DR. BONG CHUL SEO (Orcid ID : 0000-0002-3092-0241)
PROF. WITOLD KRAJEWSKI (Orcid ID : 0000-0002-3477-9281)

Article type : Technical Note

Multi-scale Hydrologic Evaluation of the National Water Model Streamflow Data Assimilation

Bong-Chul Seo, Witold F. Krajewski, and Felipe Quintero

Iowa Flood Center and IIHR—Hydroscience & Engineering (Seo, Krajewski, Quintero), The University of Iowa, Iowa City, Iowa, USA (Correspondence to Seo: bongchul-seo@uiowa.edu).

Research Impact Statement: Based on the multi-scale evaluation at 70 Iowa locations, the National Water Model streamflow data assimilation leads to improved downstream predictions, compared to open-loop and persistence methods.

ABSTRACT

Streamflow predictions derived from a hydrologic model are subject to many sources of errors, including uncertainties in meteorological inputs, representation of physical processes, and model parameters. To reduce the effects of these uncertainties and thus improve the accuracy of model prediction, the U.S. National Water Model (NWM) incorporates streamflow observations in the modeling framework and updates model-simulated values using the observed ones. This updating procedure is called streamflow data assimilation (DA). This study evaluates the prediction performance of streamflow DA realized in the NWM. We implemented the model using WRF-Hydro® with the NWM modeling elements and assimilated 15-minute streamflow data into the model, observed during 2016–2018 at 140 U.S. Geological Survey stream gauge stations in Iowa. In its current DA scheme, known as "nudging," the assimilation effect is propagated downstream only, which allows us to assess the performance of streamflow predictions generated at 70 downstream stations in the study domain. These 70 locations cover basins of a range of scales, thus enabling a multi-scale hydrologic evaluation by inspecting annual total volume, peak discharge magnitude and timing, and an overall performance indicator represented by the Kling-Gupta efficiency. The evaluation results show that DA improves the prediction skill significantly, compared to open-loop simulation, and the improvements increase with areal coverage of upstream assimilation points.

(KEYWORDS: National Water Model; streamflow assimilation; multi-scale data assimilation; flood forecasting.)

INTRODUCTION

In May 2016, the U.S. National Weather Service (NWS) has implemented and continues to run a continental-scale hydrologic model, the National Water Model (NWM), as part of its operations. The NWM is a distributed hydrologic model that simulates water cycles and predicts streamflow over the entire United States (Cosgrove et al., 2015, 2016). The operational implementation of the NWM demonstrates increasing demand for high-resolution hydrologic information. This modeling framework helps researchers simulate and understand more comprehensive aspects of the interactions between atmosphere and land-surface, which have been unexplored by conventional approaches using lumped and mesoscale models (e.g., Sorooshian et al., 1993; Cuo et al., 2011). Distributed modeling also complements current streamflow guidance provided only at designated sites and expands prediction capabilities to ungauged locations. Recent results from continental-scale retrospective simulations provide a glimpse into modeling performance and demonstrate the early success and potential of data-intensive national-scale flood forecasting (e.g., Rafieeinasab et al., 2016). A recent study by Rojas et al. (2020) documents the performance of the NWM over Iowa at independent locations from which the model included no data.

The motivation to implement streamflow data assimilation (DA) in the NWM was to improve model simulation and forecast initial conditions by correcting modeled streamflow using observations at gauging stations. However, the actual performance and capabilities of DA in the NWM has not been documented well at ungauged locations. Because the NWS has not configured the model to run in an open-loop mode without streamflow observations, and the model replaces modeled streamflow at assimilation locations with observed values in the model outputs, it has been difficult to assess the model's predictive skill. Therefore, we developed a hydrologic evaluation framework to understand the capability of and improvements by the NWM's current DA scheme. We examined multiple aspects of DA's effects on hydrologic prediction and characterized their features regarding catchment scale and fractional coverage of upstream assimilation locations.

MODEL AND DATASET

The NWM is an hourly-based, uncoupled hydrologic modeling and forecasting system built on the WRF-Hydro® community model (Gochis et al., 2018). In this study, we implemented WRF-Hydro® with the NWM configuration, similar to the one running at the NWS, for the Iowa domain where abundant water information is readily accessible via an online platform (e.g., Demir and Krajewski, 2013; Krajewski et al., 2017). In Iowa, there are many U.S. Geological Survey (USGS) stream gauges covering a wide range of drainage scales (Figure 1). This enables a comprehensive performance evaluation of NWM DA across scales. NWM retrospective analysis with streamflow DA requires meteorological forcing products (e.g., precipitation) and streamflow observations, and we collected these data for the period of 2015 to 2018. We note that several earlier studies (Seo et al., 2018; Krajewski et al., 2020; Seo and Krajewski, 2020) include a variety of evaluation (e.g., precipitation) and analyses of these data for the common temporal and spatial domain used in this study.

NWM Implementation

We acquired the NWM domain dataset for Iowa from the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI), using an application known as "domain subsetter (Castronova et al., 2019)" offline. The model grids and parameters were retrieved from the NWM version 1.2.2, rather than the current operational version, 2.0 (the version 1.2.2 was the latest one available with the application at the time of conducting this study). This is unlikely to generate serious differences in simulation results because the version upgrade focused mostly on spatial (e.g., adding Hawaii) and temporal (e.g., extended lookback hours of the analysis cycle for model calibration and regionalization) domain expansion. To implement NWM in our computational environment, we used WRF-Hydro V5.0.3, which allows operational NWM configurations, including the DA capability.

The NWM consists of the Land Surface Model (LSM) and water routing elements, each of which is executed on a different NWM grid resolution (1 km for LSM and 0.25 km for routing, respectively). The LSM represents vertical exchange of energy and water fluxes between atmosphere and land surface using the Noah Multi-Parameterization (Noah-MP) model (Niu et al., 2011; Yang et al., 2011). The routing elements encompass diffusive wave surface routing (Downer, 2002), saturated subsurface flow routing (Wigmosta et al., 1994; Wigmosta and Lettenmaier, 1999), and Muskingum-Cunge channel routing (e.g., Tang et al., 1999). The routing of surface and subsurface is fulfilled on a grid basis, whereas the channel routing functions on vectorized units (i.e., channel links) derived from NHDPlus V2 stream reaches (McKay et al., 2012). We excluded reservoir routing in our NWM configuration to simplify the model implementation and ran the model with a default hydrologic parameter set (without parameter calibration). In the NWM's DA approach (Gochis et al., 2018), parameter calibration in LSM and surface/subsurface routing is of less interest because channel flow routing from an assimilated location along the downstream river reach is the major factor determining streamflow discharge.

Dataset

Input forcing data for the Noah-MP LSM includes incoming short- and long-wave radiation, specific humidity, air temperature, surface pressure, near surface wind components, and precipitation rate. We retrieved these meteorological variables from the hourly North America Land Data Assimilation System (NLDAS) dataset (e.g., Xia et al., 2012) at a resolution of 0.125 degrees. In our forcing dataset, we replaced the NLDAS precipitation rate data with the Multi-Radar Multi-Sensor (MRMS; Zhang et al., 2016) product as a separate precipitation forcing, which includes a rain gauge correction with an enhanced resolution of 0.01 degrees. We collected these hourly NLDAS and MRMS data for 2015–2018 and resampled them onto the 1-km LSM grid for model (Noah-MP) forcing.

We collected streamflow data from 140 USGS stations in Iowa (Figure 1) where quality-controlled streamflow records are available at a 15-minute resolution. These streamflow data facilitate streamflow DA at all USGS locations and the evaluation of DA at their downstream gauge locations. As indicated in Figure 1, 70 USGS locations are available for the DA evaluation; this number varies slightly depending on the status of missing data at these stations. The streamflow records were obtained by converting measured water level (stage) into discharge using well-defined rating curves produced for each site. The USGS has developed these rating curves from periodic collection of stage-discharge measurements, especially during low- and high-flow events. In this study, we do not consider rating curve uncertainty and its effect on our DA evaluation.

METHODOLOGY

NWM Simulations

To assess the improvement made by DA, we simulated the NWM with DA and open-loop (no DA) modes for a period from August 2015 to December 2018. We used the early simulation period (August 2015 to March 2016) to warm-up the model states for the remaining analysis period. Because precipitation estimation for winter months still remains challenging (e.g., Seo et al., 2015; Souverijns et al., 2017) and thus may affect model simulation results, we limited the analysis of simulation results to the period of April through October in each year (2016–2018).

The DA scheme in NWM is knowns as "nudging" and consists of direct insertion; i.e., the observed value replaces the model value without considering the associated uncertainty. In the DA procedure, we did not account for the quality of observed streamflow in the nudging process (see Gochis et al., 2018) in that the measurement (or rating curve) uncertainties are unknown. Nudge at the assimilation location is defined as the difference between observed and model estimated streamflow (i.e., model error) with a limited temporal interpolation. In the NWM, spatial smoothing is inactive for computational efficiency, while temporal smoothing assigns a heavy weight to an observation within 15 minutes from the current time and sets e-folding time as two hours. The calculated adjustment (nudge) at each assimilation location is then propagated downstream through a channel routing procedure using the Muskingum-Cunge method:

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

where Q denotes streamflow discharge at the current (t) and previous $(t-1)$ times at the downstream (d) and upstream (u) reaches. C_1, C_2, and C_3 are coefficients calculated using routing parameters (see Tang et al., 1999), and q_l and D indicate lateral inflow and the wedge storage contribution from lateral inflow. The model includes the nudge $N_d(t-1)$ in all three streamflow terms in Eq. (1) to lessen discontinuity between the upstream and downstream reaches. However, the nudge included in the first and second terms for the upstream reach is applied only for solving downstream discharge in Eq. (1) and is not saved as part of the model output for the upstream reach. In other words, the nudge is not propagated upstream.

DA Evaluation

A meaningful evaluation of DA requires a comparison of the model-estimated streamflow (at the evaluation locations) with observations at points unused in the data assimilation. In the NWM, DA replaces model-simulated values with the observations, if valid observations are available at the gauging stations. In the NWM setup, this is challenging for DA evaluation because the model assimilates the observed values at all USGS stations shown in Figure 1, including the 70 evaluation locations, which also become assimilation points for their downstream reaches. Therefore, we decided to retrieve the simulated streamflow values (for DA evaluation) at the immediate upstream links directly connected with the evaluation point, assuming that the effects of channel routing and lateral inflow along the stream link containing the evaluation point are negligible. To explore the validity of this assumption, we conducted an experiment with two selected locations (Van Meter and Cedar Falls), which cover different scale basins as shown in Figure 1. In the experiment, we did not provide streamflow observations at Van Meter and Cedar Falls to avoid replacement of model generated streamflow with the observations (i.e., to obtain model streamflow propagated from upstream DA). The result of this experiment is presented in the next section. As reference for DA evaluation, we employed the persistence-based prediction (e.g., Krajewski et al., 2020), which assumes spatial persistence from upstream observations. If there are multiple upstream stations on different branches of the river network (see Krajewski et al. 2020 for details), a simple addition of their observations would provide a predicted value at the downstream location.

We compared the prediction performance of the NWM with DA to the performance without DA (NoDA) and persistence (indicated as "No Model"). The evaluation metrics used in the analyses are: (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). The peak errors are calculated for an annual maximum discharge. The formulas of these metrics are provided in Eqs. (2)-(5):

$$RE_V = \frac{V_{NWM} - V_{obs}}{V_{obs}} \times 100\% \quad (2)$$

$$RE_{Q_p} = \frac{Q_{p,NWM} - Q_{p,obs}}{Q_{p,obs}} \times 100\% \quad (3)$$

$$E_{t_p} = t_{p,NWM} - t_{p,obs} \quad (4)$$

$$KGE = 1.0 - \sqrt{(\rho - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (5)$$

where V, Q_p, and t_p denote total volume (m^3), peak discharge (m^3s^{-1}), and peak time (h) obtained from model simulations (NWM) and observations (obs) from April to October of each year. KGE (Gupta et al., 2009) is an overall performance indicator describing the predictive power of hydrologic models and 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. We examined these evaluation metrics, focusing on catchment scale and the analyzed performance improvements accomplished by DA (against NoDA), with respect to the areal coverage fraction defined using the assimilated upstream catchment area. The improvements are defined as simple differences in the evaluation metrics calculated with and without data assimilation.

RESULTS

The results of the experiment, conducted to learn whether using model prediction from upstream links is suitable for our analysis, are presented in Figure 2 for two gauging stations. These results show that streamflow discharge at the two locations and their upstream links, represented by blue and red solid lines, agree very well; there is little if any difference between them. The KGE values for the blue and red lines appear to be the same (0.79 and 0.91 for Van Meter and Cedar Falls, respectively). This allows us to use the modeled streamflow at the upstream links for DA evaluation. The model simulations at the location of the evaluation gauge are “corrupted” by the data collected there. Figure 2 also demonstrates that DA significantly improves model performance at the two locations, compared to open-loop simulations. For example, DA eliminated an erroneous peak observed at Van Meter in August 2016 and significantly improved KGE (0.13 vs. 0.79).

In Figure 3, we present the evaluation results focusing on the four metrics defined in Eqs. (2)-(5) for each simulation year. We assessed the NWM's prediction performance with DA and NoDA, compared to the result from the persistence method indicated as "No Model" in Figure 3. To calculate the relative peak error (RE_{Q_p}) and peak timing error (E_{t_p}), we identified an NWM simulated peak within a scale-dependent time window around the annual peak observed from the USGS streamflow data. We made this choice because the model occasionally generates an annual peak at a completely different time, as shown in the case of Van Meter in Figure 2. The search window was defined using time of concentration (i.e., the longest travel time along the river network) or 5 days, whichever is smaller. In Figure 3, DA seems to perform better at estimating runoff volume and peak discharge than NoDA and persistence do. For RE_V and RE_{Q_p}, most of the red dots representing DA stay near the no error (0%) line and within a ±50 % range, respectively, whereas NoDA and persistence show underestimations both in volume and peak discharge. Persistence leads to underestimations in volume and peak discharge, and early peak timing, as illustrated in Figure 3; drainage areas (represented by single or multiple upstream gauging stations) that are smaller than the area represented by the downstream evaluation station yield the observed underestimations and early peak. However, the overall performance (KGE) of persistence seems better at many locations than that of model simulation with NoDA, implying that the forecasting approach without models can provide useful guidance if there are reliable gauging stations upstream (see Krajewski et al., 2020). Overall, the NWM with DA outperforms persistence and NoDA based on KGE. We note that DA's slight underestimations of total volume might be the result of lateral inflow missed along the stream links of evaluation points.

We examined the scale-dependent performance of DA and persistence in Figure 4. In this analysis, we excluded the result with NoDA because its performance was lower than those of DA and persistence. As shown in Figure 4, the performance of DA- and persistence-based predictions tends to improve as catchment scale becomes larger. This scale-dependence is obviously shown in KGE, while E_{t_p}reveals wide distribution across catchment scales (many locations have timing errors outside a one-day window from the actual peak time). With increasing scale, the dispersion of RE_V and RE_{Q_p}decreases, and the mean of these errors gradually approaches negligible bias. The key findings from Figure 4 are: (1) DA outperforms persistence, particularly at smaller scales (e.g., approximately up to 5,000 km^2) for the study domain; and (2) persistence-based predictions are comparable with the ones made by DA at larger scales. This is understandable because the skill in the streamflow prediction is determined by measuring the water already in the river system.

Based on the results shown in Figures 3 and 4, we quantified the performance improvements (e.g., in terms of each evaluation metric) attained by DA in the NWM procedures. Figure 5 shows the improved model performance characterized by the areal coverage fraction presented in Figure 5(c), which describes the areal coverage of upstream assimilation stations to the entire catchment delineated by downstream evaluation station. As shown in Figure 5, the DA performance tends to improve as the upstream stations cover larger areas, indicating that fractional coverage is a primary factor in determining the performance of DA. The large variability of the KGE improvement is somewhat surprising. While the improvement is greater because more of the upstream area is being monitored, the variability is high. The variability in the improvement is partially due to the statistical effect of the relative sample size and is also a consequence of the model performance (e.g., open-loop) itself. For example, when the model works well with an open-loop mode, the expected improvement by DA is small. When the model works poorly, the potential for improvement is much higher (see Supporting Information).

As we discussed in the "NWM Implementation" section, parameter calibration in the LSM and surface/subsurface routing elements would be less impactful if this coverage fraction is sufficiently high. Streamflow assimilation diminishes uncertainties/errors generated by misinterpreted parameters in upstream catchment modeling. We recognize from Figure 5 that improving the peak estimation is challenging with large variability even at the higher coverage fraction range, although the total volume reveals relatively low variability. Figure 5 could provide insight for the potential performance of DA for other regions with landscapes similar to Iowa's (e.g., no complex terrain and natural channels).

SUMMARY AND CONCLUSIONS

This study extensively evaluated the NWM's DA performance based on our model implementation that updated the model-simulated streamflow every 15 minutes using streamflow data observed during 2016–2018 at 140 USGS stations in Iowa. Our investigation builds on a recent evaluation done by Rojas et al. (2020) on an earlier version of the NWM. Since NWM

DA evaluation is challenging with the current NWM configuration (there is no access to the open-loop prediction at the assimilation data points), we developed a novel framework to assess streamflow predictions generated by the DA procedure. To demonstrate DA's prediction capability compared to the open-loop (NoDA) and persistence (No Model) method, we measured an overall performance metric known as KGE and errors in annual total volume, peak discharge, and peak timing. The analysis results showed that DA significantly improves streamflow prediction. The improvements (DA vs. NoDA) were characterized by the areal coverage fraction of the upstream assimilation point; it tends to increase with larger fractional coverage (Figure 5). Given the large dispersion in the annual peak errors (e.g., amounts and time), predicting the peak remains challenging, even using the DA procedure. We plan to investigate this aspect further to learn if another channel routing scheme or use of a different set of parameters (e.g., calibration) with the current scheme can ameliorate the peak estimation. The tendency of prediction improvement observed in Figure 5 could be used as reference for application of DA to other regions or guidance when designing a stream sensor network for hydrologic prediction.

We used persistence-based predictions as reference to assess the DA-based prediction results. The persistence method incorporates streamflow observations from the same upstream stations used in DA and its concept is rather simple but efficient (e.g., Krajewski et al., 2020). We found that DA outperforms persistence, particularly at catchment scales smaller than 5,000 km^2 (the number might be different at different regions depending on the configuration of stream gauge network), where the coverage fraction is not as good as the one for larger scales as shown in Figure 5(c). This should come as no surprise because the model uses additional information, i.e., rainfall. Nevertheless, the performance of persistence looks impressive and reliable at larger scales, and thus could be a good alternative to save model computation time and computational resources. The multi-scale evaluation of this study revealed its scale-dependent features: (1) the prediction performance increases as catchment scale becomes larger (e.g., KGE); and (2) KGE and errors in volume and peak discharge are approaching ideal prediction (e.g., no error), and their dispersion decreases significantly at larger scales.

RECOMMENDED FUTURE RESEARCH

Numerous stage-only sensors exist that can complement the current coverage of USGS stations and thus expand DA's performance to relatively smaller basins. A good example is about 250 stream sensors (Kruger et al., 2016) operated by the Iowa Flood Center (IFC) to monitor streams and creeks near Iowa communities. The IFC has developed a procedure to build "synthetic rating curves" (Quintero et al., 2021) using hydraulic/hydrologic models. Soon we will include these stations in our NWM configuration and fill the significant scale gap (e.g., smaller than 1,000 km^2) shown in Figure 4. This incorporation will also provide an opportunity to independently evaluate the synthetic rating curves developed using the IFC's Hillslope Link Model (Krajewski et al., 2017; Quintero et al., 2020) with DA procedures different than the one used in the NWM.

SUPPORTING INFORMATION

Additional supporting information may be found online under the Supporting Information tab for this article: A figure accounting for the variability of prediction improvement shown in Figure 5.

ACKNOWLEDGMENTS

This study was supported by the Iowa Flood Center at the University of Iowa and the Hydrometeorology Testbed (HMT) Program within NOAA/OAR Office of Weather and Air Quality under Grant No. NA17OAR4590131. The authors thank Dr. Anthony Castronova at CUAHSI for providing the NWM grids and parameters. The authors are also grateful to Drs. Aubrey Dugger and David Gochis at the National Center for Atmospheric Research for guidance on our NWM implementation.

LITERATURE CITED

Castronova, A.M., D. Tijerina, A.L. Dugger, A. Rafieeinasab, M. McAllister, L.E. Condon, H. Tran, J. Zhang, D. Gochis, and R.M. Maxwell. 2019. "Improving access to continental-scale hydrology models for research and education - A subsetting adventure." Paper presented at 2019 AGU Fall meeting, American Geophysical Union, San Francisco, CA.

Cosgrove, B., D. Gochis, E.P. Clark, Z. Cui, A.L. Dugger, G.M. Fall, X. Feng, M.A. Fresch, J.J. Gourley, S. Khan, D. Kitzmiller, H.S. Lee, Y. Liu, J.L McCreight, A.J. Newman, A. Oubeidillah, L. Pan, C. Pham, F. Salas, K.M Sampson, M. Smith, G. Sood, A. Wood, D.N. Yates, W. Yu, and Y. Zhang. 2015. "Hydrologic modeling at the National Water Center: Operational implementation of the WRF-Hydro model to support National

Weather Service hydrology." Paper presented at 2015 AGU Fall meeting, American Geophysical Union, San Francisco, CA.

Cosgrove, B., D. Gochis, E.P. Clark, Z. Cui, A.L. Dugger, X. Feng, L.R. Karsten, S. Khan, D. Kitzmiller, H.S. Lee, Y. Liu, J.L McCreight, A.J. Newman, A. Oubeidillah, L. Pan, C. Pham, F. Salas, K.M Sampson, G. Sood, A. Wood, D.N. Yates, and W. Yu. 2016. "An overview of the National Weather Service National Water Model." Paper presented at 2016 AGU Fall meeting, American Geophysical Union, San Francisco, CA.

Cuo, L., T.C. Pagano, and Q.J. Wang. 2011. "A review of quantitative precipitation forecasts and their use in short- to medium-range streamflow forecasting." *Journal of Hydrometeorology* 12(5): 713–728. https://doi.org/10.1175/2011JHM1347.1

Demir, I., and W.F. Krajewski. 2013. "Towards an integrated flood information system: Centralized data access, analysis, and visualization." *Environmental Modelling and Software* 50: 77–84. https://doi.org/10.1016/j.envsoft.2013.08.009

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." *Hydrological Processes* 16(2): 255–275. https://doi.org/10.1002/hyp.338

Gochis, D., M. Barlage, A. Dugger, L. Karsten, M. McAllister, J. McCreight, J. Mills, A. Rafieeinasab, L. Read, K. Sampson, D. Yates, and W. Yu. 2018. "The WRF-Hydro modeling system technical description (Version 5.0)." NCAR Technical Note. https://doi.org/10.5065/D6J38RBJ

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." *Journal of Hydrology* 377(1–2): 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Krajewski, W.F., D. Ceynar, I. Demir, R. Goska, A. Kruger, C. Langel, R. Mantilla, J. Niemeier, F. Quintero, B.-C. Seo, S.J. Small, L.J. Weber, and N. Young. 2017. "Real-time flood forecasting and information system for the State of Iowa." *Bulletin of the American Meteorological Society* 98(3): 539–554. https://doi.org/10.1175/BAMS-D-15-00243.1

Krajewski, W.F., G.R. Ghimire, and F. Quintero. 2020. "Streamflow forecasting without models." *Journal of Hydrometeorology* 21(8): 1689–1704. https://doi.org/10.1175/JHM-D-19-0292.1

Kruger, A., W.F. Krajewski, J.J. Niemeier, D.L. Ceynar, and R. Goska. 2016. "Bridge mounted river stage sensors (BMRSS)." *IEEE Access* 4: 8948–8967. https://doi.org/10.1109/ACCESS.2016.2631172

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

Niu, G.-Y., Z.-L. Yang, K.E. Mitchell, F. Chen, M.B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari, and Y. Xia. 2011. "The community Noah land surface model with multi-parameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements." *Journal of Geophysical Research* 116: D12109. https://doi.org/10.1029/2010JD015139

Quintero, F., W.F. Krajewski, M. Muste, M. Rojas, G. Perez, S.J. Johnson, A.N. Anderson, T.J. Honemuller, W. Cappuccio, and J. Zogg. 2021. "Development of synthetic rating curves: A case study in Iowa." *Journal of Hydrologic Engineering* 26(1): 05020046. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002022

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." *Journal of Hydrology* 584: 124686. https://doi.org/10.1016/j.jhydrol.2020.124686

Rafieeinasab, A., J.L. McCreight, A.L. Dugger, D. Gochis, L.R. Karsten, Y. Zhang, B. Cosgrove, and Y. Liu. 2016. "Evaluation of streamflow forecast for the National Water Model of U.S. National Weather Service." Paper presented at 2016 AGU Fall meeting, American Geophysical Union, San Francisco, CA.

Rojas, M., F. Quintero, and W.F. Krajewski. 2020. "Performance of the National Water Model Analysis and Assimilation configuration over Iowa." *Journal of the American Water Resources Association* 56(4): 568–585. https://doi.org/10.1111/1752-1688.12820

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." *Journal of Hydrometeorology* 16(4): 1658–1675. https://doi.org/10.1175/JHM-D-14-0169.1

Seo, B.-C., and W.F. Krajewski. 2020. "Statewide real-time quantitative precipitation estimation using weather radar and NWP model analysis: Algorithm description and product evaluation." *Environmental Modelling and Software* 132: 104791. https://doi.org/10.1016/j.envsoft.2020.104791

Seo, B.-C., F. Quintero, and W.F. Krajewski. 2018. "High-resolution QPF uncertainty and its implications for flood prediction: A case study for the Eastern Iowa flood of 2016." *Journal of Hydrometeorology* 19(8): 1289–1304. https://doi.org/10.1175/JHM-D-18-0046.1

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 Resources Research* 29(4): 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." *Atmospheric Research* 196: 211–223. https://doi.org/10.1016/j.atmosres.2017.06.001

Tang, X., D. Knight, and P.G. Samuels. 1999. "Variable parameter Muskingum-Cunge method for flood routing in a compound channel." *Journal of Hydraulic Research* 37(5): 591–614. https://doi.org/10.1080/00221689909498519

Wigmosta, M.S., and D.P. Lettenmaier. 1999. "A comparison of simplified methods for routing topographically driven subsurface flow." *Water Resources Research* 35(1): 255–264. https://doi.org/10.1029/1998WR900017

Wigmosta, M.S., L.W. Vail, and D.P. Lettenmaier. 1994. "A distributed hydrology-vegetation model for complex terrain." *Water Resources Research* 30(6): 1665–1679. https://doi.org/10.1029/94WR00436

Xia, Y., K. Mitchell, M. Ek, J. Sheffield, B. Cosgrove, E. Wood, L. Luo, C. Alonge, H. Wei, J. Meng, B. Livneh, D. Lettenmaier, V. Koren, Q. Duan, K. Mo, Y. Fan, and D. Mocko. 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." *Journal of Geophysical Research: Atmospheres* 117: D03109. https://doi.org/10.1029/2011JD016048

Yang, Z.-L., G.-Y. Niu, K.E. Mitchell, F. Chen, M.B. Ek, M. Barlage, L. Longuevergne, K. Manning, D. Niyogi, M. Tewari, and Y. Xia. 2011. "The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins." *Journal of Geophysical Research* 116: D12110. https://doi.org/10.1029/2010JD15140

Zhang, J., K. Howard, C. Langston, B. Kaney, Y. Qi, L. Tang, H. Grams, Y. Wang, S. Cocks, S. Martinaitis, A. Arthur, K. Cooper, J. Brogden, and D. Kitzmiller. 2016. "Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities." *Bulletin of the American Meteorological Society* 97(4): 621–638. https://doi.org/10.1175/BAMS-D-14-00174.1

Figure Legends

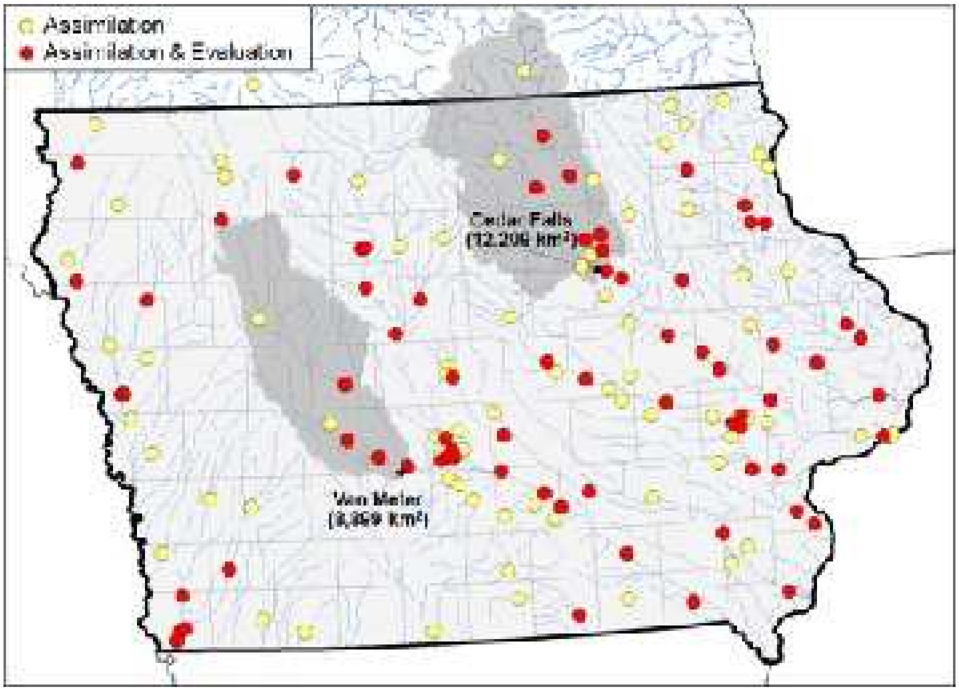
Figure 1. The locations of 140 USGS stations in the study domain where streamflow observations were assimilated into the NWM. The yellow circles represent the uppermost USGS stream gauges. The red circles indicate the evaluation points in this study. The solid blue lines represent river and stream networks. The two shaded watersheds delineate the drainage areas of two USGS stations (Van Meter and Cedar Falls) used in Figure 2.

Figure 2. Observed and NWM simulated hydrographs with DA and open-loop (NoDA) modes at Van Meter (USGS 05484500) and Cedar Falls (USGS 05463050) in Iowa.

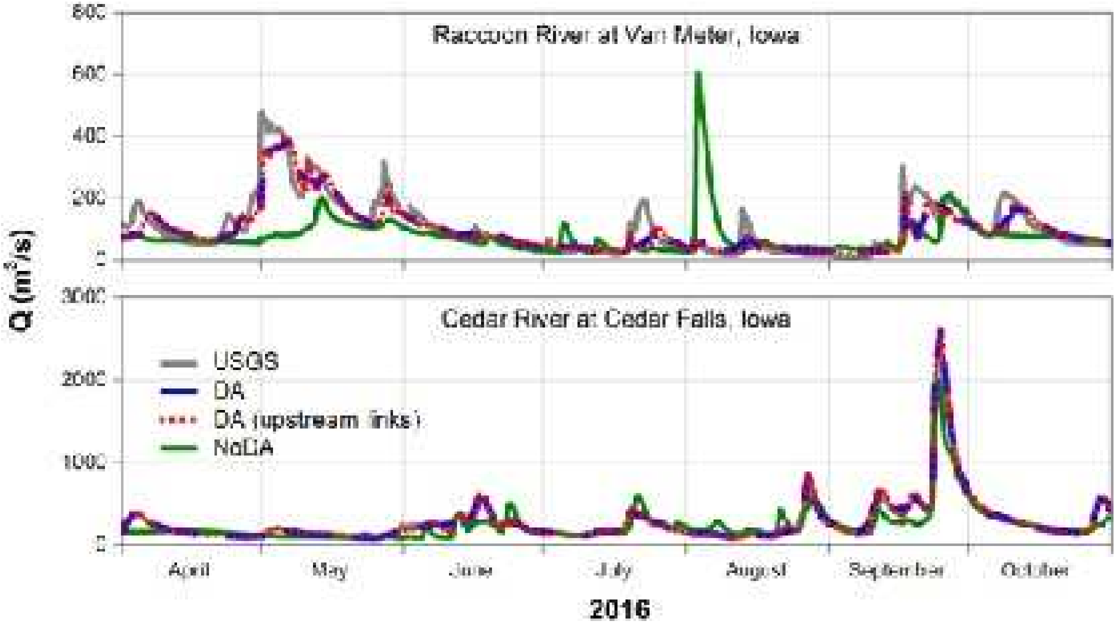
Figure 3. Performance comparison of model simulation results (DA and NoDA) with those of persistence (No Model). Each circle indicates one of 70 individual evaluation locations presented in Figure 1.

Figure 4. Performance comparison between the results of DA and persistence (No Model) regarding catchment scale.

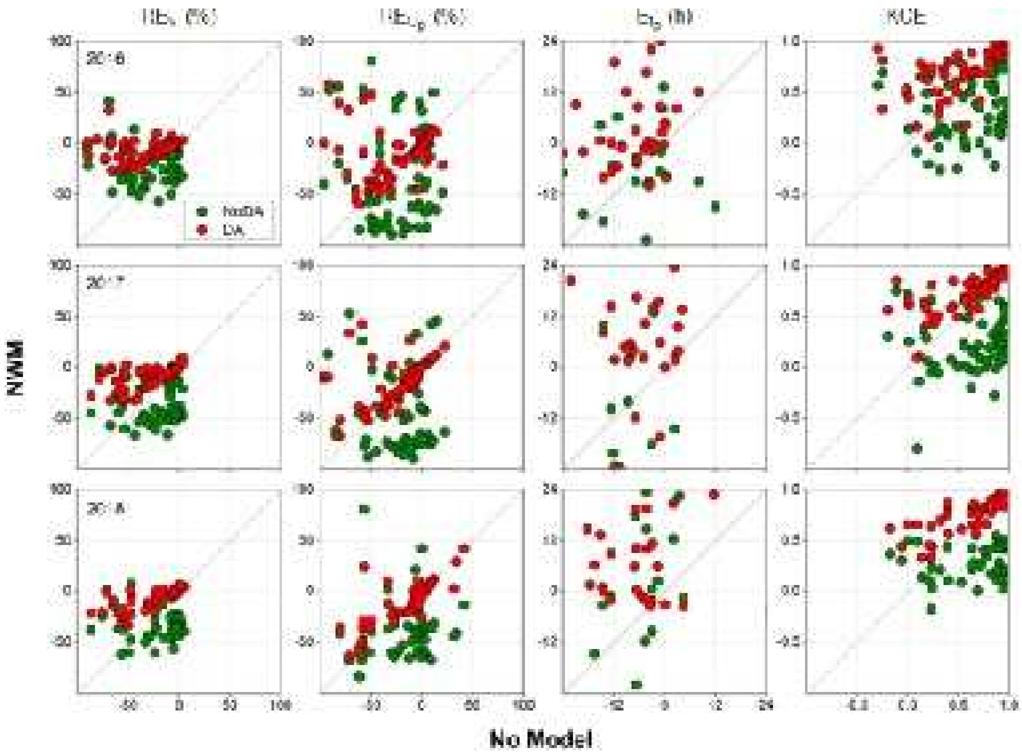
Figure 5. Performance improvement characterized by (a) the areal coverage fraction of upstream assimilation locations to a downstream evaluation location and (b) the distribution change of peak timing error. The distribution of areal coverage fraction is shown in (c).



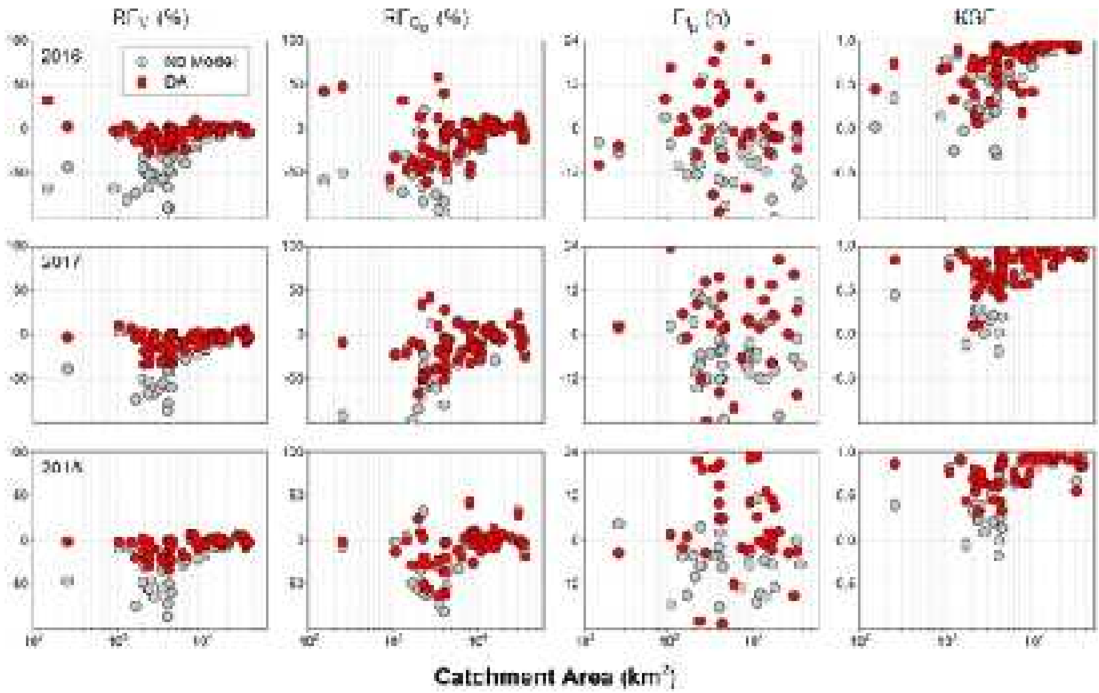
jawr_12955_f1.tif



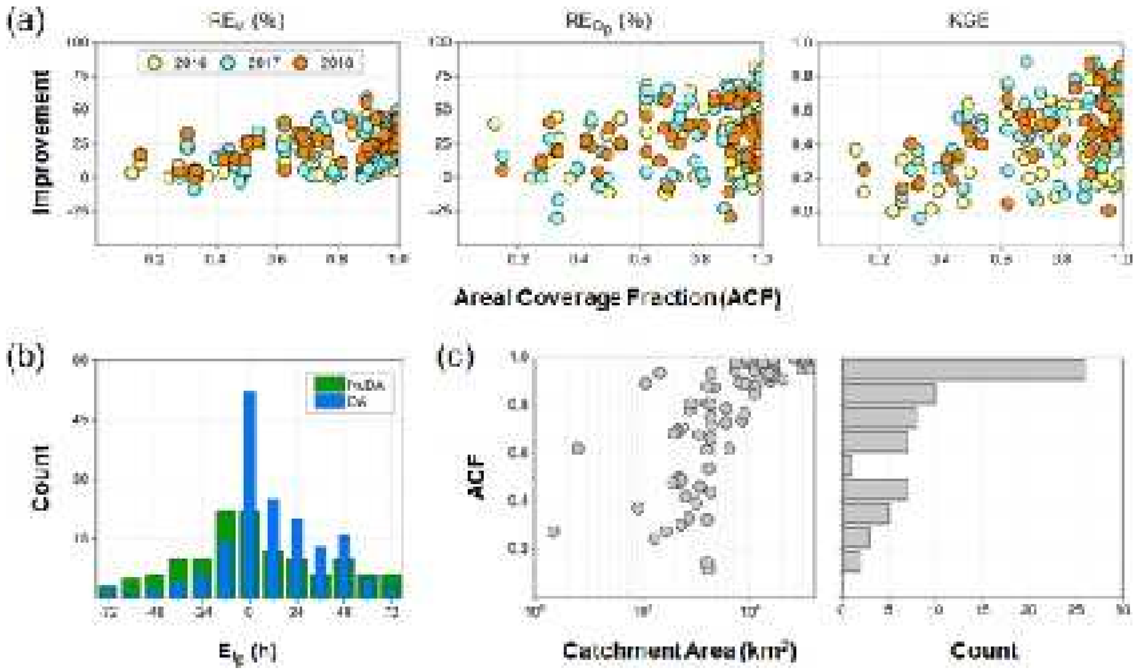
jawr_12955_f2.tif



jawr_12955_f3.tif



jawr_12955_f4.tif



jawr_12955_f5.tif