

Assessment of Streamflow Predictions Generated Using Multimodel and Multiprecipitation Product Forcing

BONG-CHUL SEO,^a WITOLD F. KRAJEWSKI,^a FELIPE QUINTERO,^a STEVE BUAN,^b AND BRIAN CONNELLY^b

^a*Iowa Flood Center and IIHR—Hydroscience and Engineering, University of Iowa, Iowa City, Iowa*

^b*North Central River Forecast Center, National Weather Service, Chanhassen, Minnesota*

(Manuscript received 21 December 2020, in final form 18 May 2021)

ABSTRACT: This study assesses streamflow predictions generated by two distributed hydrologic models, the Hillslope Link Model (HLM) and the National Water Model (NWM), driven by three radar-based precipitation forcing datasets. These forcing data include the Multi-Radar Multi-Sensor (MRMS), and the Iowa Flood Center's single-polarization-based (IFC-SP) and dual-polarization-based (IFC-DP) products. To examine forcing- and model-dependent aspects of the representation of hydrologic processes, we mixed and matched all forcing data and models, and simulated streamflow for 2016–18 based on six forcing–model combinations. The forcing product evaluation using independent ground reference data showed that the IFC-DP radar-only product's accuracy is comparable to MRMS, which is rain gauge corrected. Streamflow evaluation at 140 U.S. Geological Survey (USGS) stations in Iowa demonstrated that the HLM tended to perform slightly better than the NWM, generating streamflow with smaller volume errors and higher predictive power as measured by Kling–Gupta efficiency (KGE). The authors also inspected the effect of estimation errors in the forcing products on streamflow generation and found that MRMS's slight underestimation bias led to streamflow underestimation for all simulation years, particularly with the NWM. The less biased product (IFC-DP), which has higher error variability, resulted in increased runoff volumes with larger dispersion of errors compared to the ones derived from MRMS. Despite its tendency to underestimate, MRMS showed consistent performance with lower error variability as reflected by the KGE. The dispersion observed from the evaluation metrics (e.g., volume error and KGE) seems to decrease as scale becomes larger, implying that random errors in forcing are likely to average out at larger-scale basins. The evaluation of simulated peaks revealed that an accurate estimation of peak (e.g., time and magnitude) remains challenging, as demonstrated by the highly scattered distribution of peak errors for both hydrologic models.

KEYWORDS: Hydrologic models; Model comparison; Model evaluation/performance

1. Introduction

Recent implementation of the National Water Model (NWM) into U.S. National Weather Service (NWS) operations demonstrates increasing demand for high-resolution hydrologic modeling and forecasting. NWS's operational hydrologic forecasting has relied on a long-standing lumped model with extensive calibration (e.g., Sorooshian et al. 1993; Koren et al. 2014). Many studies (e.g., Reed et al. 2004; Smith et al. 2012) have explored the predictive capability of distributed models, suggesting some requirements for distributed models to replace or complement the lumped model. While the calibrated lumped model is still a primary forecast tool for the NWS's River Forecast Centers, high-resolution modeling can help describe scale-dependent variability and many details of interactions between the atmosphere and the land surface (e.g., Cole and Moore 2009). Conventional approaches (i.e., lumped hydrology and mesoscale weather) have addressed these poorly. Furthermore, high-resolution distributed modeling can complement current hydrologic guidance at NWS forecast points and expand forecast capability and coverage to ungauged locations (e.g., Cosgrove et al. 2016; Cohen et al. 2018).

High-resolution modeling and forecasting require model configuration using high-resolution topography and precipitation

forcing data. The NWM configuration includes (i) high-resolution modeling grids and National Hydrography Dataset (NHD) Plus V2 (McKay et al. 2012) for landscape representation and (ii) high-resolution precipitation forcing, such as the Multi-Radar Multi-Sensor (MRMS; Zhang et al. 2016) quantitative precipitation estimation (QPE) product. The results from continental-scale retrospective simulations driven by such forcing and topography data provided a glimpse into modeling performance, demonstrating an early success and the potential for the data-intensive national-scale flood forecasting (e.g., Hansen et al. 2019). However, the key features of prediction errors associated with model structure and individual routing components, basin scale, and uncertainty in precipitation forcing data have not been extensively examined since NWM's operational implementation in 2016. Different forcing products (e.g., Seo et al. 2018), different models (e.g., Reed et al. 2004; Smith et al. 2012), or different forcing–model combinations may increase or decrease predictive capability. It is crucial to understand how each element employed in the forecasting system behaves and how it contributes to skill in streamflow simulation. As such, a framework that can investigate differences in prediction accuracy derived by the different configurations of forcing products and models is required to improve our understanding and prediction skills.

In this study, we explore streamflow prediction skill and uncertainty derived from various combinations of multiple hydrologic models and precipitation forcing products. We focus on

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

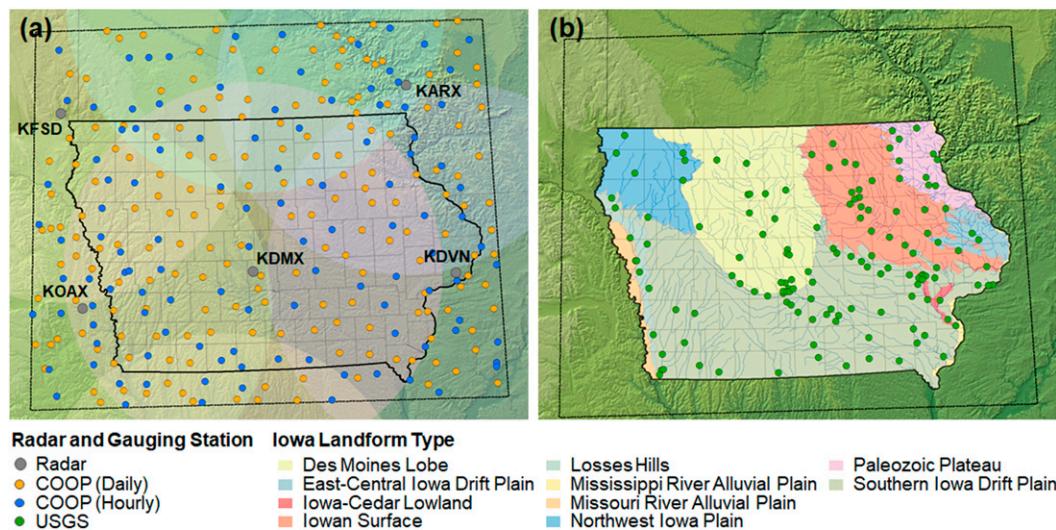


FIG. 1. (a) The locations of NWS COOP rain gauges and (b) USGS streamflow gauges in the study domain represented by the black dashed box. The circular areas in (a) demarcate 230-km ranges centered on the WSR-88D radars indicated by four-digit codes. Landform types presented in (b) are color coded to explore landscape-dependent runoff features.

streamflow prediction during warm and hot months (e.g., April–October) because streamflow during winter and early spring is primarily affected by frozen ground and snowmelt combined with winter precipitation, the data for all of which typically contain large uncertainties. This mix-and-match approach is critical, as the conventional approach of calibrating hydrologic models is challenging at the national scale (e.g., Beven 1993). Accordingly, herein we used the Iowa Flood Center (IFC; Krajewski et al. 2017) Hillslope Link Model (HLM) and the NWM for hydrologic models; and MRMS and IFC's radar-based QPE products for precipitation forcing data. The theoretical roots of the HLM are in the scaling properties of the river networks and landscape decomposition into hillslopes and channel links (Mantilla and Gupta 2005; Gupta et al. 2010, 2015). The scale of the hillslopes, where the conversion of rainfall into runoff takes place, is much smaller than the scale of the NHDPlus basins used as the topographic underlining for the NWM. At larger scales, the two models are likely compatible, only within the margin of errors due to digital elevation model (DEM) data processing and network extraction (Quintero and Krajewski 2018). The assessment of the mix-and-match simulation results presented in this study will provide valuable insights to assist researchers and operational forecasters to understand modeling uncertainties and improve prediction skills.

2. Hydrologic models and data

We applied the mix-and-match approach to the IFC's forecasting domain where a variety of hydrologic data resources are instantly accessible through a web portal known as the Iowa Flood Information System (e.g., Demir and Krajewski 2013; Krajewski et al. 2017). The climate in Iowa is described by wet springs, hot summers, and cold winters. Iowa's mean

annual precipitation and evaporation are about 860 mm (<http://www.ocs.orst.edu/>) and 580 mm (<http://www.ntsg.umt.edu/project/mod16/>), respectively. The main portion of land cover is agriculture with corn–soybean rotation. Major rivers and stream gauge stations in Iowa belong to the NWS North Central River Forecast Center (NCRFC) forecasting domain, while smaller rivers in western Iowa belong to the Missouri Basin RFC territory. We used streamflow observations from 140 U.S. Geological Survey (USGS) stream gauges in Iowa to evaluate the results of multiyear mix-and-match simulations for the period from 2016 to 2018. We also collected meteorological forcing data, including multiple precipitation products required for both HLM and NWM model simulations. The following subsections provide brief descriptions of models, model forcing products, and reference datasets (e.g., rain gauge and streamflow observations) used for the evaluation of forcing products and model simulations. The locations of rain and stream gauge stations are illustrated in Figs. 1a and 1b, respectively.

a. Hydrologic models

1) HLM

The Hillslope Link Model (HLM) is a conceptual hydrologic model that simulates the main aspects of surface processes and flood genesis. HLM is distributed in space using an irregular mesh given by the partitioning of the landscape into hillslopes and channels. In the model representation of the river network, a channel-link pair is defined as the portion of a channel between two junctions of a river network, and hillslope is the adjacent area that drains into the link. In HLM, the hillslope is the volume control unit for runoff production, and runoff propagation occurs from each hillslope to its adjacent channel link. The conceptualization of runoff production at each hillside consists of several vertical tanks representing different

water storages in a soil column. These tanks are (i) snow tank, (ii) surface ponding, (iii) topsoil, (iv) subsurface, and (v) channel. Vertical fluxes connecting these tanks are represented by the processes of precipitation, snowmelt, evapotranspiration, infiltration, and percolation. Horizontal fluxes are associated with overland flow, interflow, and base flow and provide inputs into the channel tank. The HLM physics and equations are documented in prior research (e.g., [Mantilla and Gupta 2005](#); [Gupta et al. 2010](#); [Quintero et al. 2016](#); [ElSaadani et al. 2018](#)), and recent improvements in HLM's routing elements are reported in [Ghimire et al. \(2018\)](#) and [Quintero et al. \(2020\)](#). Since IFC's establishment in 2009, the HLM has been the IFC's operational forecast model to provide real-time streamflow forecasts for Iowa communities ([Krajewski et al. 2017](#)). Mathematically, the model consists of a large system of ordinary nonlinear differential equations organized to correspond with the river network topology. This allows use of an efficient numerical solver designed for high-performance computing ([Small et al. 2013](#)) and capable of updating the forecasts as frequently as every 15 min.

2) NWM

The National Water Model (NWM) is also a highly distributed hydrologic model that simulates and forecasts streamflow over the entire United States, based on an hourly modeling cycle. The NWM is configured using hydrologic processes and routing components (e.g., subsurface, surface/terrain, and channel routing) included in the community WRF-Hydro modeling system ([Gochis et al. 2018](#)) developed at the National Center for Atmospheric Research (NCAR). The key modules of the NWM system consist of the Noah Multi-Parameterization (Noah-MP) land surface model (LSM) to represent land surface processes ([Niu et al. 2011](#); [Yang et al. 2011](#)) and separate water routing modules. The LSM simulates the vertical exchange of water and energy fluxes between the Earth surface and atmosphere interface on a 1-km grid. The routing modules encompass diffusive wave surface routing ([Downer et al. 2002](#)) and saturated subsurface routing ([Wigmsta et al. 1994](#); [Wigmsta and Lettenmaier 1999](#)), based on a 250-m grid, as well as Muskingum–Cunge channel routing (e.g., [Tang et al. 1999](#)) using the vectorized NHDPlusV2 stream units ([McKay et al. 2012](#)). To improve the model's initial states for its forecasting cycles, a simple nudging data assimilation (DA) scheme (e.g., [Gochis et al. 2018](#); [Seo et al. 2020](#), manuscript submitted to *J. Amer. Water Resour. Assoc.*) is applied to the channel routing routine using observed streamflow data. However, we excluded DA and reservoir routing in our NWM configuration for the simplicity of model implementation and a fair comparison with the HLM simulation.

b. Model forcing products

Hydrologic models require various forcing data to trigger interactions between their modeling elements (e.g., atmosphere–surface and surface–subsurface). These forcing data for distributed models include gridded precipitation and environmental variables (e.g., surface temperature), estimated using remote sensing platforms (e.g., radar and satellite) or numerical weather prediction (NWP) models. For both HLM

and NWM simulations, we used MRMS and IFC radar-based precipitation products as the main driving factor. Because NWM land surface modeling (i.e., Noah-MP) requires additional environmental variables, we retrieved those forcing data from the North America Land Data Assimilation System (NLDAS) dataset (e.g., [Xia et al. 2012](#)).

MRMS integrates base radar data with satellite, lightning, and rain gauge observations, as well as atmospheric environmental data, using NWP model analyses ([Zhang et al. 2016](#)). MRMS provides a suite of weather and QPE products (e.g., rainfall rate, accumulation, and precipitation type) with a 0.01° (approximately 1 km) resolution. The one used for this study is a rain gauge–corrected product on an hourly basis.

The IFC product is a composite of the seven U.S. Weather Surveillance Radar-1988 Doppler (WSR-88D) radars that cover the full Iowa domain (see [Fig. 1](#)). The IFC acquires Level II radar volume data through Local Data Manager (e.g., [Kelleher et al. 2007](#)), processes the data with its own QPE algorithms, and creates a real-time composite product ([Krajewski et al. 2017](#); [Seo and Krajewski 2020](#)). The IFC recently switched the key QPE algorithm to the specific attenuation-based one (e.g., [Cocks et al. 2019](#); [Wang et al. 2019](#); [Seo et al. 2020c](#)) to take advantage of its capability for full polarimetric observations. The forcing dataset in the mix-and-match configuration includes the IFC's former single-polarization-based (IFC-SP) and the new dual-polarization-based (IFC-DP) estimates. We generated both IFC products using a tool ([Seo et al. 2019](#)) that acquires the Level II data from the Amazon's Big Data archive ([Ansari et al. 2017](#)) and delivers a customized QPE product for the space-time domain of this study. These products are radar-only estimates (not corrected with rain gauge data) with 5-min and approximately 0.5-km resolutions. For this study, we aggregated the estimates to hourly resolution to be compatible with the MRMS product. The resolution and QPE algorithm differences among MRMS and the two IFC products are summarized in [Table 1](#).

Additional forcing data for the Noah-MP LSM encompass incoming shortwave and longwave radiation, specific humidity, air temperature, surface pressure, near-surface wind components, and precipitation rate. We retrieved all these forcing data from the NLDAS dataset at 0.125° resolution, except for precipitation rate, for which we used the radar-based forcing products listed in [Table 1](#). The HLM also uses another forcing data, averaged evapotranspiration (ET) from the Moderate Resolution Imaging Spectroradiometer (MODIS; e.g., [Mu et al. 2011](#)). Monthly averages of the MODIS actual ET over the entire State (see [Fig. 1](#)) for the past 10 years were used for HLM streamflow simulation. The different forcing product spatial resolutions (e.g., MRMS versus IFC) were resampled onto the 1-km LSM grid and employed for NWM simulations, while the HLM is more flexible for forcing product resolution.

c. Rain gauge and streamflow data

To evaluate the precipitation forcing products, we acquired ground reference rain gauge data from the NWS Cooperative Observer Program (COOP; [Mosbacher et al. 1989](#)) network for the period from 2016 to 2018. As illustrated in [Fig. 1a](#), the COOP gauges are well distributed over the study domain, and

TABLE 1. QPE algorithm comparison of precipitation forcing products. All products are primarily based on the WSR-88D radar observations. N/A = not applicable.

QPE Product	Quality control	Algorithm		Ancillary data
		Hydrometeor/precipitation type classification		
MRMS (1h and 1.0 km)	Correlation coefficient (ρ_{hv}) filter and heuristic rules (Zhang et al. 2016)	Convective Stratiform Tropical Snow	$R_{\text{conv}}(Z) = 0.017Z^{0.714}$ $R_{\text{stra}}(Z) = 0.0365Z^{0.625}$ $R_{\text{trop}}(Z) = 0.010Z^{0.833}$ $R_{\text{snow}}(Z) = 0.115Z^{0.5}$ $\max(R_{\text{stra}}, R_{\text{snow}})$ $R(Z) = 0.017Z^{0.714}$	Rapid Refresh model analyses (surface and wet-bulb temperature and freezing-level height) and rain gauge data from the hydrometeorological automated data system (HADS)
IFC-SP (1 h and 0.5 km)	ρ_{hv} and differential phase (Φ_{dp}) filters (Seo et al. 2015)	Warm/cold stratiform N/A	$R(A) = 4120A^{1.03}$ $R(Z) = 0.017Z^{0.714}$ 0.6 $R(Z)$	None
IFC-DP (1 h and 0.5 km)	ρ_{hv} and Φ_{dp} filters combining with vertically integrated liquid (VIL) threshold (Seo et al. 2020)	Rain Dry snow Wet snow Ice Hail	$R(A) = 4120A^{1.03}$ $R(Z) = 0.017Z^{0.714}$ 0.6 $R(Z)$ 2.8 $R(Z)$ $R(K_{dp}) = 29.0K_{dp}^{0.77}$	Rapid Refresh model analyses (temperature sounding)

about 40% of them provide hourly observations. The rest report daily totals measured by human observers (local volunteers). Because of the recognized report timing error of these daily records (see, e.g., Seo et al. 2013), we aggregated the daily values over a longer time span (e.g., annual) and used them to evaluate the precipitation product. We did not include observations from a nationwide network of hourly-basis rain gauges known as the hydrometeorological automated data system (HADS; e.g., Kim et al. 2009) in the reference data to allow for an independent evaluation of MRMS, which contains a bias correction using the HADS data.

We assessed the streamflow simulation results generated from the mix-and-match approach using streamflow data observed at 140 USGS stations, as shown in Fig. 1b. These stations offer reliable, quality-controlled streamflow data with a 15-min interval and cover a wide range of drainage scales, which allows us to develop a multiscale performance evaluation of simulated streamflow. The streamflow records were provided by converting measured water level (stage) into discharge using well-defined rating curves. The USGS has developed these rating curves by periodically measuring stage–discharge relationship, particularly during several low- and high-flow events. We assumed that the USGS rating curves are accurate and did not consider their uncertainty in our analysis.

3. Methodology

In this section, we briefly describe the concept of the mix-and-match approach using multiple precipitation forcing products and hydrologic models. We discuss a modeling framework implemented for a fair comparison/evaluation of streamflow predictions generated from different modeling elements. This section also outlines tactics to assess the accuracy/performance of precipitation forcing products and mix-and-match simulation results (i.e., streamflow predictions) at relevant space and time scales using well-known evaluation metrics.

a. Mix-and-match simulation

The purpose of mix-and-match simulations is to explore all possible combination scenarios of the modeling elements (precipitation forcing products and hydrologic models) and to identify the best forcing-model combination. In this study, we used the MRMS, IFC-SP, and IFC-DP products to drive the HLM and NWM, which results in six forcing-model combinations. To capture the model states for an initial condition, we first performed a continuous simulation for a period from August 2015 to December 2018, using MRMS precipitation forcing for both models. The antecedent period from August 2015 to March 2016 was used to spin up the model states. We then saved the model states on 1 April of each year as the initial conditions for each individual year simulation. We limited our evaluation analysis to the period of April–October in each year because radar-based precipitation estimates for winter months are likely affected by large uncertainties (e.g., Seo et al. 2015; Souverijns et al. 2017). Furthermore, winter precipitation does not immediately contribute to streamflow discharge (e.g., Fontaine et al. 2002). The MRMS-generated initial states were then used for model simulations driven by the two IFC products to avoid

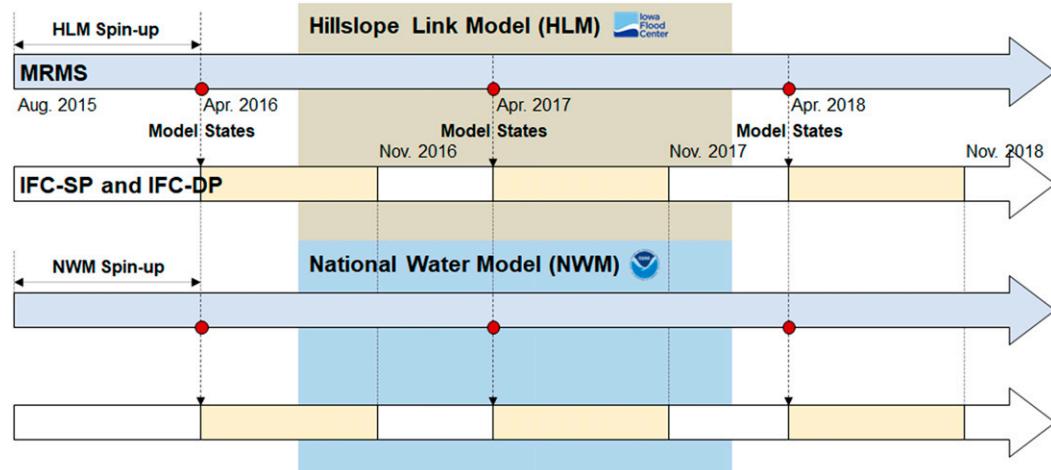


FIG. 2. A schematic view of the mix-and-match modeling framework. Initial states of both models (on 1 Apr, each individual year) for simulations driven by the IFC-SP and IFC-DP products were taken from model simulations driven by MRMS.

any discrepancies generated by different forcing products. Figure 2 demonstrates the mix-and-match modeling strategy, generating consistent initial conditions when comparing the performance of different models driven by multiple precipitation forcing products.

We configured WRF-Hydro (version 5.0.3) as closely as possible to the NWM, which is running at the NWS. As part of this effort, we acquired exactly the same NWM domain grids and associated data for the study area, with the help of the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI). The domain grids and parameters were retrieved from the NWM version 1.2.2 using a CUAHSI-developed tool “domain subsetter” (Castranova et al. 2019) offline. Because the NWM’s upgrade to the current operational version (2.0) focused mainly on spatial and temporal domain expansion, the version discrepancy does not likely engender any major differences in model simulation results.

We note that modeling procedures in this study did not include model parameter calibration and streamflow DA. Calibration of model parameters may conceal prediction uncertainties derived from different precipitation forcing products and prevent us from understanding the propagation of these uncertainties through a hydrologic model. The benefit of streamflow DA in streamflow prediction is documented in F. Quintero et al. (2020, unpublished manuscript) and Seo et al. (2020, manuscript submitted to *J. Amer. Water Resour. Assoc.*), separately for the two models.

b. Evaluation

1) FORCING PRODUCTS

We evaluated the three precipitation forcing products using the COOP rain gauge observations at two temporal scales (e.g., yearly and hourly). To avoid the effect of timing errors in daily readings, we accumulated individual gauge records over a year (April–October) and compared yearly gauge totals with

corresponding estimates collocated in the forcing products. We also performed hourly radar–gauge (R – G) comparisons for those locations where hourly observations are available, as shown in Fig. 1a. We assumed that the gauge representativeness error (e.g., Morrissey et al. 1995) is negligible at the temporal (hourly and yearly) and spatial (MRMS: 1 km; IFC-SP and IFC-DP: 0.5 km) scales analyzed. The independent evaluation at both hourly and yearly scales employs three statistical metrics to quantify the accuracy of precipitation estimates: multiplicative bias (B) defined as a ratio (R/G), Pearson correlation coefficient (r), and mean absolute error (MAE). Detailed formulas for the three metrics are provided in Eqs. (1)–(3):

$$B = \frac{\sum_{i=1, j=1}^{n, t} R_{i,j}}{\sum_{i=1, j=1}^{n, t} G_{i,j}}, \quad (1)$$

$$r = \frac{\sum_{i=1, j=1}^{n, t} (R_{i,j} - \bar{R})(G_{i,j} - \bar{G})}{\sqrt{\sum_{i=1, j=1}^{n, t} (R_{i,j} - \bar{R})^2 \sum_{i=1, j=1}^{n, t} (G_{i,j} - \bar{G})^2}}, \quad (2)$$

$$MAE = \frac{1}{n, t} \sum_{i=1, j=1}^{n, t} |R_{i,j} - G_{i,j}|, \quad (3)$$

where $R_{i,j}$ and $G_{i,j}$ indicate the j th hour/day radar estimate and gauge observation at the i th gauge location, respectively; n and t denote the number of rain gauges and total hours/days in the evaluation period; and \bar{R} and \bar{G} are the radar and gauge mean values.

2) STREAMFLOW PREDICTIONS

We assessed streamflow predictions generated by the mix-and-match simulations at 140 USGS stream gauge stations, as

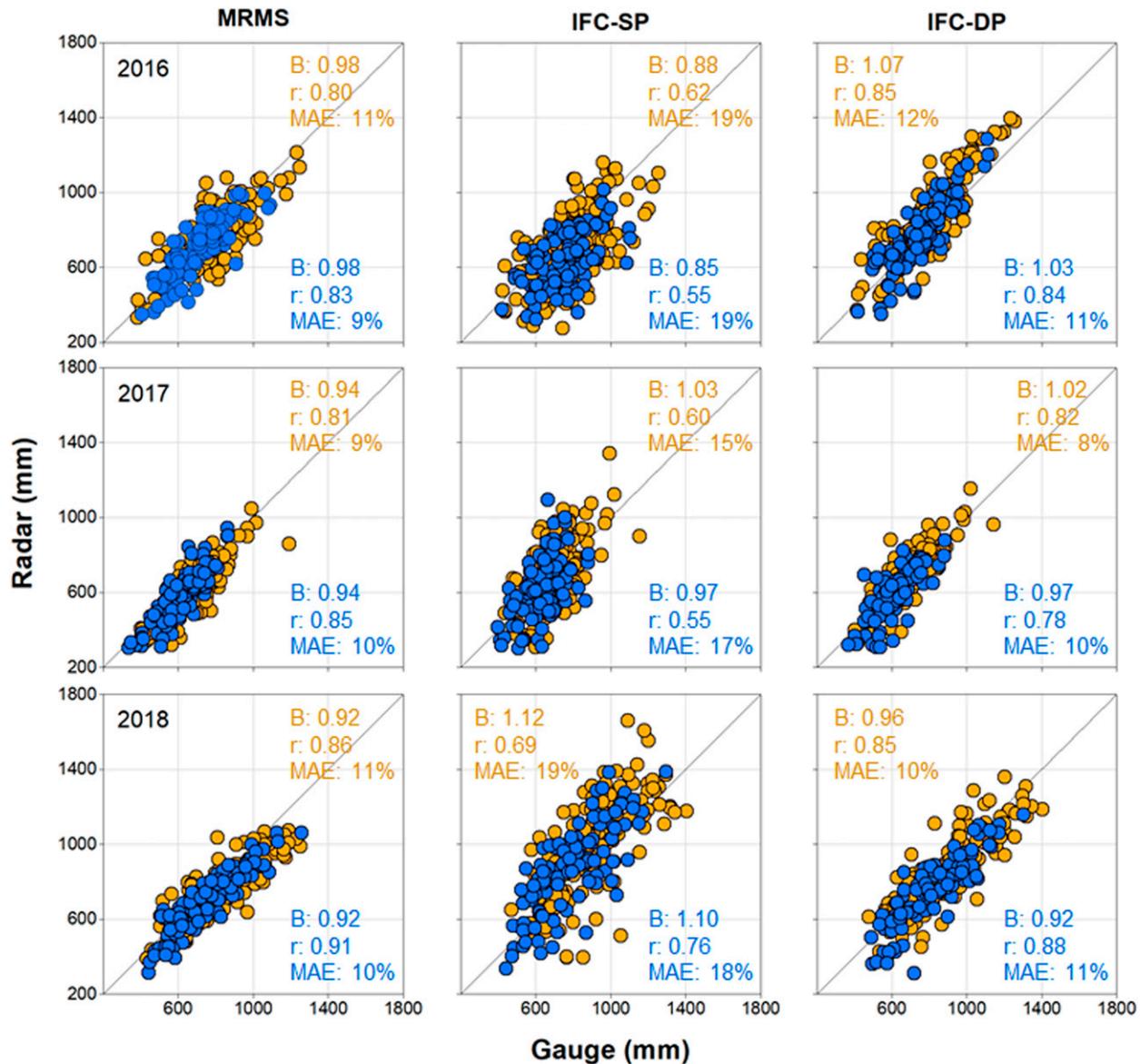


FIG. 3. Quantitative evaluation of the three precipitation forcing products (MRMS, IFC-SP, and IFC-DP) using hourly and daily COOP observations at annual (April–October) total scale. The presented evaluation metrics in each scatterplot are separately color coded for hourly (blue) and daily (orange) rain gauges shown in Fig. 1a. MAE was normalized by the gauge mean and presented as a percentage.

shown in Fig. 1b. The simulated streamflow with a 15-min time interval was analyzed using common statistical evaluation metrics such as correlation (r), MAE, and root-mean-square error (RMSE). The correlation and MAE calculations are similar to those in Eqs. (2) and (3), and RMSE is provided in Eq. (4):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Q_{t,\text{sim}} - Q_{t,\text{obs}})^2}, \quad (4)$$

where Q_t denotes observed (obs) and model simulated (sim) discharge at time t , and n indicates the number of time steps.

We further refined prediction performance based on the major hydrologic indicators such as runoff volume and peak flow, with respect to different combinations and upstream drainage scales. This analysis permitted us to define prediction uncertainties (or performance) as a function of such variables as basin scale and regional landscape/geology. Iowa's diverse landscape is represented by several major landform types as presented in Fig. 1b. The evaluation metrics that expose the central features of hydrologic prediction include (i) relative volume error (RE_V), (ii) relative peak error (RE_{Q_p}), (iii) peak timing error (E_p), and (iv) Kling–Gupta efficiency (KGE). KGE (Gupta et al. 2009) describes an overall predictive power

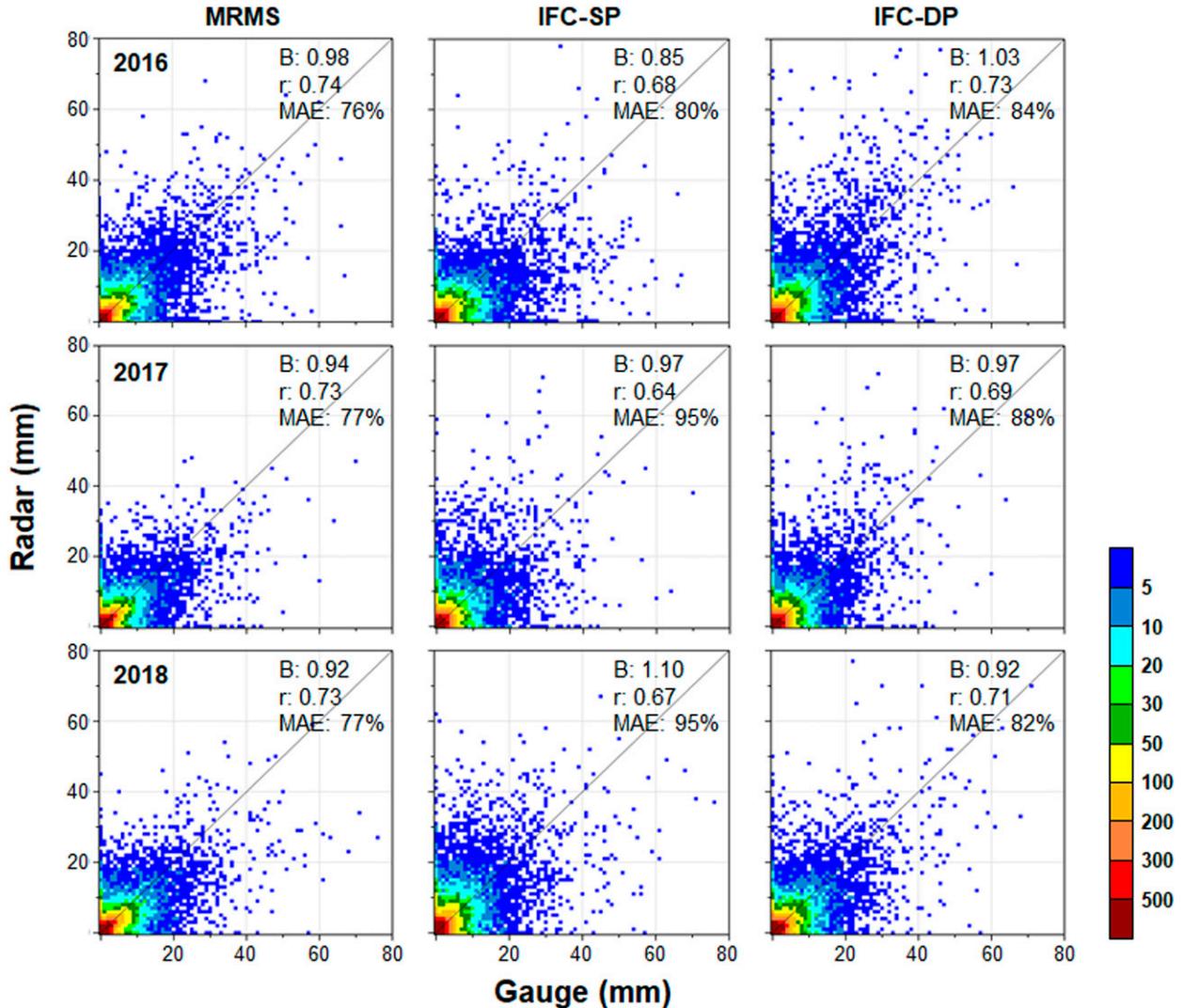


FIG. 4. Precipitation forcing product evaluation at hourly scale. The 2D histograms show the number of occurrences for given rainfall ranges of radar-gauge pairs.

of hydrologic models and was proposed to improve deficiencies in Nash–Sutcliffe efficiency (Nash and Sutcliffe 1970). The four metrics are defined in Eqs. (5)–(8):

$$\text{RE}_V = \frac{V_{\text{sim}} - V_{\text{obs}}}{V_{\text{obs}}} \times 100\%, \quad (5)$$

$$\text{RE}_{Q_p} = \frac{Q_{p,\text{sim}} - Q_{p,\text{obs}}}{Q_{p,\text{obs}}} \times 100\%, \quad (6)$$

$$E_{t_p} = t_{p,\text{sim}} - t_{p,\text{obs}}, \quad (7)$$

$$\text{KGE} = 1.0 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2}, \quad (8)$$

where V , Q_p , and t_p denote total volume (m^3), peak discharge ($\text{m}^3 \text{s}^{-1}$), and peak time (h) obtained from model simulations (sim) and observations (obs) during the period from April to October of each year. KGE is represented as a function of

correlation (r), the ratio of standard deviation (α), and the ratio of mean (β) between simulated and observed streamflow.

4. Results

a. Evaluation of precipitation forcing products

We evaluated three precipitation forcing products employed in the mix-and-match simulations using ground reference data attained from the NWS COOP network within the study domain illustrated in Fig. 1a. In the domain, there are 289 COOP gauges in total (109 hourly and 180 daily gauges), and we accumulated precipitation records from these gauges for the period from April to October for a yearly evaluation. Figure 3 shows the yearly evaluation of MRMS, IFC-SP, and IFC-DP products with three statistical metrics defined in Eqs. (1)–(3). The MAE values shown in Fig. 3 were normalized by the gauge mean and presented as a percentage. The dots represent

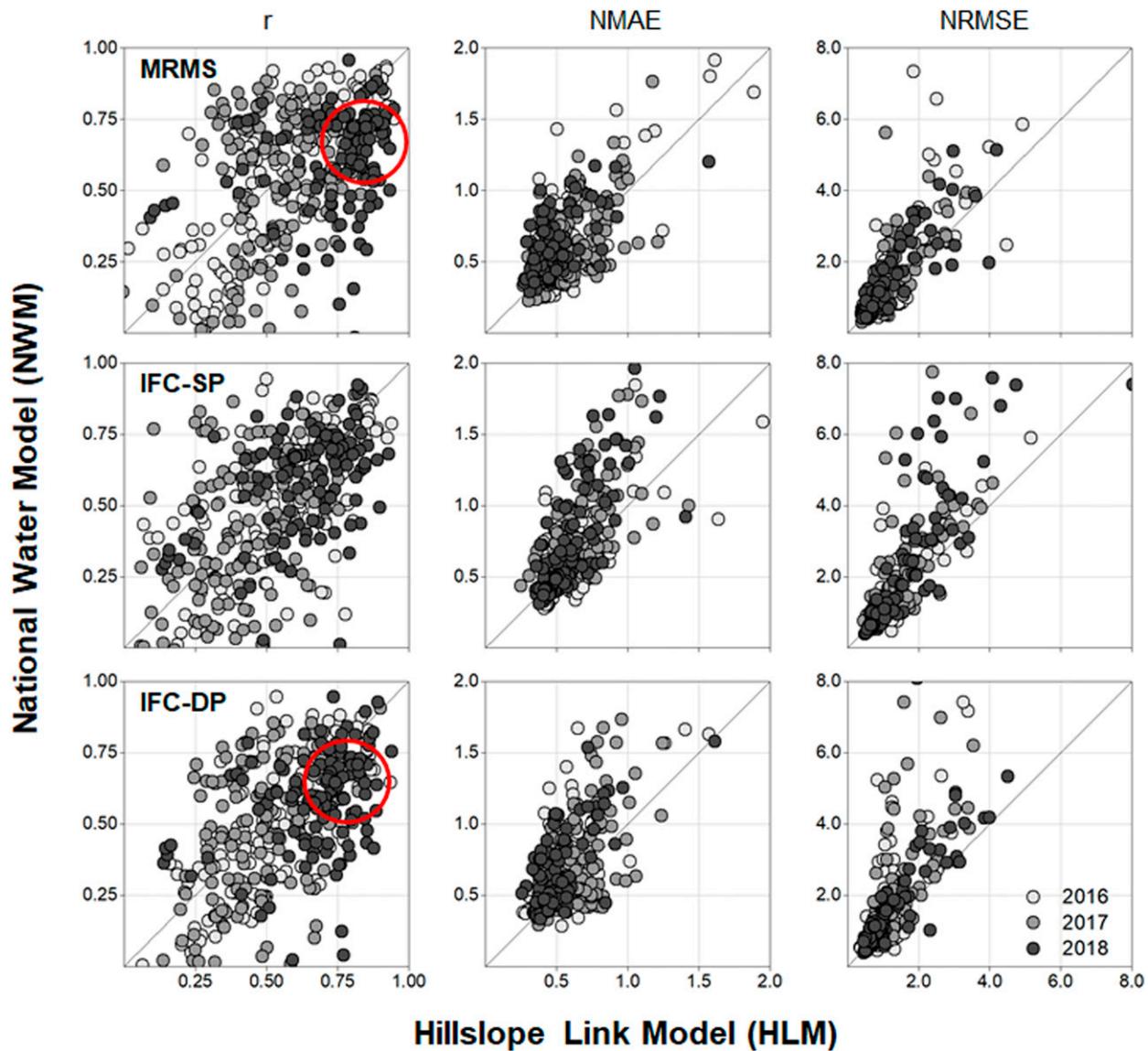


FIG. 5. Performance comparison of the mix-and-match simulation results characterized by correlation (r), MAE, and RMSE. MAE and RMSE were normalized by annual mean streamflow at individual USGS stations. The solid red circles in the correlation plots indicate a high-density cluster in 2018 to compare the performance of MRMS and IFC-DP.

individual gauge and corresponding radar grid totals with the same color code used in Fig. 1a to distinguish hourly (blue) and daily (orange) gauges. As shown in Fig. 3, the dots indicating MRMS and IFC-DP tend to closely align along the one-to-one line, with slightly different degrees of linear dependence and dispersion, while the tendency of the dots from IFC-SP looks slanted from the line with relatively larger dispersion. We note that MRMS and IFC-DP show quite comparable performance, although IFC-DP does not contain a bias correction using rain gauge records, which is included in the MRMS product. The estimates of IFC-DP shown in Fig. 3 were derived from the specific attenuation algorithm (Seo et al. 2020c); this reveals significant accuracy improvement against its predecessor, the reflectivity-based estimates IFC-SP. Overall, the performance

of MRMS looks slightly better than that of IFC-DP, although its bias values for all three years denote consistent underestimations (i.e., $B < 1.0$). The behavior of bias seems to be better for IFC-DP when compared to MRMS. Comparisons of rain gauge records from different time scales reveal that the yearly totals from daily gauges show larger dispersion than those from hourly ones. We speculate that this is due to the errors in human reading (daily) versus automatic sensing (hourly). The hourly evaluation results are presented in Fig. 4 with two-dimensional histograms representing the occurrence of hourly values between gauge observations and radar-based estimates. The observed degrees of overall bias for each product shown in Fig. 3 are consistent with those in Fig. 4. Figure 4 clearly shows that: (i) the dispersion of IFC-DP is somewhat larger than that

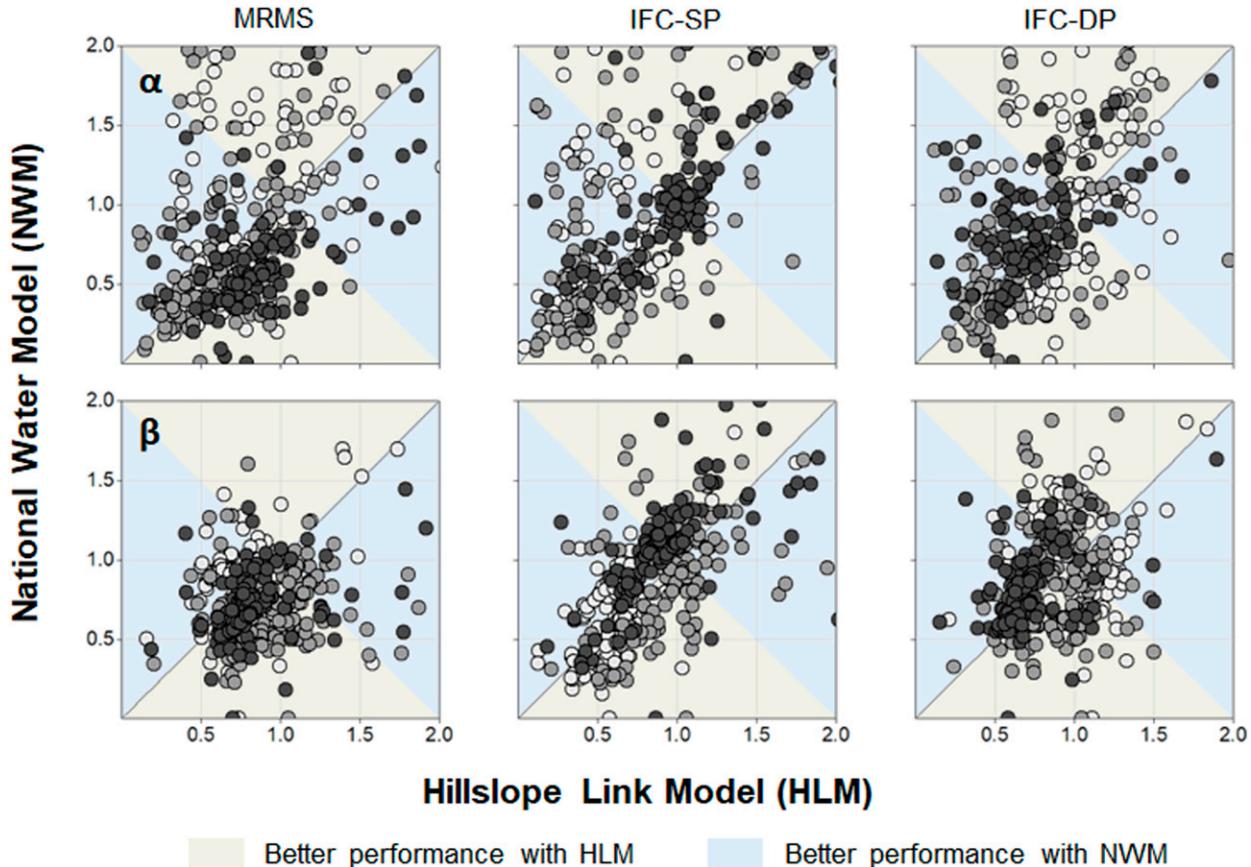


FIG. 6. Performance comparison of the mix-and-match simulation results characterized by the ratios of standard deviation (α) and mean (β). The two different colored zones indicate each model's superiority. The dots are color coded regarding different analysis years as labeled in Fig. 5.

of MRMS at hourly scale; and (ii) the three products are characterized by different uncertainty features conditioned on rainfall magnitude (e.g., Ciach et al. 2007; Seo et al. 2018).

b. Evaluation of mix-and-match simulations

We generated streamflow predictions from six mix-and-match combinations using three precipitation forcing products (MRMS, IFC-SP, and IFC-DP) and two distributed hydrologic models (HLM and NWM). Figures 5 and 6 show basic statistical metrics (correlation, MAE, RMSE, and the ratios of standard deviation and mean) for all six simulation evaluations from 140 USGS stations, as shown in Fig. 1b for the study years (2016–18). MAE and RMSE were normalized by the annual mean of observed streamflow for each year to offer practical insight regarding the degree of errors. For model performance comparison, there are more dots placed below the one-to-one line in correlation with MAE and RMSE, showing an opposite tendency for all forcing products. This implies the superiority of HLM in streamflow prediction. The ratios of standard deviation (α) and mean (β) shown in Fig. 6 also reveal that the distributions of high-density clusters with NWM are wider, and their centers appear farther from the unity (1.0) axis than HLM's do (the closer to unity for α and β , the better the

agreement with observations). Concerning forcing product comparisons represented by streamflow simulations, MRMS and IFC-DP seem to perform better than IFC-SP, as shown by the forcing product evaluation results presented in Figs. 3 and 4. IFC-SP shows more dots in a low correlation zone (e.g., $r < 0.5$ for both models) and fewer dots in a low error zone (e.g., $\text{NMAE} < 0.5$ for both models). It is not clear whether MRMS or IFC-DP performs better based on the presented metrics in Figs. 5 and 6. While the scatter of IFC-DP seems slightly smaller in correlation and errors (NMAE and NRMSE), the high-density cluster in 2018 marked by the solid red circles illustrates that MRMS simulations, particularly with HLM, agree slightly better with streamflow observations (i.e., higher correlation).

In Fig. 7, we organized the same results (for 2016 only) presented in Figs. 5 and 6, with respect to upstream drainage/catchment scale covered by individual USGS stations. This rearrangement enables inspection of forcing product and model performance in streamflow generation at a variety of basin scales. MAE and RMSE, presented in Fig. 7, were not normalized to disclose distinct scaling behavior, with larger errors as the scale increases. One can also recognize from Fig. 7 that correlation and the variability of α and β tend to gradually

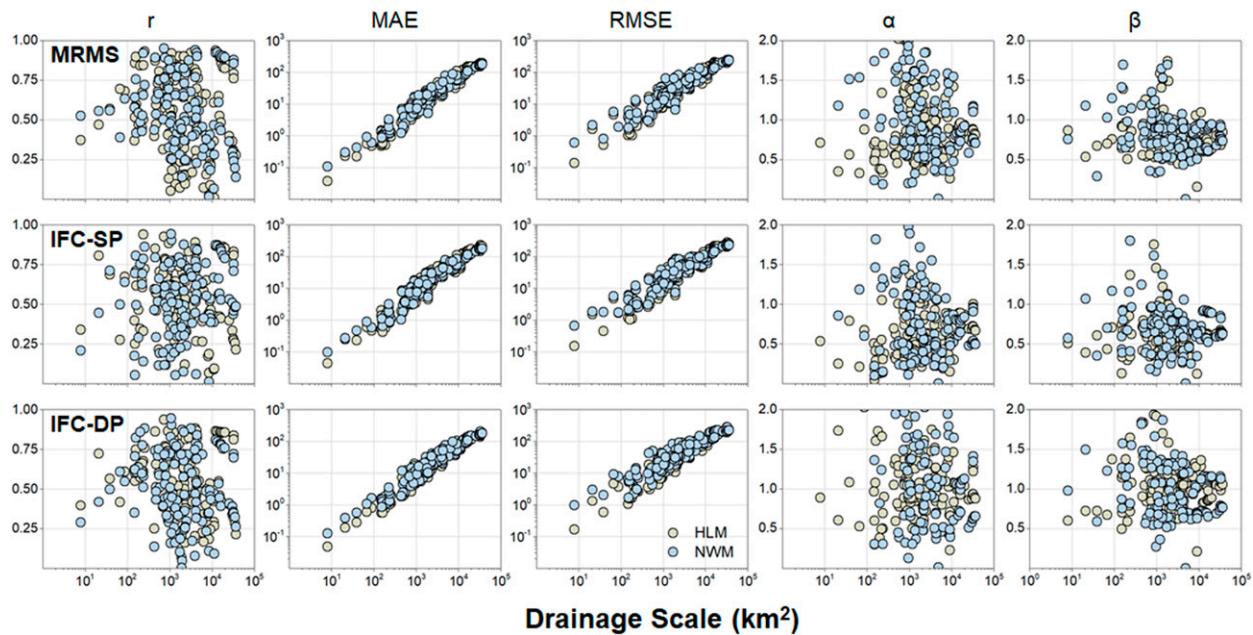


FIG. 7. Performance comparison of the mix-and-match simulation results regarding drainage scale covered by individual USGS stations. The results for 2016 shown in Fig. 5 were reorganized.

increase and decrease, respectively, as drainage scale becomes larger. Because the superiority of forcing product and model performance changes from year to year, it is hard to decide which product or model generates better streamflow predictions based on the results shown in Fig. 7. Observations of α and β gleaned from Figs. 5–7 include the following: (i) the dispersion of NWM results looks wider than that of HLM results (several α values for NWM at small scales are not within the presented range) and (ii) the center of HLM clusters for all forcing products are closer to unity. Figure 8 shows examples of observed and simulated hydrographs, with estimated α and

β values at different basin scales. In Fig. 8, one of α and β for both models is in a good range (close to unity) at each different location, whereas the other shows performance differences. Good estimations of β at Redfield were contributed by the erroneous peak (compensating early misses in April–May) detected in August, which decreased correlation significantly. HLM's underestimation in β at Wapello seems to arise from (i) the early recession during the peak event detected in late September and (ii) the initial condition and simulated discharge during an early period (April–mid-June) lower than the observed. Although NWM's result at Wapello also revealed

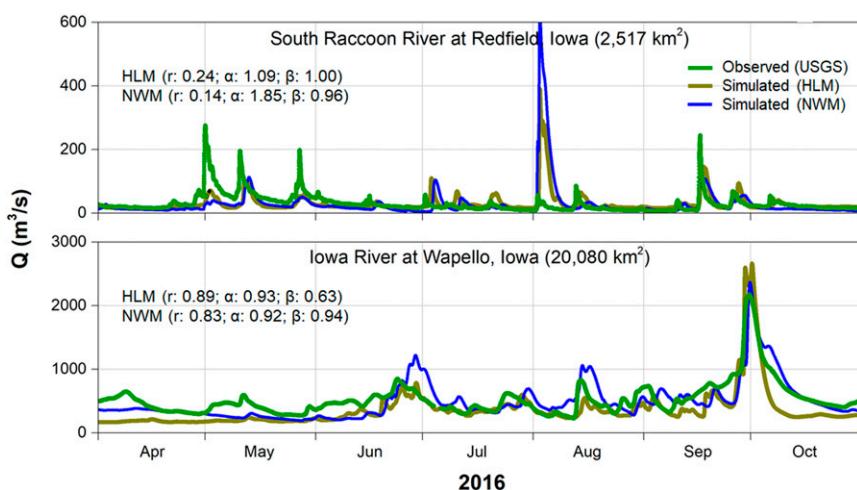


FIG. 8. Example cases showing observed (USGS) and simulated hydrographs (HLM and NWM) at Redfield (USGS 05484000) and Wapello (USGS 05465500) in Iowa.

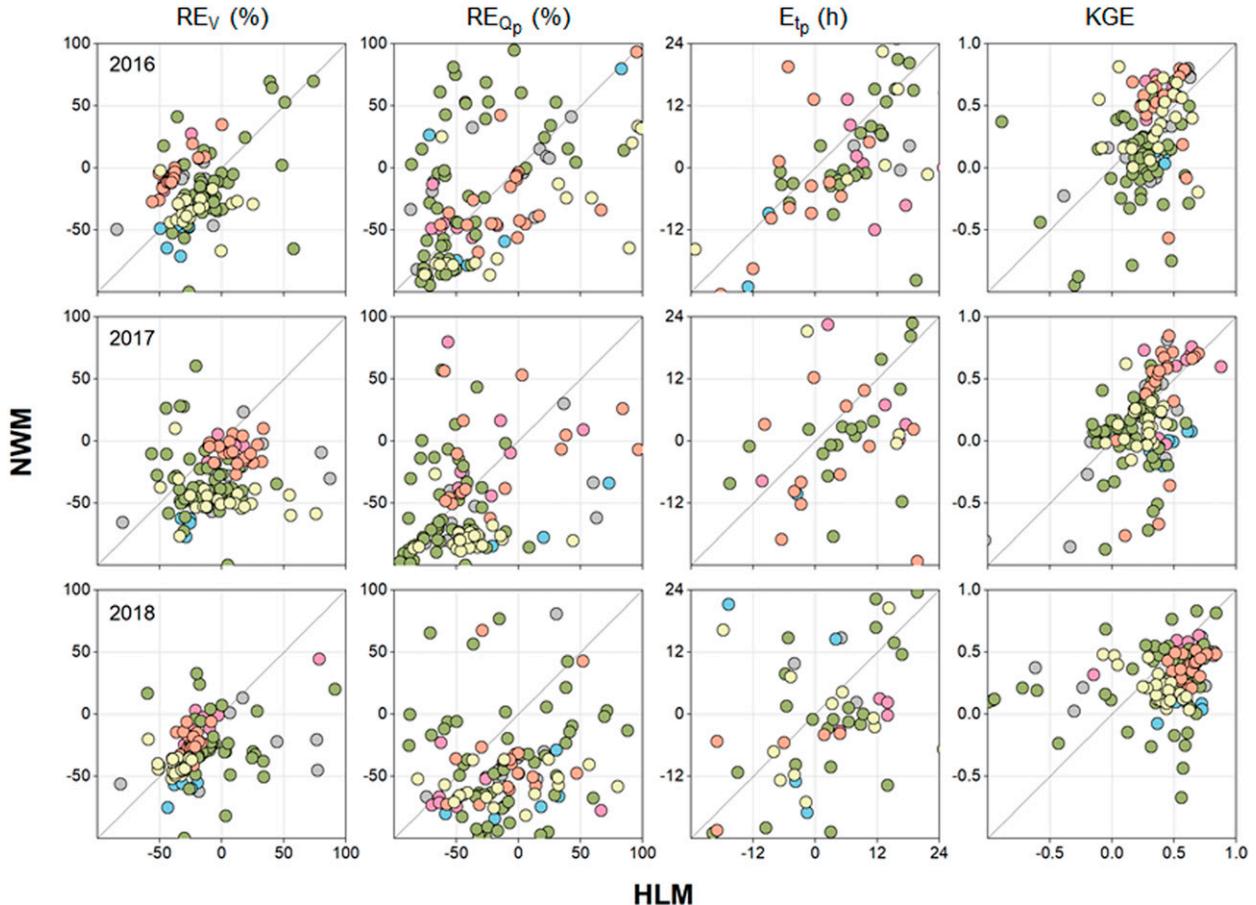


FIG. 9. Model performance comparison driven by MRMS based on hydrologic evaluation metrics defined in Eqs. (5)–(8). The metrics were color coded by different landforms using the same colors shown in Fig. 1b. Minor landform regions (e.g., east-central Iowa drift plain, Iowa-Cedar lowland, Losses hills, and Missouri River alluvial plain) into which only few USGS stations are assigned are indicated by the gray dots.

underestimated discharge during the same period, several overestimations from June to September compensated for the early misses. Correlation, α , and β are the factors that determine the overall performance of streamflow prediction represented by KGE defined in Eq. (8). The next few figures present the estimated KGE to further assess the mix-and-match simulation results, along with significant hydrologic features associated with runoff volume and peak discharge.

We further refined the evaluation results using the performance metrics defined in Eqs. (5)–(8), which allow us to examine major hydrologic aspects of simulated streamflow. Model simulations driven by MRMS and IFC-DP are compared in Figs. 9 and 10, respectively. We excluded IFC-SP in this analysis because of its relatively low accuracy, discovered in Figs. 3 and 5. To determine relative peak (RE_{Q_p}) and peak timing (E_{tp}) errors, we identified a model-simulated peak within the same event period containing an annual peak observed at each individual USGS station (hydrologic models occasionally generate an annual peak at a completely different time than the actual peak time observed). We color coded the calculated metrics according to different landform types; the

corresponding colors are illustrated in Fig. 1b. As shown in Figs. 9 and 10, no landform type shows consistent performance with both models; neither model seems to favor a specific type of landscape or geology in describing surface or subsurface processes. For example, the relative performance of the two models varies (above and below the one-to-one line) from year to year based on the KGE values for the “Des Moines Lobe” and “Iowan Surface” landforms covering north-central Iowa. In Fig. 9, the HLM seems slightly better in estimating runoff volume and peak discharge than the NWM does: (i) more dots are closer to the vertical 0% (no error) line, indicating that the HLM better estimates these quantities, and (ii) the NWM shows underestimations (dots below the horizontal 0% line) in RE_V and RE_{Q_p} at most USGS locations. IFC-DP forcing (Fig. 10) significantly improved the underestimation tendency observed in Fig. 9 and “redeployed” many dots in RE_V and RE_{Q_p} into the positive area, particularly for 2016 and 2018. Given the results shown in Figs. 9 and 10, we acknowledge that the estimation of peak (time and magnitude) is more challenging than that of volume: (i) the variability of RE_{Q_p} and E_{tp} looks much larger than that of RE_V and

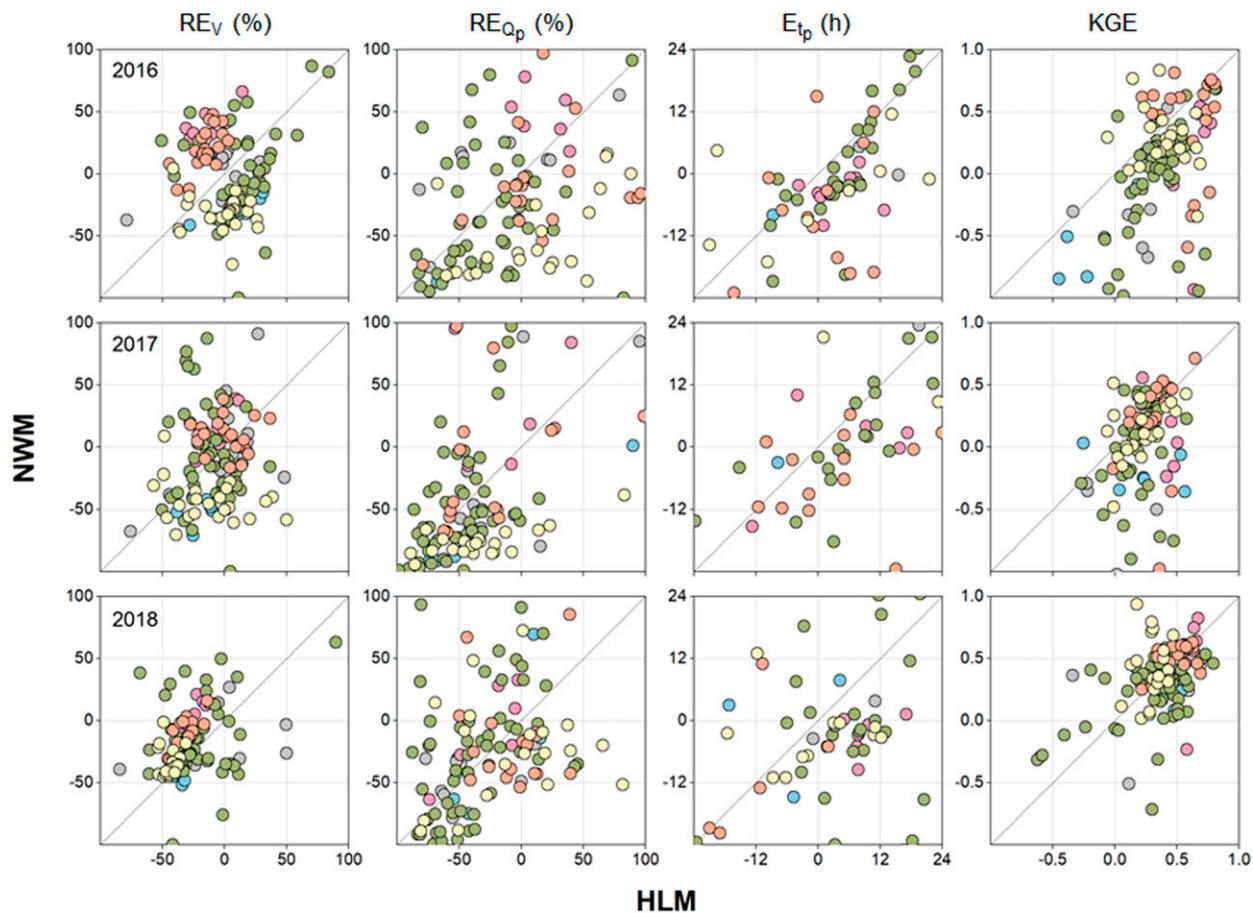


FIG. 10. As in Fig. 9, but driven by IFC-DP.

(ii) model predictions at many locations miss a presented range (± 24 h) for E_{tp} . Overall, the HLM shows slightly better performance for both MRMS and IFC-DP forcing, considering the greater number of KGE dots below the one-to-one line in Figs. 9 and 10.

To describe scale-dependent performance of the forcing-model combinations, we rearranged the hydrologic evaluation results presented in Figs. 9 and 10 regarding drainage scale. Figures 11 and 12 demonstrate scale-dependent features of the evaluation metrics for respective HLM and NWM prediction results driven by the MRMS and IFC-DP products. Because we found that the prediction results of HLM were superior to those of NWM from Figs. 9 and 10, the analysis shown in Figs. 11 and 12 focuses on comparing prediction results driven by different forcing products. In Figs. 11 and 12, IFC-DP leads to increased runoff volumes compared to the volumes generated by MRMS, which consistently stay below the 0% line, indicating underestimations. As we discussed earlier, peak estimation is quite challenging. It is difficult to distinguish which forcing product performs better in capturing the observed peaks. The highly scattered patterns shown in both RE_{Qp} and E_{tp} are likely to diminish their scale-dependent features, which are clearly exposed in RE_V and KGE. As drainage scale increases, the variability of RE_V and KGE tends

to decrease and gradually approach ideal conditions (i.e., 0% for RE_V and 1.0 for KGE). Figure 11 shows few exceptional cases with some negative KGE values (<-0.5) at medium- and large-scale basins (e.g., in 2018). We found that these points are located downstream from reservoirs; both HLM and NWM did not include reservoir routing and controls to simplify model configuration and implementation. The results from NWM (Fig. 12) at these locations also resulted in low performance (e.g., negative KGE for medium-scale basins and slightly above 0.0 for large-scale basins). Reservoir controls vary from year to year depending on different circumstances (e.g., flood control or irrigation) and may lead to inconsistent prediction performance at these locations during the years shown in Figs. 11 and 12.

5. Conclusions

This study reports the assessment results of streamflow predictions generated by mix-and-match combinations using three precipitation forcing data (MRMS, IFC-SP, and IFC-DP) and two hydrologic models (HLM and NWM). All three forcing data are radar-based gridded products, which offer spatially variable information to activate grid-based processes and routing realized in the two distributed models. We evaluated

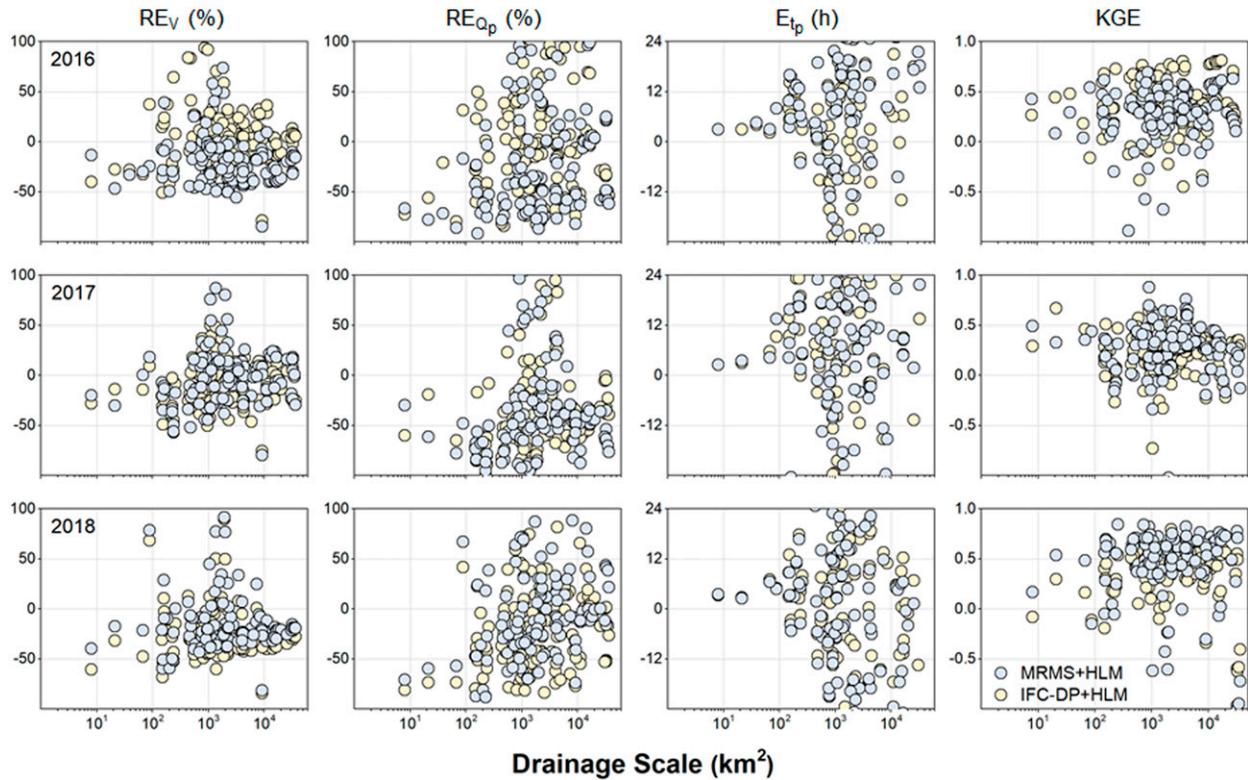


FIG. 11. HLM performance comparison resulted from different forcing products (MRMS vs IFC-DP) regarding drainage scale.

these forcing products and model-generated streamflows using rainfall and streamflow observations acquired from about 289 NWS COOP rain gauges and 140 USGS stations. The objectives of this assessment are (i) to improve our understanding of forcing- and model-dependent prediction capabilities and (ii) to define the best forcing–model combination and understand the reasons.

Forcing product evaluation presented in Fig. 3 demonstrated that one of the radar-only products (i.e., IFC-DP) is comparable with the gauge-corrected one (i.e., MRMS). This implies that the state-of-the-art polarimetric estimation (e.g., Wang et al. 2019; Seo et al. 2020c) significantly improved QPE accuracy against the widespread reflectivity-based estimation (e.g., Fulton et al. 1998). It also enables the application of finer temporal resolution forcing (e.g., 15 or 30 min) with greater accuracy to hydrologic applications (gauge-corrected ones are barely available at these scales). The resulting effects of forcing data resolution are closely related to the scale of the basin being simulated (e.g., Aronica et al. 2005; Lyu et al. 2018). For a bias perspective, IFC-DP looked closer to the reference data, and MRMS consistently showed slight underestimations. Interestingly, the dispersion shown in R – G comparisons (Figs. 3 and 4) appeared smaller for MRMS. We will discuss this tendency of bias and dispersion between MRMS and IFC-DP and its effects on the errors in streamflow prediction.

Based on our extensive comparison analyses, we conclude that the HLM performs slightly better in streamflow generation than the NWM does: the runoff volume errors of HLM

shown in Fig. 11 are more closely distributed to the 0% line, with smaller variability than those of NWM in Fig. 12. As such, we conjecture that the modeling elements of HLM addressing precipitation losses (e.g., evapotranspiration and subsurface process) tend to better capture what happens in nature. HLM’s modeling element for precipitation losses based on the concept of linear reservoirs (see, e.g., Quintero et al. 2016) is much simpler than detailed land surface processes described by Noah-MP (Niu et al. 2011) in the NWM. This indicates that complex modeling of land surface processes using additional data resources (e.g., NLDAS in this study) implemented in the NWM does not necessarily lead to more accurate streamflow generation. The snowmelt process (e.g., Fontaine et al. 2002) is likely not included in the analyzed water volumes because our analysis excluded the winter months. The evaluation results regarding peak time and discharge are combined effects of such complex processes as the aforementioned losses, and surface, subsurface, and channel routing. As shown in Figs. 9–12, it is not clear how the process components included in the two models differ in describing the simulated peaks because of the massive dispersion observed in peak errors. One obvious observation from the comparison between Figs. 11 and 12 is the relatively smaller peaks with the NWM regardless of forcing data. We note that our prior research comparing the two models, e.g., channel routing schemes (ElSaadani et al. 2018) and representation of river network (Quintero and Krajewski 2018), partially accounts for the observed differences in the models’ performance.

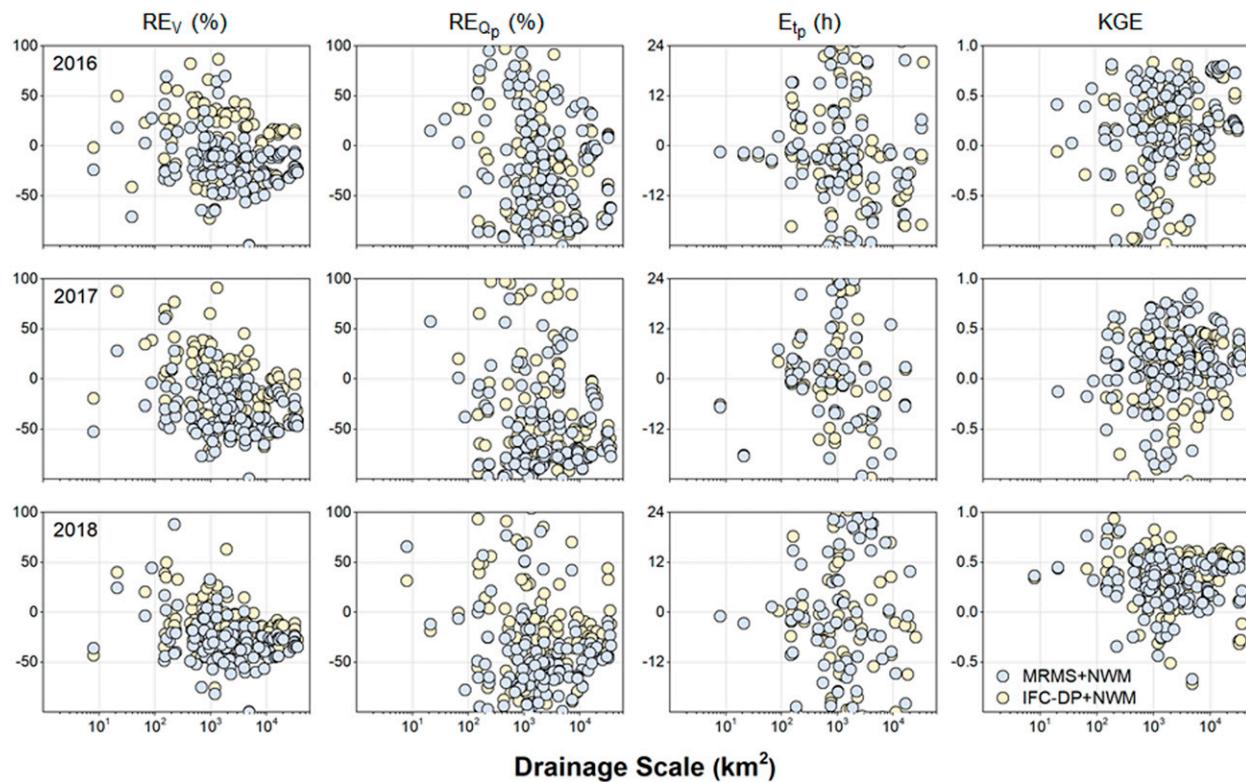


FIG. 12. NWM performance comparison resulted from different forcing products (MRMS vs IFC-DP) regarding drainage scale.

We discovered that bias in precipitation forcing products is closely related to the volume of model-generated streamflow: slight underestimations of MRMS presented in Fig. 3 led to streamflow underestimations for all three years in the NWM simulation, while the less biased forcing product (i.e., IFC-DP) resulted in increased volumes (see Fig. 12). The observed variability (dispersion) of the forcing product errors is also likely associated with model performance. Fewer variable errors in MRMS show smaller scatter in both model simulations, as shown in Figs. 11 and 12. This scatter seems to decrease as scale becomes larger (e.g., volume error and KGE) because random errors in precipitation tend to average out at larger-scale basins (e.g., Vivoni et al. 2007; Cunha et al. 2012). Given our evaluation and discussion, we concluded that MRMS performs slightly better than IFC-DP, focusing on the behavior of KGE (with some performance variations observed from year to year). Therefore, we selected the combination of MRMS and HLM as the best mix-and-match set in this study.

As we demonstrated in Figs. 7, 11, and 12, there are not many evaluation points at the smaller scale (e.g., $<1000 \text{ km}^2$), which is particularly useful for flash flood forecasting (e.g., Gourley et al. 2013). This scale gap could be addressed by about 280 stage-only sensors (Kruger et al. 2016) managed by the IFC to monitor streams and creeks near Iowa communities. To use these sensor measurements in a variety of hydrologic applications, the IFC has developed a framework to build synthetic rating curves (Quintero et al. 2021).

We will soon incorporate the data from these sensors into our evaluation to fill the significant scale gap, as well as into HLM and NWM configuration to improve and expand streamflow prediction capability using streamflow DA (e.g., Seo et al. 2020, manuscript submitted to *J. Amer. Water Resour. Assoc.*).

We recognize that recently proposed approaches can improve hydrologic predictions by (i) averaging of multiple precipitation forcing data (Schreiner-McGraw and Ajami 2020) and (ii) averaging of simulated outputs generated from multiple models and forcing data (Zhu et al. 2019). We plan to explore and test these approaches to understand how they affect prediction skills, particularly in small scale basins where the temporal and spatial variability plays a primary role in streamflow generation.

Acknowledgments. This study was supported by the Hydrometeorology Testbed (HMT) Program within NOAA/OAR Office of Weather and Air Quality under Grant NA17OAR4590131. The authors are grateful to Dr. Anthony Castranova at CUAHSI for providing the NWM grids and parameters and to Drs. Aubrey Dugger and David Gochis at the National Center for Atmospheric Research for guidance on our NWM implementation.

REFERENCES

- Ansari, S., and Coauthors, 2017: Unlocking the potential of NEXRAD data through NOAA's big data partnership. *Bull. Amer. Meteor. Soc.*, **99**, 189–204, <https://doi.org/10.1175/BAMS-D-16-0021.1>.

- Aronica, G., G. Freni, and E. Oliveri, 2005: Uncertainty analysis of the influence of rainfall time resolution in the modelling of urban drainage systems. *Hydrol. Processes*, **19**, 1055–1071, <https://doi.org/10.1002/hyp.5645>.
- 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](https://doi.org/10.1016/0309-1708(93)90028-E).
- Castranova, A. M., and Coauthors, 2019: Improving access to continental-scale hydrology models for research and education—A subsetting adventure. *2019 AGU Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract H43I-2146.
- Ciach, G. J., W. F. Krajewski, and G. Villarini, 2007: Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data. *J. Hydrometeor.*, **8**, 1325–1347, <https://doi.org/10.1175/2007JHM814.1>.
- Cocks, S. B., and Coauthors, 2019: A prototype quantitative precipitation estimation algorithm for operational S-band polarimetric radar utilizing specific attenuation and specific differential phase. Part II: Performance verification and case study analysis. *J. Hydrometeor.*, **20**, 999–1014, <https://doi.org/10.1175/JHM-D-18-0070.1>.
- Cohen, S., S. Praskiewicz, 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>.
- Cole, S. J., and R. J. Moore, 2009: Distributed hydrological modelling using weather radar in gauged and ungauged basins. *Adv. Water Resour.*, **32**, 1107–1120, <https://doi.org/10.1016/j.advwatres.2009.01.006>.
- Cosgrove, B., and Coauthors, 2016: An overview of the National Weather Service National Water Model. *2016 AGU Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract H42B-05.
- Cunha, L. K., P. V. Mandapaka, W. F. Krajewski, R. Mantilla, and A. A. Bradley, 2012: Impact of radar-rainfall error structure on estimated flood magnitude across scales: An investigation based on a parsimonious distributed hydrological model. *Water Resour. Res.*, **48**, W10515, <https://doi.org/10.1029/2012WR012138>.
- Demir, I., and W. F. Krajewski, 2013: Towards an integrated flood information system: Centralized data access, analysis, and visualization. *Environ. Model. 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. *Hydrol. Processes*, **16**, 255–275, <https://doi.org/10.1002/hyp.338>.
- ElSaadani, M., W. F. Krajewski, R. Goska, and M. B. Smith, 2018: An investigation of errors in distributed models' stream discharge prediction due to channel routing. *J. Amer. Water Resour. Assoc.*, **54**, 742–751, <https://doi.org/10.1111/1752-1688.12627>.
- Fontaine, T. A., T. S. Cruickshank, J. G. Arnold, and R. H. Hotchkiss, 2002: Development of a snowfall–snowmelt routine for mountainous terrain for the Soil Water Assessment Tool (SWAT). *J. Hydrol.*, **262**, 209–223, [https://doi.org/10.1016/S0022-1694\(02\)00029-X](https://doi.org/10.1016/S0022-1694(02)00029-X).
- Fulton, R. A., J. P. Breidenbach, D.-J. Seo, D. A. Miller, and T. O'Bannon, 1998: The WSR-88D rainfall algorithm. *Wea. Forecasting*, **13**, 377–395, [https://doi.org/10.1175/1520-0434\(1998\)013<0377:TWRA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1998)013<0377:TWRA>2.0.CO;2).
- Ghimire, G. R., W. F. Krajewski, and R. Mantilla, 2018: A power law model for river flow velocity in Iowa basins. *J. Amer. Water Resour. Assoc.*, **54**, 1055–1067, <https://doi.org/10.1111/1752-1688.12665>.
- Gochis, D. J., and Coauthors, 2018: The 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>.
- Gourley, J. J., and Coauthors, 2013: A unified flash flood database across the United States. *Bull. Amer. Meteor. Soc.*, **94**, 799–805, <https://doi.org/10.1175/BAMS-D-12-00198.1>.
- 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>.
- Gupta, V. K., R. Mantilla, B. M. Troutman, D. Dawdy, and W. F. Krajewski, 2010: Generalizing a nonlinear geophysical flood theory to medium size river basins. *Geophys. Res. Lett.*, **37**, L11402, <https://doi.org/10.1029/2009GL041540>.
- , T. Ayalew, R. Mantilla, and W. F. Krajewski, 2015: Classical and generalized Horton laws for peak flows in rainfall-runoff events. *Chaos*, **25**, 075408, <https://doi.org/10.1063/1.4922177>.
- Hansen, C., J. Shafiei Shiva, S. McDonald, and A. Nabors, 2019: Assessing retrospective National Water Model streamflow with respect to droughts and low flows in the Colorado River basin. *J. Amer. Water Resour. Assoc.*, **55**, 964–975, <https://doi.org/10.1111/1752-1688.12784>.
- Kelleher, K. E., and Coauthors, 2007: A real-time delivery system for NEXRAD Level II data via the internet. *Bull. Amer. Meteor. Soc.*, **88**, 1045–1058, <https://doi.org/10.1175/BAMS-88-7-1045>.
- Kim, D., B. Nelson, and D.-J. Seo, 2009: Characteristics of reprocessed hydrometeorological automated data system (HADS) hourly precipitation data. *Wea. Forecasting*, **24**, 1287–1296, <https://doi.org/10.1175/2009WAF222227.1>.
- 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>.
- Lyu, H., G. Ni, X. Cao, Y. Ma, and F. Tian, 2018: Effect of temporal resolution of rainfall on simulation of urban flood processes. *Water*, **10**, 880, <https://doi.org/10.3390/w10070880>.
- Mantilla, R., and V. K. Gupta, 2005: A GIS numerical framework to study the process basis of scaling statistics in river networks. *IEEE Geosci. Remote Sens. Lett.*, **2**, 404–408, <https://doi.org/10.1109/LGRS.2005.853571>.
- McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Rea, 2012: NHDPlus Version 2: User Guide. U.S. Environmental Protection Agency, 173 pp., https://nctc.fws.gov/courses/references/tutorials/geospatial/CSPT306/Readings/NHDPlusV2_User_Guide.pdf.
- Morrissey, M. L., J. A. Maliekal, J. S. Greene, and J. Wang, 1995: The uncertainty of simple averages using rain gauge networks. *Water Resour. Res.*, **31**, 2011–2017, <https://doi.org/10.1029/95WR01232>.
- Mosbacher, R., W. E. Evans, and E. W. Friday Jr., 1989: Cooperative station observations. National Weather Service Observing Handbook No. 2, NOAA, 83 pp.
- Mu, Q., M. Zhao, and S. W. Running, 2011: Improvements to a MODIS global terrestrial evapotranspiration algorithm.

- Remote Sens. Environ.*, **115**, 1781–1800, <https://doi.org/10.1016/j.rse.2011.02.019>.
- 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](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., and W. F. Krajewski, 2018: Mapping outlets of Iowa Flood Center and National Water Center river networks for hydrologic model comparison. *J. Amer. Water Resour. Assoc.*, **54**, 28–39, <https://doi.org/10.1111/1752-1688.12554>.
- , —, R. Mantilla, S. Small, and B.-C. Seo, 2016: A spatial-dynamical framework for evaluation of satellite rainfall products for flood prediction. *J. Hydrometeor.*, **17**, 2137–2154, <https://doi.org/10.1175/JHM-D-15-0195.1>.
- , —, 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**, 1–12, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0002022](https://doi.org/10.1061/(ASCE)HE.1943-5584.0002022).
- Reed, S., V. Koren, M. Smith, Z. Zhang, F. Moreda, and D.-J. Seo, 2004: Overall distributed model intercomparison project results. *J. Hydrol.*, **298**, 27–60, <https://doi.org/10.1016/j.jhydrol.2004.03.031>.
- Schreiner-McGraw, A. P., and H. Ajami, 2020: Impact of uncertainty in precipitation forcing data sets on the hydrologic budget of an integrated hydrologic model in mountainous terrain. *Water Resour. Res.*, **56**, e2020WR027639, <https://doi.org/10.1029/2020WR027639>.
- 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. *Environ. Modell. Software*, **132**, 104791, <https://doi.org/10.1016/j.envsoft.2020.104791>.
- , L. K. Cunha, and W. F. Krajewski, 2013: Uncertainty in radar-rainfall composite and its impact on hydrologic prediction for the eastern Iowa flood of 2008. *Water Resour. Res.*, **49**, 2747–2764, <https://doi.org/10.1002/wrcr.20244>.
- , 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>.
- , and Coauthors, 2018: Comprehensive evaluation of the IFloodS radar-rainfall products for hydrologic applications. *J. Hydrometeor.*, **19**, 1793–1813, <https://doi.org/10.1175/JHM-D-18-0080.1>.
- , M. Keem, R. Hammond, I. Demir, and W. F. Krajewski, 2019: A pilot infrastructure for searching rainfall metadata and generating rainfall product using the big data of NEXRAD. *Environ. Modell. Software*, **117**, 69–75, <https://doi.org/10.1016/j.envsoft.2019.03.008>.
- , W. F. Krajewski, and Y. Qi, 2020a: Utility of vertically integrated liquid water content for radar-rainfall estimation: Quality control and precipitation type classification. *Atmos. Res.*, **236**, 104800, <https://doi.org/10.1016/j.atmosres.2019.104800>.
- , —, and A. Ryzhkov, 2020c: Evaluation of the specific attenuation method for radar-based quantitative precipitation estimation: Improvements and practical challenges. *J. Hydrometeor.*, **21**, 1333–1347, <https://doi.org/10.1175/JHM-D-20-0030.1>.
- Small, S. J., L. O. Jay, R. Mantilla, R. Curtu, L. K. Cunha, M. Fonley, and W. F. Krajewski, 2013: An asynchronous solver for systems of ODEs linked by a directed tree structure. *Adv. Water Resour.*, **53**, 23–32, <https://doi.org/10.1016/j.advwatres.2012.10.011>.
- Smith, M. B., and Coauthors, 2012: Results of the DMIP 2 Oklahoma experiments. *J. Hydrol.*, **418–419**, 17–48, <https://doi.org/10.1016/j.jhydrol.2011.08.056>.
- 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. 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>.
- Vivoni, E. R., D. Entekhabi, and R. N. Hoffman, 2007: Error propagation of radar rainfall nowcasting fields through a fully distributed flood forecasting model. *J. Appl. Meteor. Climatol.*, **46**, 932–940, <https://doi.org/10.1175/JAM2506.1>.
- Wang, Y., S. Cocks, L. Tang, A. Ryzhkov, P. Zhang, J. Zhang, and K. Howard, 2019: A prototype quantitative precipitation estimation algorithm for operational S-band polarimetric radar utilizing specific attenuation and specific differential phase. Part I: Algorithm description. *J. Hydrometeor.*, **20**, 985–997, <https://doi.org/10.1175/JHM-D-18-0071.1>.
- Wigmsta, 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>.
- Zhu, Q., X. Gao, Y.-P. Xu, and Y. Tian, 2019: Merging multi-source precipitation products or merging their simulated hydrologic flows to improve streamflow simulation. *Hydrol. Sci. J.*, **64**, 910–920, <https://doi.org/10.1080/02626667.2019.1612522>.