. 2015, 2016; Quintero et al. 2020; Seo et al. 2021b).

NWM's high-resolution distributed modeling and forecasting system requires model configuration based on high-resolution topography and forcing datasets: 1) the model basis is the landscape representation using National Hydrography Dataset (NHD) Plus V2 (McKay et al. 2012) along with two separate modeling grids, the land surface model (LSM) and routing grids (e.g., 1.0 and 0.25 km, respectively); and 2) the configuration for analysis and forecast modeling cycles consists of radar- and NWP-based high-resolution forcing data. This high-resolution distributed modeling does not necessarily guarantee improved hydrologic prediction; prediction errors are associated with numerous factors such as uncertainties in forcing inputs, conceptual representation of physical processes, model parameters, and their intricate interactions at relevant space–time scales (e.g., Beven 1993; Carpenter and Georgakakos 2004). As such, it is rather challenging to account for these uncertainty factors individually or calibrate model parameters to accommodate all uncertainties together, particularly at a national scale due to large variability across various hydrologic regions/conditions.

As a simple way to mitigate the effect from these uncertainties on the model predictions, the NWM applies a data assimilation (DA) procedure, known as streamflow "nudging"

Corresponding author: Bong-Chul Seo, bongchul-seo@uiowa.edu

DOI: 10.1175/WAF-D-22-0050.1

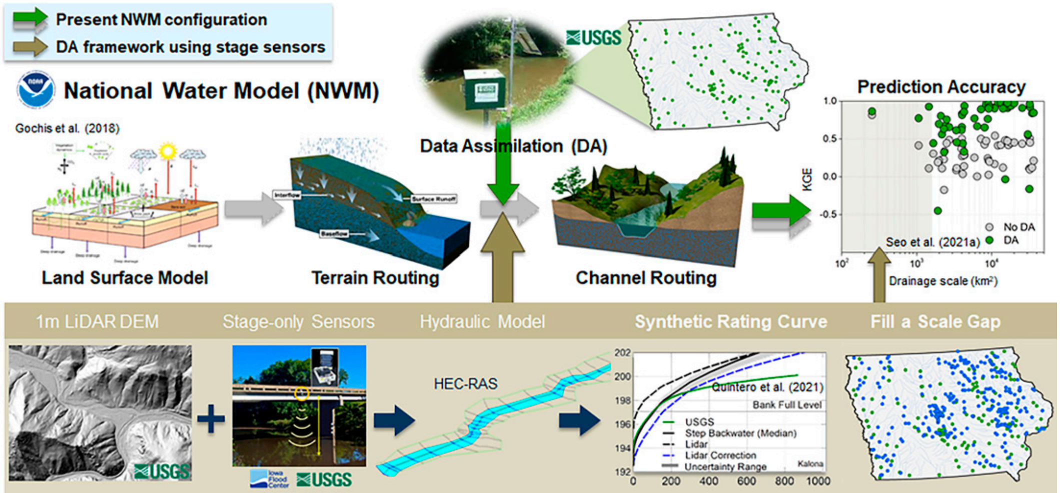

FIG. 1. Schematic overview of the NWM configuration to incorporate real-time stage measurements into its DA procedure.

(Gochis et al. 2018). This is a straightforward and inexpensive (i.e., computationally) approach, which assimilates streamflow observations into the model and replaces model states (at gauging stations) to improve model forecasts as these updated states are used as initial conditions for the NWM forecast cycles. Our recent evaluation of the multiyear NWM simulations demonstrates that the DA approach leads to improved predictions compared to the open-loop simulation (Seo et al. 2021a). However, these improvements are limited only to certain scale of basins, for which streamflow measurements and rating curves are regularly managed by the U.S. Geological Survey (USGS). As evidenced in Seo et al. (2021a), relatively small drainage areas (e.g., <1000 km^2) benefit less from DA due to the scarcity of observed streamflow data at these scales. Accordingly, it would be valuable to expand DA's proven effectiveness to these small scales and fill a scale gap by incorporating real-time low-cost (e.g., stage-only) measurements into the NWM's DA procedure. This can lead to NWM's coverage expansion with enhanced prediction skills because there are approximately thousands of stage-only sensors nationwide (CRS 2019).

The use of stage measurements through DA in hydrologic prediction has been limited due to the requirement of rating curves. Few exceptions include direct forecasting of water levels using stochastic models via an update of model parameters (e.g., Romanowicz et al. 2006; Ziliani et al. 2019). In this study, we develop a framework to include local stage observations in the NWM modeling procedure, which will expand DA's enhanced prediction capability at a wide range of basin scales. Figure 1 illustrates the schematic overview of this study and its individual elements required to build the research framework. The framework uses synthetic rating curves (SRC) derived using a well-established hydraulic model (i.e., HEC-RAS; Brunner 2010), channel geometry retrieved from high-resolution airborne lidar-based topography data, and local stage measurements (e.g., Quintero et al. 2021). The development of SRC comprises several steps to account for uncertainties in the estimation process (e.g., errors in lidar-based topography data and channel roughness). We develop SRC for stage-only stream gauges deployed in part of the Upper Mississippi River and the Missouri River basins centered on Iowa, in which abundant local measurement data from about 280 Iowa Flood Center (IFC) and 30 USGS stage-only stations are available as shown in Fig. 2a. The main objectives of this study are 1) to incorporate local stage observations into the streamflow DA routine implemented in the NWM's channel routing element; and 2) to examine and demonstrate potential benefit of these observations in streamflow prediction.

2. NWM and hydrologic dataset

The NWM is a high-resolution distributed hydrologic modeling and forecasting system built on the WRF-Hydro community model (Gochis et al. 2018). In this section, we briefly describe WRF-Hydro modeling elements and our model implementation for the study domain using the NWM configuration, similar to the one running at the NWS. To implement NWM in our computational environment, we used WRF-Hydro V5.0.3. Because streamflow DA is a key factor of our NWM implementation, we provide detailed descriptions regarding NWM's DA scheme and its governing equations. The hydrologic dataset used for the NWM simulation and analysis includes meteorological forcing products (e.g., precipitation) and streamflow observations. We collected these forcing and streamflow data from 2016 to 2021 and used them for the NWM retrospective analysis. The model forcing data collected for an additional period for August–December in 2015 were used to spin up the model states.

a. NWM configuration

The central modules of the WRF-Hydro modeling system consist of the LSM and water routing elements (i.e., surface, subsurface, and channel routing). These elements are executed on two different spatial resolutions (i.e., 1.0 km for LSM and 0.25 km for grid-based routing) of the NWM

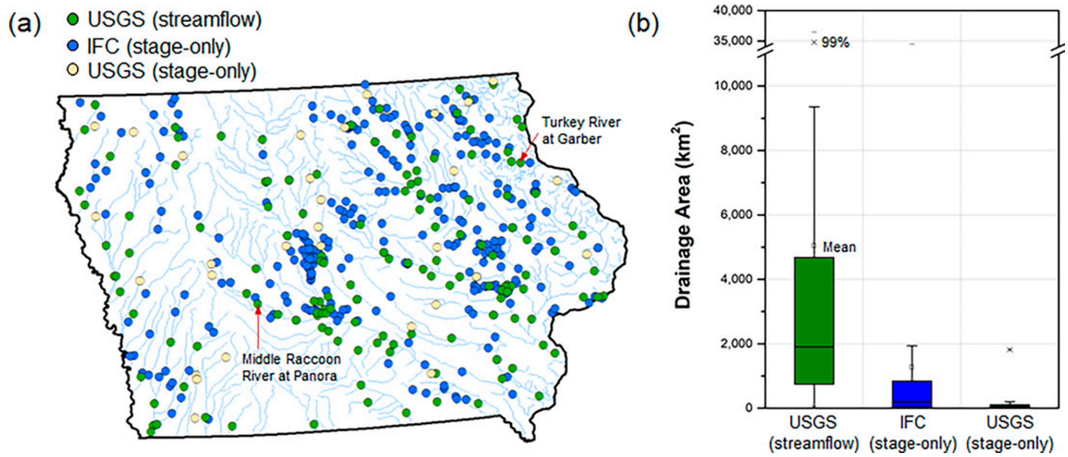

FIG. 2. (a) USGS streamflow and IFC and USGS stage-only gauge locations in Iowa and (b) a box-and-whisker plot describing the distributions of drainage areas covered by each gauge network.

domain grids. We attained the NWM domain grid data for the Iowa domain by using "domain subsetter" (Castronova et al. 2019) offline, with the help of the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI). The model grids and parameters including a user-defined mapping information for catchment aggregation were retrieved from the NWM version 2.0. In the NWM, the Noah Multi-Parameterization (Noah-MP) LSM (Niu et al. 2011; Yang et al. 2011) is selected to represent one-dimensional (vertical) exchange of energy and water fluxes between atmosphere and land surface. The groundwater/base flow module of the NWM uses a conceptual bucket (a groundwater reservoir in each subbasin) model, describing an exponential storage-discharge function (Gochis et al. 2018). The NWM's grid-based routing encompasses diffusive wave surface routing (Downer et al. 2002) and saturated subsurface flow routing (Wigmosta et al. 1994; Wigmosta and Lettenmaier 1999). Apart from the grid-based ones, the NWM adopts Muskingum–Cunge channel routing (e.g., Tang et al. 1999) based on vectorized stream units (i.e., channel links) derived from NHDPlus V2 (McKay et al. 2012) stream reaches.

To reduce the effect of uncertainties in forcing inputs, model physics, and model parameters and thus improve model predictions, the NWM applies simple "nudging" DA within its channel routing procedures (Gochis et al. 2018). This DA scheme updates model initial states for NWM's forecasting cycles (i.e., short-, medium-, and long-range; refer to https://water.noaa.gov/about/nwm) by replacing model computation with streamflow observations at corresponding assimilation points. For a given space (i.e., stream reach) and time domain, the Muskingum–Cunge routing equation is

$$Q_d(t) = C_1 Q_u(t-1) + C_2 Q_u(t) + C_3 Q_d(t-1) + \left(\frac{q_l dt}{D}\right), \tag{1}$$

where Q denotes streamflow discharge at the current (t) and previous ($t-1$) times at the upstream (u) and downstream (d) locations of a stream reach. The terms C_1, C_2, and C_3 are coefficients calculated using routing parameters (see Tang et al. 1999). The terms q_l and D indicate lateral inflow and the wedge storage contribution from lateral inflow. A nudge (N) at an assimilation point is represented by the difference between observed and model estimated streamflow (i.e., model error, $Q_t - \hat{Q}_t$) with a limited temporal interpolation:

$$N = \sum_{t=1}^{n_t} q_t w^2(t)\,(Q_t - \hat{Q}_t) \Bigg/ \sum_{t=1}^{n_t} w^2(t), \tag{2}$$

where q_t denotes a quality coefficient of observed streamflow. In our NWM implementation, we did not consider the uncertainty of observed streamflow (i.e., $q_t = 1$) because streamflow measurement (e.g., rating curve) uncertainties are unknown. The term $w(t)$ indicates a temporal smoothing function that assigns a heavy weight to an observation within 15 min from the current time and sets e-folding time as 2 h (see Gochis et al. 2018 for details). In the current NWM configuration, no spatial smoothing is active for computational efficiency. The calculated nudge at time $t-1$ at the downstream reach, $N_d(t-1)$, is then applied to all upstream and downstream reach terms in Eq. (1):

$$\begin{aligned} Q_d(t) &= C_1[Q_u(t-1) + N_d(t-1)] + C_2[Q_u(t) + N_d(t-1)] \\ &\quad + C_3[Q_d(t-1) + N_d(t-1)] + \left(\frac{q_l\, dt}{D}\right). \end{aligned} \tag{3}$$

Equation (3) is the current form of nudging formula used in the NWM, and the nudge added to all upstream and downstream terms reduces discontinuity between upstream and downstream reaches. However, the nudge applied to the upstream reach does not contribute to the model output values of the upstream reach and is used only when solving downstream discharge. Consequently, the nudging effect is propagated downstream only through this Muskingum–Cunge channel routing. Note that we excluded reservoir routing in our NWM configuration to simplify the model implementation.

We ran the model using the hydrologic parameter set retrieved from the NWM 2.0 with no further calibration. Parameter calibration in LSM and surface/subsurface routing tends to be less impactful on DA because observed streamflow at an assimilation point and channel flow routing along the downstream river reach are the primary factors determining streamflow discharge.

b. *Forcing products*

We used two meteorological forcing products to drive the NWM: 1) Multi-Radar Multi-Sensor (MRMS) precipitation estimates (J. Zhang et al. 2016); and 2) North America Land Data Assimilation System (NLDAS) dataset (e.g., Xia et al. 2012). MRMS provides a wide range of products (e.g., severe weather, transportation, and precipitation) for hazardous weather monitoring and forecasting. The MRMS system integrates multiple radar data with surface and upper air observations, lightning detection, satellite observations, and numerical weather prediction model analysis. MRMS includes a suite of quantitative precipitation estimation (QPE) products with 0.01° spatial resolution and a variety of temporal (e.g., 2 min and 1, 3, 6, 12, 24, 48, and 72 h) resolutions. In this study, we used an hourly MRMS QPE product that contains a bias correction using rain gauge data. The real-time MRMS product is available from the National Centers for Environmental Prediction via a web protocol (https://mrms.ncep.noaa.gov/data/).

The Noah-MP LSM requires additional meteorological forcing such as incoming shortwave and longwave radiation, specific humidity, air temperature, surface pressure, and near-surface wind components. We retrieved these forcing data from the hourly NLDAS dataset at a resolution of 0.125 decimal degrees. The NLDAS dataset is available from the National Aeronautics and Space Administration's Goddard Earth Sciences Data and Information Services Center (https://disc.gsfc.nasa.gov/datasets/). We collected both MRMS and NLDAS data for the study period from 2015 to 2021 and resampled them onto the 1-km LSM grid to drive the NWM (i.e., Noah-MP).

c. *Streamflow and stage observations*

For streamflow DA in the NWM, we used streamflow observations from 140 U.S. Geological Survey (USGS) stations and stage measurements from 280 Iowa Flood Center (IFC) and 30 USGS stage-only sensors in Iowa (see Fig. 2). The use of stage measurements aims to complement the current coverage of the USGS streamflow stations included in the NWM DA routine and expand DA's proven performance (e.g., Seo et al. 2021a) to relatively small-scale basins and local communities. Hereafter, we use "USGS stations" to indicate locations where quality-assured rating curves and discharge data are provided and distinguish them from "USGS stage-only stations."

We collected quality-controlled streamflow data with a 15-min interval from 140 USGS stations for the period from 2016 to 2021. These streamflow data enable streamflow DA at all corresponding assimilation locations, as well as the evaluation of DA effect propagated to their downstream stations. The streamflow data were acquired by transforming recorded river stage into discharge using well-defined rating curves estimated for each individual station. The USGS has developed the rating curves through periodic stage-discharge measurements, especially during low- and high-flow events. As we discussed in section 2a, we do not account for the rating curve uncertainty and its effect on the DA evaluation in this study.

During the last decade, IFC has provided streamflow predictions over the entire state of Iowa (Krajewski et al. 2017). IFC has deployed bridge-mounted stream sensors to monitor streams and creeks near Iowa communities for which IFC makes streamflow predictions to complement NWS's forecasts issued at the limited number of forecast points. The number of sensors is consistently increasing upon requests from local agencies and communities. These sensors measure water elevation every 15 min using an ultrasonic sensor and transfer the observations to a data server via cell phone communication (Kruger et al. 2016). The sensors are autonomous: i.e., equipped with a battery recharged by a solar panel, a datalogger, a GPS receiver, and a cell modem to relay the data. IFC has developed a procedure to build SRC (Quintero et al. 2021) to incorporate these data into model-based streamflow prediction. A brief description regarding the development of SRC is provided in section 3. We collected IFC's stream sensor data (stage and discharge) for the study period. The available data period for individual stations varies according to their deployment time. The real-time data of IFC stream sensors can be visualized via the web-based Iowa Flood Information System.

While most USGS stations provide discharge observations, some remain river stage-only gauges because of limited funds available to regularly develop and maintain rating curves. The stations are equipped with modern electronic stage sensors and water-level recorders to collect river stages. We collected data from 30 USGS stage-only stations in Iowa for the year of 2021 and developed SRC (these data are available online for the latest 120 days only). The UGGS streamflow and stage-only data are available from the National Water Information System (https://waterdata.usgs.gov). While we assimilated discharge data obtained from SRC into the NWM simulations, we did not use them for the DA evaluation due to uncertainty contained in SRC, degree of which is site specific.

3. Methodology

In this section, we describe approaches to develop SRC and evaluate the effect of streamflow DA. These approaches are the essential factors needed to accomplish the main objectives of this study discussed in section 1.

a. *SRC*

The USGS's standard to develop and maintain the stage-discharge rating curve entails frequent site visits to acquire direct discharge and stage measurements for various river flow conditions. These site visits need to partially account for occasional changes in channel bed (e.g., sedimentation and erosion). Although this standard procedure results in the most accurate stage–discharge relationship, it is also prohibitively

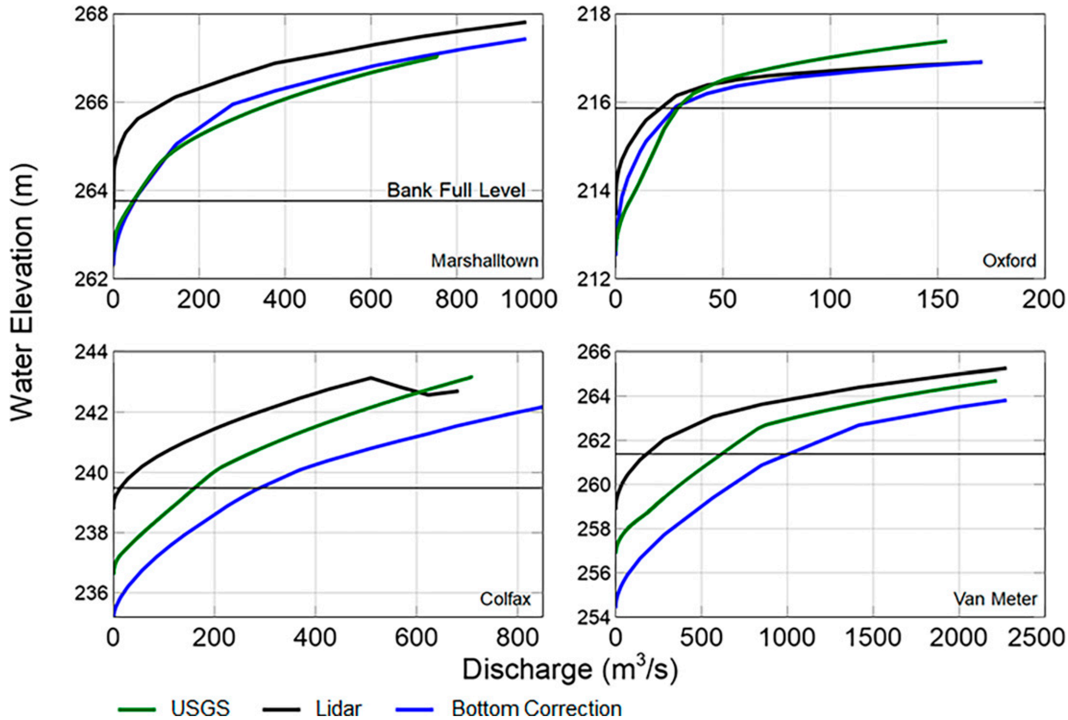

FIG. 3. A preliminary evaluation of SRC derived from the lidar data (with and without the bottom corrections) compared to the USGS rating curves at four selected sites in Iowa.

expensive. IFC in partnership with the Iowa Silver Jackets Program (2016) developed a methodology economically feasible to extend the capabilities of stage-only sensors. This method allows the translation of stage readings into discharge estimates by using SRC derived from hydraulic models, particularly HEC-RAS (Quintero et al. 2021). The implementation of this approach requires one visit to the sites to collect brief information on the channel section geometry, sensor location, and its adjacent sections upstream and downstream.

We used the lidar-based topography data to delineate relevant channel sections required for a hydraulic model. Because lidar beam cannot penetrate the water column, our methodology includes a procedure to correct the elevation of channel bottom based on the stage measurements during dry seasons. The visit to the sites of interest also allows us to measure and achieve the bottom elevation for a better representation of channel cross sections. Figure 3 shows a brief evaluation result of SRC developed at four locations where USGS and IFC sensors are collocated; reference streamflow data exist at these USGS stations. In Fig. 3, the correction of channel bottom leads to an improved rating curve estimation at Marshalltown and Oxford (top panel), while the correction at the other two sites (bottom panel) is not as efficient as the ones shown in the top panel. The uncertainty of SRC is caused by various factors, and we explore how this uncertainty results in streamflow DA in this study. However, the investigation on the uncertainty factors themselves in the rating curves is not within the scope of this study. For detailed development processes of SRC and their extensive evaluation, refer to Quintero et al. (2021).

Following the same methodology, we also developed SRC for the 30 USGS stage-only stations. To obtain the topographic information at these locations, we visited each site and collected a cross section downstream of the bridge, the slope of the channel, and other information useful to set up the hydraulic model to complement the lidar data. However, the data availability for these sites is limited to 1 year only and their coverage area is very small as shown in Fig. 2b.

b. Evaluation

We simulated the NWM for the 7-yr period with streamflow DA and open-loop (No DA) modes. We used the first year (i.e., 2015) simulation to spin up the model states for the remaining analysis period of 6 years. In the DA mode, we assimilated streamflow discharge in two ways: 1) NWM's current DA configuration using the streamflow data collected only from the 140 USGS stations (referred to as "USGS-only" and not including data from USGS stage-only sensors); and 2) new DA configuration using all combined network gauge data (referred to as "Combined" and the number of stations is about 450 in total). These separate simulations were designed to evaluate potential benefit of stage sensors in streamflow prediction by comparing the results of Combined with those of USGS-only at the evaluation stations that possess assimilation points upstream. In the analyses, we limited the evaluation period to relatively warm months during April–October in each year. Streamflow during winter and early

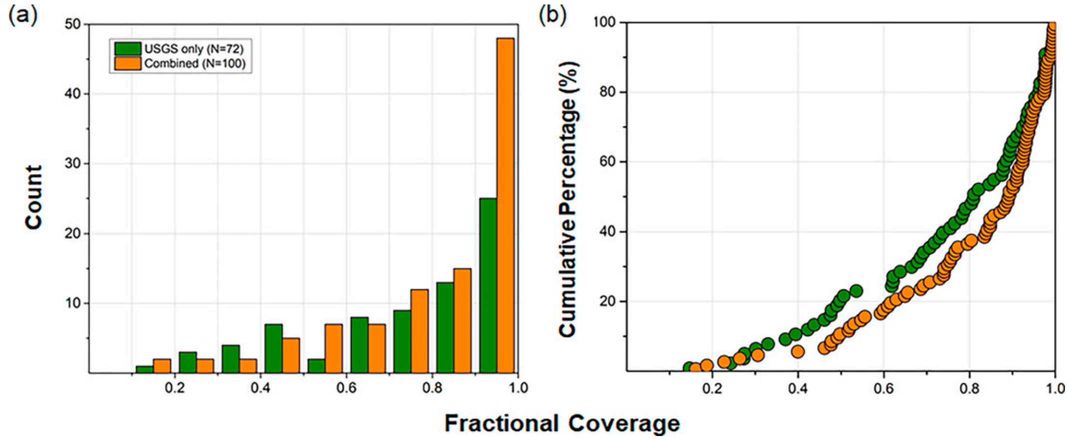

FIG. 4. (a) Distributions and (b) cumulative percentages of the fractional coverage for different DA configurations (USGS only and Combined). The same colors are applied to (a) and (b) to compare the two configurations.

spring in Iowa is primarily affected by frozen ground and snowmelt combined with winter precipitation, the observations/estimations for all of which are typically subject to large uncertainties (e.g., Seo et al. 2015; J. L. Zhang et al. 2016; Souverijns et al. 2017).

For a meaningful DA assessment at the evaluation points, the model-computed streamflow propagated from upstream assimilation points should be compared with the observed streamflow at the location. However, this is infeasible with the current setup of NWM's DA routine: the execution of DA replaces model-simulated streamflow with the observed values at all gauging stations where valid observations are fed into the model. This setup is challenging for DA assessment because most evaluation points also become assimilation points for their downstream reaches, implying that observed streamflow at the evaluation points overwrites model values propagated from upstream reaches. For that reason, we retrieved the model-simulated values at the closest upstream links (one or multiple) directly connected with the corresponding evaluation point to evaluate the effects of DA. The underlying assumption here is that the effects of channel routing and later inflow along a stream segment (link) to which the evaluation point belongs are negligible. This assumption was validated in Seo et al. (2021a) through an experiment using two pilot basins in Iowa.

To quantitatively measure the hydrologic prediction skills of DA, we use four performance evaluation metrics: 1) relative volume error (RE_V); 2) relative peak error (RE_{Q_p}); 3) peak timing error (E_{t_p}); and 4) Kling–Gupta efficiency (KGE; Gupta et al. 2009). The formulas of these metrics are provided:

$$\mathrm{RE}_V = \frac{V_{\mathrm{NWM}} - V_{\mathrm{obs}}}{V_{\mathrm{obs}}} \times 100\%, \tag{4}$$

$$\mathrm{RE}_{Q_p} = \frac{Q_{p,\mathrm{NWM}} - Q_{p,\mathrm{obs}}}{Q_{p,\mathrm{obs}}} \times 100\%, \tag{5}$$

$$E_{t_p} = t_{p,\mathrm{NWM}} - t_{p,\mathrm{obs}}, \tag{6}$$

$$\mathrm{KGE} = 1.0 - \sqrt{(\rho - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2}, \tag{7}$$

where V, Q_p, and t_p denote total water volume (m^3), peak discharge ($\mathrm{m}^3\ \mathrm{s}^{-1}$), and peak time (h), respectively, obtained from model simulations (NWM) and observations (obs) in an annual basis (April–October). KGE is represented as a function of correlation (ρ), the ratio of standard deviation (α), and the ratio of mean (β) between simulated and observed streamflow. This metric is often used as an overall performance indicator describing the predictive power of hydrologic models to avoid some deficiencies (e.g., skewness and variability due to sampling) of Nash–Sutcliffe efficiency (NSE; Nash and Sutcliffe 1970). While there are many discussions regarding the advantages and disadvantages of KGE and NSE (e.g., Lamontagne et al. 2020; Clark et al. 2021), we decided to use KGE because KGE is easily understood by examining its contributing components (i.e., ρ, α, and β).

4. Results

The main factor impacting the performance of DA is the fractional coverage of upstream assimilation points to the entire contributing area delineated by the downstream evaluation point (e.g., Seo et al. 2021a). As illustrated in Fig. 4, the fractional coverage within the study domain improved considerably by including stage-only sensors in the DA framework. The number of evaluation points that require at least an assimilation point upstream increased from 72 to 100, implying that 28 farther upstream areas benefit additionally from DA using the stage-only sensors. Figure 4 shows the change of fractional coverage distribution and cumulative percentage between the two DA configurations discussed in section 3b. We recognize from Fig. 4a that the number of evaluation points decreases at a lower coverage range (e.g., 0.2–0.5) and increases oppositely at a higher range (e.g., 0.5–1.0) after including stage sensors. This is due to the effect of additional coverage added by the stage sensors that monitor upstream tributaries and local communities. The cumulative percentage

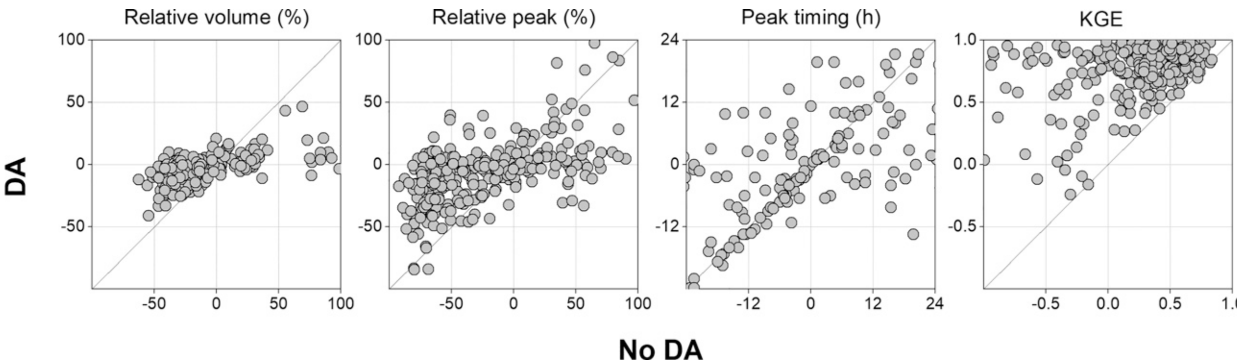

FIG. 5. Performance comparison of the DA simulation results with those of the open-loop simulation (No DA). Each circle corresponds to an annual estimate at an individual evaluation location, and all 6-yr (2016–21) results are presented in the figure.

in Fig. 4b represents an "exceedance" proportion (e.g., similar to exceedance probability) for a given coverage value. For example, about 60% of areas in the combined network coverage have a fractional coverage value greater than 0.8, while there are about 50% of those with USGS-only for the same coverage threshold.

a. *DA with USGS-only*

The results of DA simulation with USGS-only are evaluated against those of open-loop simulation (No DA). We calculated the four metrics defined in Eqs. (4)–(7) and present them in Fig. 5. All these metrics were estimated in an annual basis because the data availability of upstream assimilation points was not consistent year to year, and thus the fractional coverage may vary year to year for some locations. Each circle in Fig. 5 indicates an annual estimate of the metrics at an evaluation point, and we placed all 6-yr (2016–21) results together in Fig. 5. To identify an annual peak from the simulated NWM time series, we defined a scale-dependent time span around the annual peak observed from the USGS streamflow data. The errors in peak (RE_{Q_p}) and its timing (E_{t_p}) are then estimated from an NWM simulated peak specified within the time span to avoid cases where a model simulation generates an annual peak from an entirely different event (see e.g., Seo et al. 2021a). We determined the time span by estimating time of concentration (i.e., the travel time along the longest stream network within a specific drainage area) or 5 days, whichever is smaller.

It is obvious that DA improves in estimating runoff volume and peak discharge, compared to the open-loop simulation. The circles with DA in the relative volume error are densely placed around the no error (0%) line, whereas the ones with open-loop are widely distributed within a range from −50% to 100%. The error in relative peak shows a similar tendency to relative volume, but with a little higher variability and more underestimations. We speculate that DA's slight underestimations in total runoff volume and peak discharge might be the effect of lateral inflow ignored along a stream segment to which the evaluation points belong. The estimation of peak timing seems to be quite challenging for both DA and open-loop simulations: the circles are broadly scattered, and many of them tend to align along the one-to-one line, meaning that the results of DA and open-loop are very close. We note that there is a smaller number of circles in the peak timing than those in other metrics because we present the error only within a 24-h time span. The overall performance of streamflow simulations represented by KGE demonstrates the superiority of DA over open-loop: almost all circles are placed above the one-to-one line that indicates equal-performance of the two simulations.

In Fig. 6, we reorganize the results shown in Fig. 5 to demonstrate the scale-dependent performance of NWM simulations. As demonstrated in Fig. 6, DA tends to improve prediction skills as drainage scale becomes larger. This scale dependence of DA looks prominent, particularly with the relative volume error and KGE: 1) the variability of volume error gradually decreases and the mean of the error approaches no bias with increasing scale; and 2) KGE increases sharply and approaches the unity as drainage scale increases. We observe some exceptions at the largest scale with unexpected dispersion in the peak error and KGE. This dispersion might be the effect from the misrepresentation of catchment aggregation and stream networks implemented in the NWM, discussed in Rojas et al. (2020).

b. *DA with Combined*

We combined data from all stream gauges together shown in Fig. 2 and forced them into the NWM DA data stream to evaluate the utility of stage sensors in the model prediction. To distinguish the new DA effect of stage sensors from the existing one by USGS-only, we specify three types of stream gauge network configuration (referred to as "NC") that may affect DA performance, as illustrated in Fig. 7. NC1 demonstrates unprecedented value of stage sensors in the DA framework. While the DA effect of NC1 can be evaluated only at 28 locations as shown in Fig. 4, there are much more NC1 areas that cannot be evaluated in the study domain because it requires a USGS station at their direct downstream reach. In NC2, the stage sensors provide additional observation coverage to the evaluation point downstream. In the study domain, the USGS stations (e.g., a USGS assimilation point in NC2) typically monitor the main stream reach of large areas, and thus their fractional coverage tends to be higher than that of stage sensors that are mostly located in small creeks and

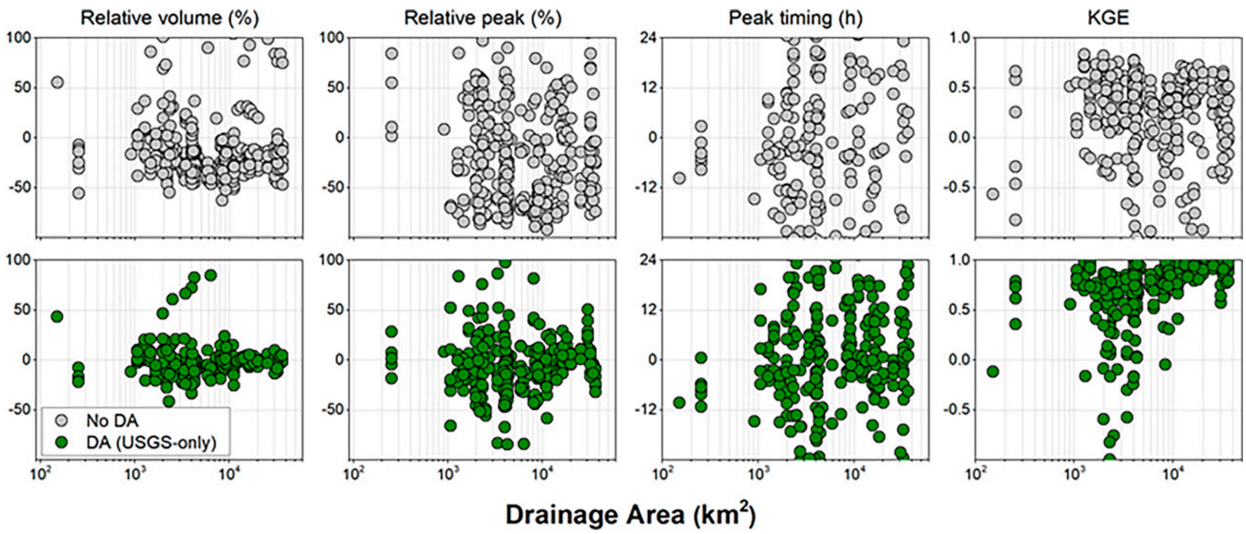

FIG. 6. Performance comparison between the DA and open-loop (No DA) simulations with respect to drainage scale.

streams. In NC3, the stage sensors are located between the USGS assimilation (upstream) and evaluation (downstream) points. DA at the stage sensors replaces the effect from the USGS assimilation points with the stage sensor data; the model overwrites the result propagated from the upstream USGS streamflow observations. The number of NC2 and NC3 gauges in the study domain are 12 and 34, respectively.

We calculated the four metrics for the three NC cases and present them in Fig. 8. The circles in Fig. 8 are annual estimates for the entire 6-yr period. Since the circles seem more scattered than those in Fig. 5, we expose the different colored zones in the volume and peak errors to readily distinguish the superiority of each DA configuration. The zones shaded by light green and orange colors indicate that USGS-only and Combined performs better than the other, respectively.

In fact, the DA effect in NC1 is actually compared with the result of open-loop simulation because all these evaluation points were the farthest upstream stations in the USGS network, and thus no DA result was available with USGS-only. It is not clear from Fig. 8 whether DA (Combined) for NC1 performs better than open-loop does in estimating total water volume and peak discharge. On the other hand, the majority of evaluation points show significant improvement in KGE as most circles are placed over the one-to-one line. We note that the annual (April–October) KGE estimated in this study may include a sampling issue (e.g., variability and skewness) documented in Lamontagne et al. (2020). One can reduce this sampling issue by estimating a KGE value for the time series of entire 6-yr data. However, each deployment time of IFC stage sensors varies, and the data availability of upstream assimilation points is not consistent year to year for the Combined DA configuration. This fact made harder to estimate a consistent and robust performance metric value using the entire time series over the 6-yr period. Although we attempted to calculate lognormal NSE (LNSE) with the same annual basis, we decided to use KGE because its performance is understandable by interpreting its constituting components. The use of LNSE may raise a similar issue regarding sampling biases (Lamontagne et al. 2020).

The fractional coverage in NC2 slightly increased by the additional areas monitored by the stage sensors. We do not observe meaningful changes from Fig. 8, except two cases marked

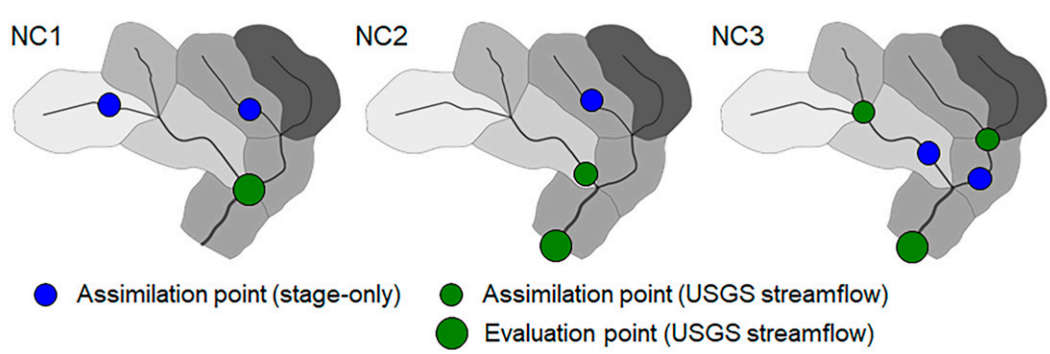

FIG. 7. Three example types of stream gauge network configuration when adding stage-only sensors on the current USGS (streamflow) network. NC stands for network configuration.

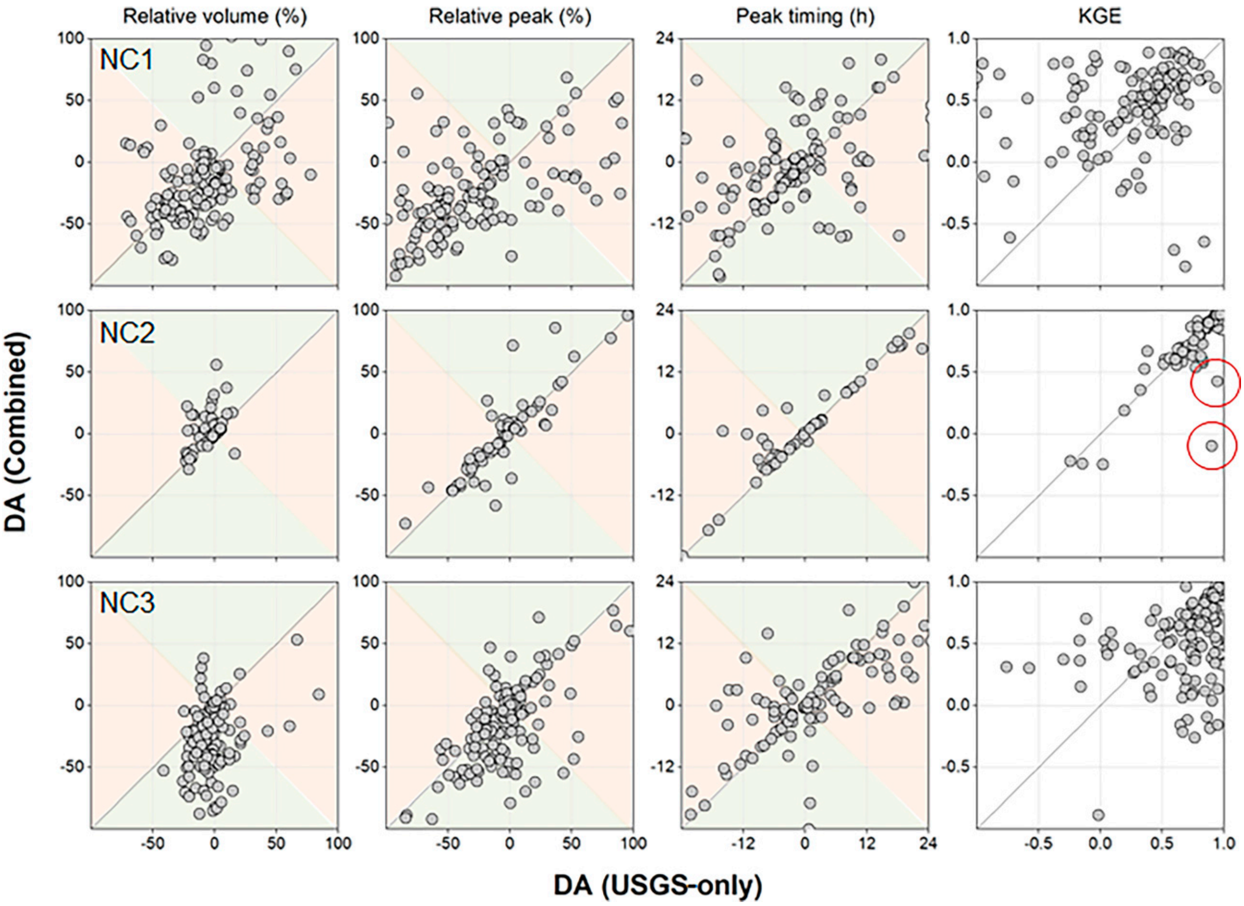

FIG. 8. Performance comparison of DA (Combined) with DA (USGS-only) for the different stream gauge network configuration (NC). The two different colored zones represent the superiority of each DA configuration (light green: USGS-only; and light orange: Combined).

by the red circles, probably because a small portion of coverage combined cannot change major features of streamflow. We investigated the two exceptions indicated in NC2 that show a significant performance drop in KGE. Figure 9 shows observed and simulated hydrographs at the two exceptional locations (the Middle Raccoon River at Panora and the Turkey River at Garber) for the period in 2020. In both locations, DA (Combined) shows systematic shifting that causes a considerable change in bias factor β in Eq. (7) while the other two terms in Eq. (7) remain at a similar range. This leads to a significant drop in KGE. The observed systematic overestimations seem to be an issue with SRC developed for these specific sites using lidar information in the absence of a detailed survey that describes fairly the local characteristics of the channel. Findings from Quintero et al. (2021) suggest that a better topographic description of the channel alleviates this problem.

The evaluation points included in NC3 cover relatively large areas. As we expected, the DA (Combined) performance measured by KGE becomes worse in most locations. In these NC3 locations, streamflow data estimated using SRC replace the assimilation effect propagated from the USGS stations upstream. Therefore, the quality of SRC directly affects DA performance in the NC3 locations, and the derivation of lidar-based rating curves is challenging for these large drainage areas.

We reorganize the results of NC1 in Fig. 8 and present them with respect to drainage scale in Fig. 10. NC2 and NC3 were excluded in this scale-dependent analysis because the benefit of DA using the stage sensors in NC2 and NC3 is not meaningful. As demonstrated in Fig. 10, NC1 includes many small-scale areas that were not covered by the USGS stream gauges (see Fig. 6 for a comparison of scales covered). In Fig. 10, we do not detect any visible scale dependence on the prediction performance from each DA and open-loop simulation. Instead, the results of DA (combined) show better performance with the errors in peak and KGE particularly at scales smaller than 1000-km^2 areas.

5. Conclusions

This study builds on the preliminary performance evaluation of NWM DA documented in Seo et al. (2021a). In this study, we demonstrate a framework to expand DA's proven performance to relatively small-scale areas and improve

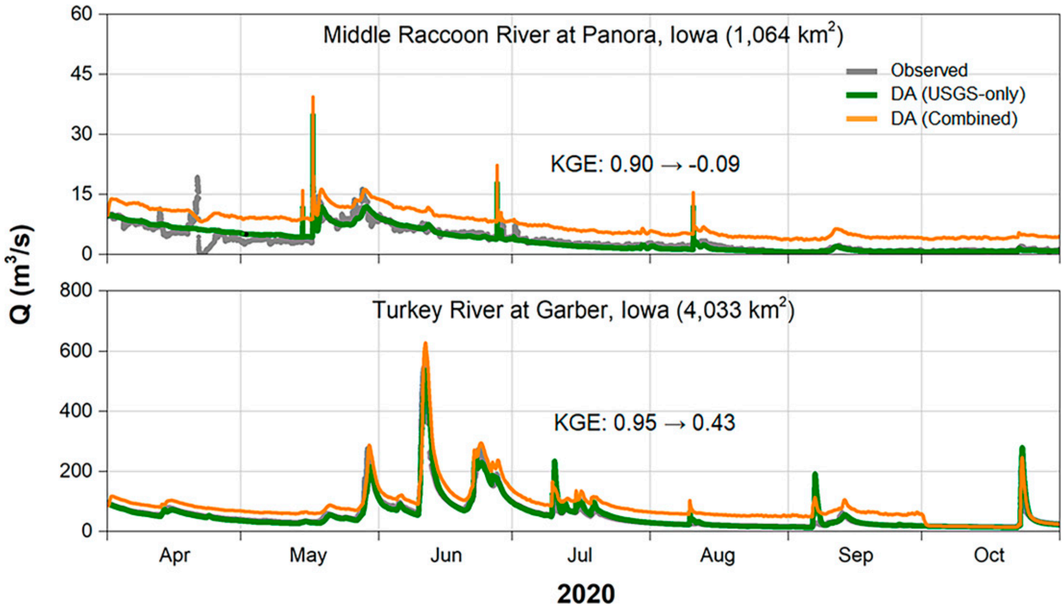

FIG. 9. Observed and simulated hydrographs at the two locations marked by the solid red circles in Fig. 8. The two locations are indicated in Fig. 2a.

streamflow predictions by including low-cost stage measurements in NWM's DA routine. We tested our hypothesis ("incorporation of local stage measurements into the NWM modeling procedure will expand and improve the current prediction capability") for a study domain, part of the Upper Mississippi River and the Missouri River basins centered on Iowa. For streamflow DA and its performance assessment, we used data from 140 USGS gauging stations, 280 IFC stream stage sensors (Kruger et al. 2016), and 30 USGS stage-only stations in Iowa for the period from 2016 to 2021. To derive streamflow discharge from the stage measurements collected from 310 stage-only sensors, SRC were developed based on hydraulic modeling using channel geometry information retrieved from lidar-based elevation data (see Quintero et al. 2021).

The simulations using NWM's current DA configuration ("USGS-only") clearly showed that DA improves streamflow prediction skills at the 72 evaluation points. DA led to improved runoff volume and peak discharge with smaller bias and variability, compared to the open-loop simulation (No DA). The overall performance measure, KGE, shown in Fig. 5 confirmed the superiority of DA to open-loop; the

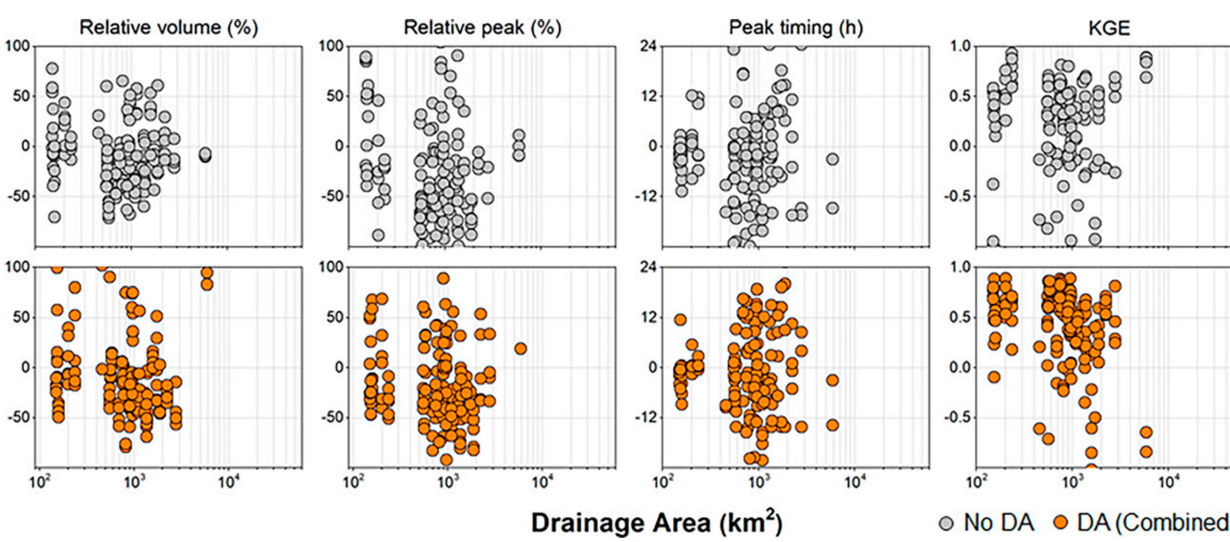

FIG. 10. Performance comparison for NC1 between the DA (Combined) and open-loop (No DA) simulations with respect to drainage scale.

superior performance tends to increase as drainage scale becomes larger.

We combined all available streamflow data including the ones derived from stage measurements and fed them into the DA stream within the NWM channel routing process. To assess the actual benefit of stage sensors in streamflow prediction, we defined three types of gauge network configuration (NC) that may affect DA performance. These NC types illustrated in Fig. 7 distinguish the DA effects of the new configuration (Combined) from those of NWM's current one. The assessment results at 100 evaluation points revealed that the stage sensors deployed at the upstream reaches of uppermost USGS stations (NC1) improve streamflow prediction. We also observed that the stage sensors located in between USGS stations (NC3) tend to degrade prediction performance due to the inherent limitation of lidar-based SRC over relatively large drainage area. These findings confirm that our hypothesis is valid, and low-cost stage sensors are particularly useful for streamflow prediction at smaller-scale basins.

The quality of SRC is a key element to determine the performance of streamflow prediction generated by DA using the stage sensor data, as presented in Fig. 9. The estimation of SRC is affected by a variety of uncertainty factors, and we demonstrated the effect from one of the most significant factors (i.e., bottom of channel geometry) on the quality of rating curves. In the future, we hope to extend the comparison shown in Fig. 3 to many more USGS gauging stations where quality-assured rating curves are maintained. Perhaps, one way that may mitigate the uncertainty effect of SRC on the DA predictions (e.g., for NC3) is to apply a weighing scheme to combine the measurements from stage sensors and propagation from USGS assimilation points upstream. The weights might be determined through a consistency check by comparing streamflow time series at previous times and fractional coverage between an evaluation point and the two assimilation points upstream.

While we developed SRC for the USGS stage-only stations, the DA effect from these sensors was quite limited because of their small coverage areas (see Fig. 2) and short data period (i.e., one year). The results from stage sensors presented in this study were mostly contributed by the ones operated by IFC. IFC has made a consistent effort to develop (for new sites) and improve the SRC (Krajewski et al. 2017; Quintero et al. 2021). To more effectively utilize the data from the USGS stage-only stations, we think that a similar degree of effort and investment (e.g., data quality control and maintenance) is required by the USGS. The framework demonstrated in this study will increase the utility of stage-only sensors, which have had limited potential for hydrologic applications, by developing SRC. Hopefully, this framework would be readily applied to include thousands of stage-only stream gauges nationwide (e.g., CRS 2019) operated by the USGS or local agencies in the NWM modeling and forecasting procedures.

In this study, we were able to evaluate the benefit of stage-only sensors for streamflow prediction at the limited number of USGS stations. However, we think that the effect and actual benefit of these sensors are much broader and more significant than what we presented in this study. For example, Fig. 11 illustrates the Iowa communities and river reaches monitored by the current USGS stations and IFC sensors. The extended coverage offered by IFC sensors will complement NWS's present forecasting domain and provide useful hydrologic guidance for local communities that are not currently benefitted from NWS's streamflow forecasts with the current DA method.

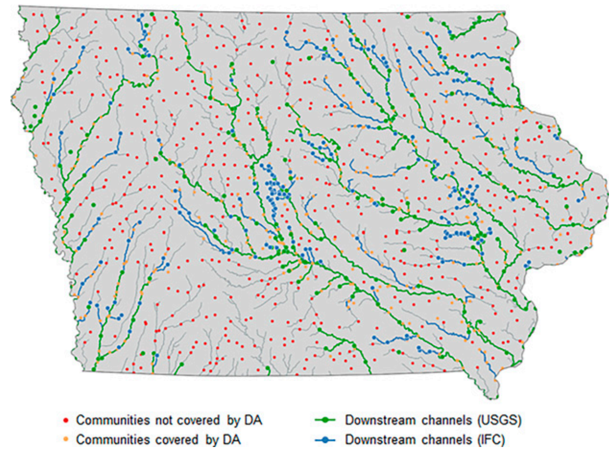

FIG. 11. A map showing about 1000 Iowa communities located near streams and rivers influenced by the current assimilation (USGS) points and the new assimilation points (stage-only).

Acknowledgments. This work was supported by the Iowa Flood Center and the NWC Cooperative Project program under UCAR Grant SUBAWD002640. The authors are grateful to Dr. Anthony Castronova at CUAHSI for providing the NWM 2.0 domain data and Kaleb Young at the University of Iowa for managing the stage data to develop SRC. We also would like to thank Jon Nania at the USGS Central Midwest Water Science Center for useful discussions and site information on the USGS stage-only stations required for the SRC development.

Data availability statement. The meteorological forcing (MRMS and NLDAS) and streamflow data including stage-only observations are publicly available through corresponding URLs provided in section 2.

REFERENCES

Beven, K., 1993: Prophecy, reality and uncertainty in distributed hydrological modeling. *Adv. Water Resour.*, **16**, 41–51, https://doi.org/10.1016/0309-1708(93)90028-E.

Brunner, G. W., 2010: HEC-RAS River analysis system: Hydraulic reference manual, version 4.1. U.S. Army Corps of Engineers Tech. Rep. CPD-69, 417 pp.

Carpenter, T. M., and K. P. Georgakakos, 2004: Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrologic model. *J. Hydrol.*, **298**, 202–221, https://doi.org/10.1016/j.jhydrol.2004.03.036.

Castronova, A. M., and Coauthors, 2019: Improving access to continental-scale hydrology models for research and education—A subsetting adventure. *2019 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract H43I-2146, https://agu.confex.com/agu/fm19/meetingapp.cgi/Paper/619688.

Clark, M. P., and Coauthors, 2021: The abuse of popular performance metrics in hydrologic modeling. *Water Resour. Res.*, **57**, e2020WR029001, https://doi.org/10.1029/2020WR029001.

Cohen, S., S. Praskievicz, and D. R. Maidment, 2018: Featured collection introduction: National Water Model. *J. Amer. Water Resour. Assoc.*, **54**, 767–769, https://doi.org/10.1111/1752-1688.12664.

Congressional Research Service, 2019: U.S. Geological Survey (USGS) streamgaging network: Overview and issues for Congress. CRS Rep. R45695, 31 pp., https://crsreports.congress.gov/product/pdf/R/R45695.

Cosgrove, B., and Coauthors, 2015: Hydrologic modeling at the National Water Center: Operational implementation of the WRF-hydro model to support National Weather Service hydrology. *2015 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract H53A-1649, https://agu.confex.com/agu/fm15/meetingapp.cgi/Paper/66151.

——, and Coauthors, 2016: An overview of the National Weather Service National Water Model. *2016 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract H42B-05, https://agu.confex.com/agu/fm16/meetingapp.cgi/Paper/181304.

Downer, C. W., F. L. Ogden, W. D. Martin, and R. S. Harmon, 2002: Theory, development, and applicability of the surface water hydrologic model CASC2D. *Hydrol. Processes*, **16**, 255–275, https://doi.org/10.1002/hyp.338.

Gochis, D., and Coauthors, 2018: The NCAR WRF-Hydro modeling system technical description (version 5.0). NCAR Tech. Note, 107 pp., https://ral.ucar.edu/sites/default/files/public/WRF-HydroV5TechnicalDescription.pdf.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, **377**, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003.

Iowa Silver Jackets Program, 2016: Iowa Bridge Sensor Demonstration Project Phase I and Phase II Executive Summary Report. U.S. Army Corps of Engineers Tech. Rep., 11 pp., https://www.mvr.usace.army.mil/Portals/48/docs/FRM/Iowa%20Bridge%20Sensor%20Demonstration%20Project%20Executive%20Summary.pdf?ver=2017-05-05-125939-557.

Koren, V., M. Smith, and Z. Cui, 2014: Physically-based modifications to the Sacramento Soil Moisture Accounting model. Part A: Modeling the effects of frozen ground on the runoff generation process. *J. Hydrol.*, **519**, 3475–3491, https://doi.org/10.1016/j.jhydrol.2014.03.004.

Krajewski, W. F., and Coauthors, 2017: Real-time flood forecasting and information system for the State of Iowa. *Bull. Amer. Meteor. Soc.*, **98**, 539–554, https://doi.org/10.1175/BAMS-D-15-00243.1.

Kruger, A., W. F. Krajewski, J. J. Niemeier, D. L. Ceynar, and R. Goska, 2016: Bridge mounted river stage sensors (BMRSS). *IEEE Access*, **4**, 8948–8966, https://doi.org/10.1109/ACCESS.2016.2631172.

Lamontagne, J. R., C. A. Barber, and R. M. Vogel, 2020: Improved estimators of model performance efficiency for skewed hydrologic data. *Water Resour. Res.*, **56**, e2020WR027101, https://doi.org/10.1029/2020WR027101.

Maidment, D. R., 2016: Conceptual framework for the National Flood Interoperability Experiment. *J. Amer. Water Resour. Assoc.*, **53**, 245–257, https://doi.org/10.1111/1752-1688.12474.

McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Rea, 2012: NHDPlus Version 2: User guide. U.S. Environmental Protection Agency, 181 pp., ftp://ftp.horizon-systems.com/NHDplus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf.

Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models Part I—A discussion of principles. *J. Hydrol.*, **10**, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6.

Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multi-parameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116**, D12109, https://doi.org/10.1029/2010JD015139.

Quintero, F., W. F. Krajewski, B.-C. Seo, and R. Mantilla, 2020: Improvement and evaluation of the Iowa flood center hillslope link model (HLM) by calibration-free approach. *J. Hydrol.*, **584**, 124686, https://doi.org/10.1016/j.jhydrol.2020.124686.

——, and Coauthors, 2021: Development of synthetic rating curves: Case study in Iowa. *J. Hydrol. Eng.*, **26**, 05020046, https://doi.org/10.1061/(ASCE)HE.1943-5584.0002022.

Rojas, M., F. Quintero, and W. F. Krajewski, 2020: Performance of the national water model in Iowa using independent observations. *J. Amer. Water Resour. Assoc.*, **56**, 568–585, https://doi.org/10.1111/1752-1688.12820.

Romanowicz, R. J., P. C. Young, and K. J. Beven, 2006: Data assimilation and adaptive forecasting of water levels in the river Severn catchment, United Kingdom. *Water Resour. Res.*, **42**, W06407, https://doi.org/10.1029/2005WR004373.

Seo, B.-C., B. Dolan, W. F. Krajewski, S. Rutledge, and W. Petersen, 2015: Comparison of single and dual polarization based rainfall estimates using NEXRAD data for the NASA Iowa flood studies project. *J. Hydrometeor.*, **16**, 1658–1675, https://doi.org/10.1175/JHM-D-14-0169.1.

——, W. F. Krajewski, and F. Quintero, 2021a: Multi-scale hydrologic evaluation of the national water model streamflow data assimilation. *J. Amer. Water Resour. Assoc.*, **57**, 875–884, https://doi.org/10.1111/1752-1688.12955.

——, ——, ——, S. Buan, and B. Connelly, 2021b: Assessment of streamflow predictions generated using multi-model and multi-precipitation product forcing. *J. Hydrometeor.*, **22**, 2275–2290, https://doi.org/10.1175/JHM-D-20-0310.1.

Sorooshian, S., Q. Duan, and V. K. Gupta, 1993: Calibration of rainfall- runoff models: Application of global optimization to the Sacramento soil moisture accounting model. *Water Resour. Res.*, **29**, 1185–1194, https://doi.org/10.1029/92WR02617.

Souverijns, N., A. Gossart, S. Lhermitte, I. V. Gorodetskaya, S. Kneifel, M. Maahn, F. L. Bliven, and N. P. M. van Lipzig, 2017: Estimating radar reflectivity—Snowfall rate relationships and their uncertainties over Antarctica by combining disdrometer and radar observations. *Atmos. Res.*, **196**, 211–223, https://doi.org/10.1016/j.atmosres.2017.06.001.

Tang, X., D. W. Knight, and P. G. Samuels, 1999: Variable parameter Muskingum–Cunge method for flood routing in a compound channel. *J. Hydraul. Res.*, **37**, 591–614, https://doi.org/10.1080/00221689909498519.

Uccellini, L. W., and J. E. Ten Hoeve, 2019: Evolving the National Weather Service to build a weather-ready nation: Connecting observations, forecasts, and warnings to decisionmakers through impact-based decision support services. *Bull. Amer. Meteor. Soc.*, **100**, 1923–1942, https://doi.org/10.1175/BAMS-D-18-0159.1.

Wigmosta, M. S., and D. P. Lettenmaier, 1999: A comparison of simplified methods for routing topographically driven subsurface flow. *Water Resour. Res.*, **35**, 255–264, https://doi.org/10.1029/1998WR900017.

——, L. W. Vail, and D. P. Lettenmaier, 1994: A distributed hydrology-vegetation model for complex terrain. *Water Resour. Res.*, **30**, 1665–1679, https://doi.org/10.1029/94WR00436.

Xia, Y., and Coauthors, 2012: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *J. Geophys. Res.*, **117**, D03109, https://doi.org/10.1029/2011JD016048.

Yang, Z.-L., and Coauthors, 2011: The community Noah land surface model with multi-parameterization options (Noah-MP): 2. Evaluation over global river basins. *J. Geophys. Res.*, **116**, D12110, https://doi.org/10.1029/2010JD015140.

Zhang, J., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 621–638, https://doi.org/10.1175/BAMS-D-14-00174.1.

Zhang, J. L., Y. P. Li, G. H. Huang, C. X. Wang, and G. H. Cheng, 2016: Evaluation of uncertainties in input data and parameters of a hydrological model using a Bayesian framework: A case study of a snowmelt–precipitation-driven watershed. *J. Hydrometeor.*, **17**, 2333–2350, https://doi.org/10.1175/JHM-D-15-0236.1.

Ziliani, M. G., R. Ghostine, B. Ait-El-Fquih, M. F. McCabe, and I. Hoteit, 2019: Enhanced flood forecasting through ensemble data assimilation and joint state-parameter estimation. *J. Hydrol.*, **577**, 123924, https://doi.org/10.1016/j.jhydrol.2019.123924.