Hydrology and
Earth System
Sciences

Ensemble streamflow data assimilation using WRF-Hydro and DART: novel localization and inflation techniques applied to Hurricane Florence flooding

Mohamad El Gharamti[1], **James L. McCreight**[2], **Seong Jin Noh**[3], **Timothy J. Hoar**[1], **Arezoo RafieeiNasab**[2], **and Benjamin K. Johnson**[1]

[1]NCAR, Computational and Information Systems Laboratory (CISL), Boulder CO, USA
[2]NCAR, Research Application Laboratory (RAL), Boulder CO, USA
[3]Civil Engineering, Kumoh National Institute of Technology, Gumi, South Korea

Correspondence: Mohamad El Gharamti (gharamti@ucar.edu)

Abstract. Predicting major floods during extreme rainfall events remains an important challenge. Rapid changes in flows over short timescales, combined with multiple sources of model error, makes it difficult to accurately simulate intense floods. This study presents a general data assimilation framework that aims to improve flood predictions in channel routing models. Hurricane Florence, which caused catastrophic flooding and damages in the Carolinas in September 2018, is used as a case study. The National Water Model (NWM) configuration of the WRF-Hydro modeling framework is interfaced with the Data Assimilation Research Testbed (DART) to produce ensemble streamflow forecasts and analyses. Instantaneous streamflow observations from 107 United States Geological Survey (USGS) gauges are assimilated for a period of 1 month.

The data assimilation (DA) system developed in this paper explores two novel contributions, namely (1) along-the-stream (ATS) covariance localization and (2) spatially and temporally varying adaptive covariance inflation. ATS localization aims to mitigate not only spurious correlations, due to limited ensemble size, but also physically incorrect correlations between unconnected and indirectly connected state variables in the river network. We demonstrate that ATS localization provides improved information propagation during the model update. Adaptive prior inflation is used to tackle errors in the prior, including large model biases which often occur in flooding situations. Analysis errors incurred during the update are addressed using posterior inflation. Results

show that ATS localization is a crucial ingredient of our hydrologic DA system, providing at least 40 % more accurate (root mean square error) streamflow estimates than regular, Euclidean distance-based localization. An assessment of hydrographs indicates that adaptive inflation is extremely useful and perhaps indispensable for improving the forecast skill during flooding events with significant model errors. We argue that adaptive prior inflation is able to serve as a vigorous bias correction scheme which varies both spatially and temporally. Major improvements over the model's severely underestimated streamflow estimates are suggested along the Pee Dee River in South Carolina, and many other locations in the domain, where inflation is able to avoid filter divergence and, thereby, assimilate significantly more observations.

1 Introduction

Affecting nearly a 100 million people worldwide per year, flooding is the most common natural disaster (Guha-Sapir et al., 2013). Flooding impacts human life, livelihood, and property. Improved streamflow flood forecasts can benefit the public in a variety of ways, from planning to emergency management. The topic of flood forecasting remains an area of active of research and operational development. This study contributes to improving short-term (hourly) streamflow flooding forecasts by minimizing error in their initial conditions through streamflow data assimilation. We focus

on the rainfall-driven streamflow flooding caused by Hurricane Florence in 2018. Within the context of an operational and spatially distributed hydrologic model, we examine the data assimilation (DA) challenges of dominant errors (bias) arising from the precipitation boundary conditions (forcings) and of improving information propagation from the observations into the model ensemble (background).

Streamflow is one of the most commonly observed hydrologic variables. Its earliest measurements date back to the late 19th century (Ashman et al., 2004). In more recent decades, the assimilation of streamflow observations into hydrologic models has followed various DA approaches and covered the gamut of applications (e.g., Wood and Szöllösi-Nagy, 1978; Kitanidis and Bras, 1980; Moradkhani et al., 2005; Weerts and El Serafy, 2006; Clark et al., 2008; Pauwels and De Lannoy, 2009; Seo et al., 2009; DeChant and Moradkhani, 2012; Noh et al., 2013; McMillan et al., 2013; Rafieeinasab et al., 2014; Lee and Seo, 2014; Sun et al., 2015; Ercolani and Castelli, 2017; Abbaszadeh et al., 2018; Ziliani et al., 2019, and references therein). Meanwhile, hydrologic models have evolved from lumped to high resolution and spatially distributed, given the increase in computational resources and the availability of high-resolution terrain and forcing data.

While research studies have shown success of streamflow and even multivariate data assimilation, operational flood forecasting systems do not typically employ data assimilation (Emerton et al., 2016). Liu et al. (2012) detail the hurdles between hydrologic forecasting research and operations. The authors provide a wealth of recommendations on transitioning DA into hydrologic forecasting. A primary reason DA research is not commonly applied in operational settings is that the methods can perform poorly in the presence of large model errors. Probabilistic DA approaches are optimal only when the model is unbiased (Dee, 2005). When large biases or systematic errors are present, the methods commonly result in filter divergence, which is the case when observations are unable to provide state updates to the model background (e.g., DeChant and Moradkhani, 2012). In advocating for the adoption of DA methods in operational hydrologic forecasting, Seo et al. (2009) and Liu et al. (2012) suggest blending DA procedures with other kinds of interventions in order to balance the need for operational robustness against the multiple advantages (skill improvements, reproducibility, etc.) offered by automated DA methods.

Divergence and other filter-related issues are often pronounced in rainfall-driven flooding systems. This is because the large errors arising at the model boundary are often not part of the prognostic state of the system, have no memory, and cannot be constrained during the analysis. To overcome this, multiple bias correction strategies have been pursued in the context of DA for flood forecasting. Joint state–parameter estimation is often applied to help mitigate model errors (e.g., Abbaszadeh et al., 2018). Multiplicative bias correction parameters, to adjust forcing errors, are sometimes estimated alongside the physical state and parameters (e.g.,

Seo et al., 2003). Bias-aware Kalman filters are applied to estimate model and observation bias by implementing a separate update for two moments, i.e., the mean and the bias (e.g., Drécourt et al., 2006; Rasmussen et al., 2016; Ridler et al., 2018). In addition, a conditional bias-penalized Kalman filter was developed for improved estimation and prediction of hydrologic extremes. The filter operates by minimizing a weighted sum of error variances and type II squared errors, which are different from the conventional Kalman filter which is based on least square minimization (Seo et al., 2018; Lee et al., 2019; Jozaghi et al., 2019). In a recent study, Emery et al. (2020a) proposed updating the boundary fluxes based on the differences between observed and prior streamflow. The authors rerun the assimilation forecast step with the updated boundary conditions to produce a second prior without involving the streamflow state in the update. In this study, we explore the use of spatially and temporally varying adaptive covariance inflation (El Gharamti, 2018) as a way to mitigate bias in the context of extreme flood simulations. Inflation helps restore spread in the ensemble, which can yield a better fit to the observations during the analysis. In addition, spatially varying inflation can help enhance the rank of the sample background covariance matrix (El Gharamti, 2018). In soil and groundwater hydrology, using inflation was reported successful by multiple studies (e.g., Bauser et al., 2018; Jamal and Linker, 2020). In surface water hydrology, however, the impact of inflation on streamflow predictions is not fully understood. This is the first study of its kind where spatially and temporally adaptive inflation are applied to streamflow forecasting. This study further explores the use of prior versus posterior inflation and investigates the effect of each scheme on the performance of the flood prediction ensemble framework. We note that the approach of Emery et al. (2020a), a temporally fixed inflation parameter (scalar) as a means of tuning static background error covariances, is significantly different from the temporally and spatially adaptive inflation applied to time evolving background error covariances in this paper.

It is no surprise that many hydrologic and flood forecasting DA studies have highlighted the importance of estimating accurate background error covariances. Model bias, as discussed above, and sampling error hinders proper estimation of error covariances. The nonlinear relationship between variables in hydrologic modeling makes it more challenging to update unobserved state variables. Filtering approaches only consider the instantaneous error covariances. Smoothers, on the other hand, can be applied to remedy this problem (e.g., Pauwels and De Lannoy, 2006; Li et al., 2013). Even when employing an ensemble smoother for flood forecasting, Rakovec et al. (2015) concluded that the elimination of the strongly nonlinear relation between soil moisture and discharge observations improved flood forecasts. Clark et al. (2008) commented that modeled error correlations were much larger than observed error correlations and that inadequacies in modeling the spatial variability in hydrological

processes hindered the transfer of observational information to ungauged basins. In this study, we revisit the spatial basis for information propagation of observations via error covariances. We investigate updating distributed hydrologic states and propose a new topologically based localization strategy for stream networks. The method is called along-the-stream (ATS) covariance localization, and it confines state updating to directly connected (defined below) hydrological states. Information propagation to ungauged basins (e.g., Sivapalan et al., 2003) within our strategy requires such basins to be upstream of observations.

The development of the data assimilation framework in this paper begins from NOAA's National Water Model (NWM; https://water.noaa.gov/about/nwm, last access: 27 September 2019) configuration of the WRF-Hydro hydrological framework (Gochis et al., 2020). The NWM is a spatially distributed hydrologic model that produces operational forecasts and analyses of distributed hydrologic states, including streamflow, over the continental USA (also, more recently, with separate implementations in various other regions). The operational products are not evaluated in this study. However, the real-time NWM forcing fields from Hurricane Florence are run through a model configuration very close to the operational analysis configuration, providing an open-loop (no DA) deterministic analysis very close to what the NWM would have produced in real-time (if such an open-loop run were operational). Using one-way fluxes from this analysis, we drive a channel+bucket submodel of the NWM that includes streamflow and conceptual bucket storage states. The reduced computational cost of this submodel, the perturbation of its parameters, and the time-varying perturbations applied to the deterministic fluxes from the land surface model provide an ensemble basis for our DA experiments. The resulting NWM channel+bucket modeling system is interfaced with the Data Assimilation Research Testbed (DART). DART, developed and maintained at the (United States) National Center for Atmospheric Research, is an open-source community facility that provides software tools for data assimilation research, development, and education (Anderson et al., 2009). Streamflow observations from the United States Geological Survey (USGS) gauges, retrieved from the (United States) National Water Information System (NWIS; https://waterdata.usgs.gov/nwis, last access: 20 August 2019), are used to update the spatially distributed ensemble states of streamflow and groundwater bucket head. The resulting analyses could serve as the initial conditions for short-term flood forecasts. The analyses are evaluated in terms of the assimilation priors (i.e., the 1 h forecast). Hydrographs and other time series assessment tools (including errors, bias, ensemble spread, etc.) are utilized to investigate the performance of the DA framework. Streamflow distribution in space, resulting from DA, is also studied and compared to the model's estimate.

The rest of the paper is organized as follows. Section 2 presents the Hurricane Florence subdomain, the NWM sub-

model and its components, the uncertainties incorporated into the ensemble design, and the USGS observations assimilated. DART is briefly introduced in Sect. 3, and then ATS localization and adaptive inflation are described in Sects. 3.2 and 3.3, respectively. Spatial assessment with particular focus on bias correction is given in Sect. 3.6. A summary of the findings and further discussions are found in Sect. 4.
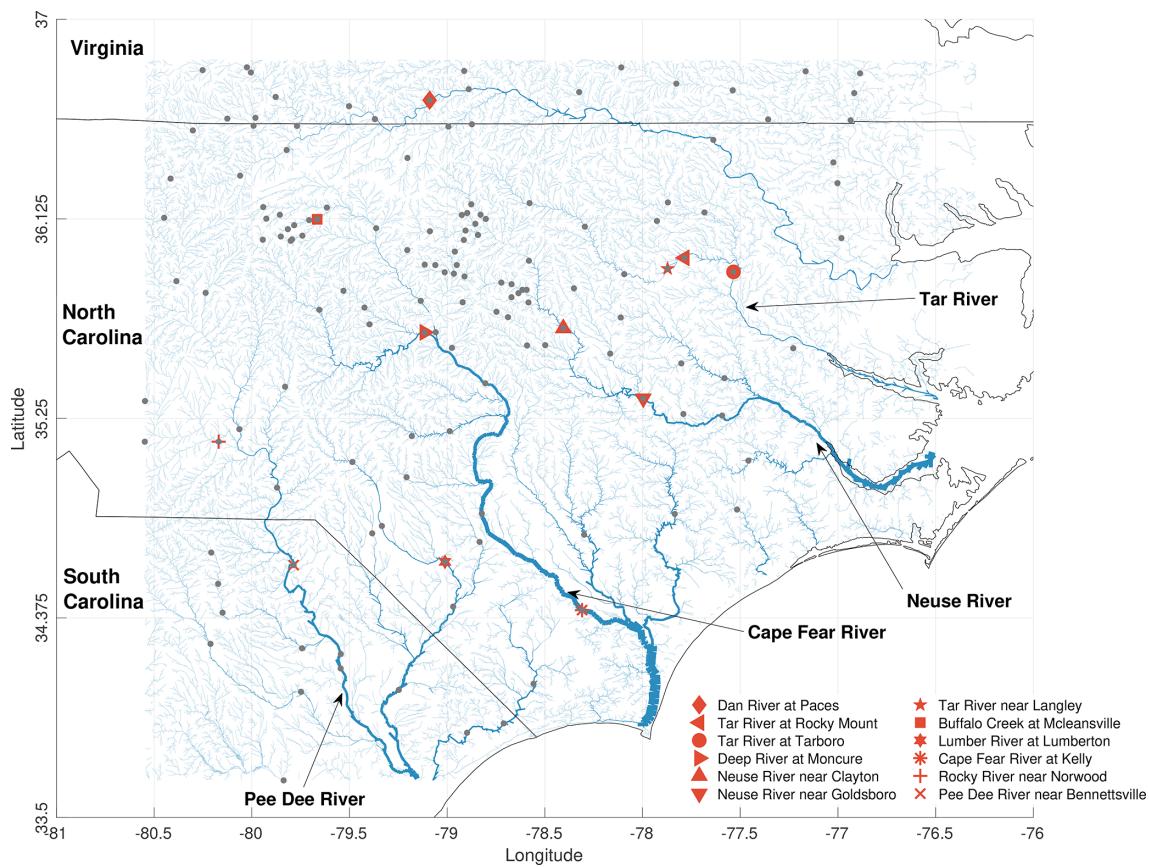
2 Hydrologic model and data

2.1 National Water Model subdomain: Hurricane Florence

In this study, we focus on a regional subdomain of the NWM CONUS (Continental United States) domain affected by Hurricane Florence in September 2018. Figure 1 shows this domain located over the states of North Carolina and South Carolina. Hurricane Florence reached Saffir–Simpson category 4 strength on two separate dates prior to landfall. Though it weakened to category 1 by the time it made landfall, it wreaked over USD 20 billion of damage largely attributed to inland freshwater flooding resulting from extreme rainfall. Many of the largest flood peaks occurred after the dissipation of the hurricane on 19 September as water concentrated along its course to the sea. A timeline of observed and modeled attributes is shown in Table 1.

Figure 1 shows the roughly 67 000 reaches of the NWM stream channel network in this subdomain. The NWM channel network is based largely on USGS's National Hydrography Dataset, namely NHDPlus version 2 (McKay et al., 2012). The rectangular extent of the subdomain indicates the region over which atmospheric forcing data are used to drive the Noah-MP (Niu et al., 2011) land surface model (1 km) and its two-way coupling to lateral surface and subsurface flow routing schemes (250 m; Gochis and Chen, 2003) used by the NWM Analysis and short-range forecast configurations (https://water.noaa.gov/about/nwm, last access: 27 September 2019). As shown in Fig. 2, the lateral flow components are one-way coupled to the streamflow, reservoir, and bucket models of the NWM. Detailed description of this submodel can be found in following section.

All model code (https://github.com/NCAR/wrf_hydro_nwm_public/releases/tag/nwm-v2.0, last access: 27 September 2019), domain data, and parameter sets (https://www.nco.ncep.noaa.gov/pmb/codes/nwprod/, last access: 27 September 2019) used in this study correspond to NWM version 2.0. The single exception is the groundwater bucket model formulation and parameters, which are based on NWM version 2.1. We run the equivalent of the NWM standard analysis and assimilation cycle without the streamflow nudging used by the NWM. This is the configuration also used for the short- and medium-range forecast cycles. During the time of the study, the NWM extended analysis cycle had not yet been implemented.

Figure 1. Hurricane Florence model domain showing the river network with locations of the main Cape Fear and Neuse rivers in North Carolina. The assimilated 107 gauges, from USGS, are denoted by gray dots. The red markers denote 12 gauges that are used for diagnostic and validation purposes. The borders between the states of Virginia, North Carolina, and South Carolina are also shown in black. The thickness of the river reaches denote the strength of streamflow (resulting from an open-loop run and averaging over the month of September in 2018), such that larger thickness means higher streamflow.

Table 1. Hurricane Florence timeline in the Carolinas, USA, 2018. NLDAS-2 denotes the forcing data for phase 2 of the North American Land Data Assimilation System.
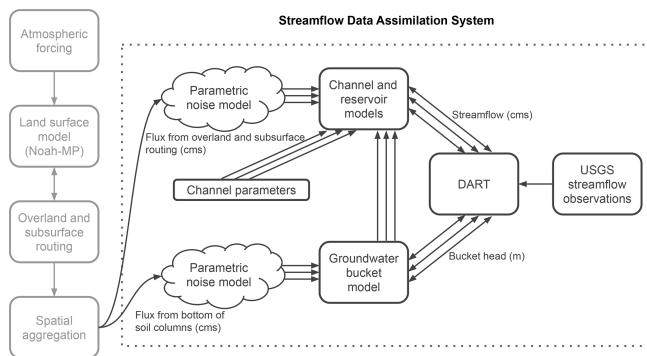
Observed and modeled timeline		
Modeled	NLDAS2-based restart; advance with operational forcings	1 August
Observed	Hurricane forms at sea	31 August
Modeled	Hourly data assimilation starts	1 September
Observed	Hurricane landfall	14 September
Observed	Hurricane dissipates	19 September
Modeled	End of simulation	15 October

The major rivers in this region are labeled in Fig. 1. To the west, the Pee Dee River has its headwaters between Charlotte and Winston–Salem and flows from North Carolina into South Carolina. Its major tributary, the Lumber River, meets Pee Dee just before reaching the edge of the domain. The Cape Fear River, with its headwaters near Greensboro and joining the sea near Wilmington, is seen in the center of the domain. Further to the east, the Neuse River flows by Durham and Raleigh (the capital of North Carolina) and

flows in to the Pamlico Sound. The Tar River (not labeled, except by its cluster of three gauges in the legend) lies to the North of the Neuse and also flows into the Pamlico Sound.

2.2 Channel+bucket submodel

We run the so-called channel+bucket submodel of the NWM. Figure 2 illustrates the one-way runoff fluxes to the streamflow and groundwater bucket models from the up-

Figure 2. Streamflow data assimilation system overview. Vertical boxes on the left show the deterministic NWM model chain from forcing through aggregation of (overland, subsurface, and column drainage) routing output fluxes. These output fluxes are the inputs to the data assimilation system used in this paper, shown inside the dotted box. Random noise is applied to these inputs to generate ensemble forcings. Ensembles are denoted by groups of three arrows (the ensemble size is much larger than three). The ensemble fluxes drive the ensemble model components (channel and reservoir model and groundwater bucket model) used in the assimilation. The depicted time-invariant a priori error distribution of channel parameters provides a multiphysics streamflow ensemble. The groundwater states produce additional fluxes to the channel and reservoir model. The spatially distributed streamflow and bucket head states comprise the state vector passed to DART for updating by USGS streamflow observations. The unit cms denotes cubic meters per second.

stream model components (land surface model, overland and subsurface routing, and spatial aggregations). The fluxes are saved as forcing files for running the channel+bucket submodel used in our assimilation approach. To generate these fluxes, the full NWM is first run once with its own set of atmospheric boundary conditions.

The use of the one-way coupled submodel means that no error covariances with the upstream components of the model will be considered (e.g., soil moisture, surface head, etc.) and that the control vector will consist of two spatially distributed states, i.e., streamflow and groundwater bucket head. Reservoir states, embedded in the stream network calculations, are not considered in the state updating.

2.3 Forcings, spin-up, and simulation

The full model was run (with no data assimilation) using NLDAS-2 (North American Land Data Assimilation System phase 2; Xia et al., 2014) forcings from 1 January 2010 to 1 July 2018. This open-loop run was then continued from 1 July through 15 October 2018 with the NWM operational analysis and assimilation cycle forcings which were collected in real-time from NOMADS (NOAA Operational Model Archive and Distribution System; https://nomads.ncep.noaa.gov/, last access: 27 September 2019). This real-time forcing product is based on magnetic res-

onance mass spectrometry (MRMS; https://www.nssl.noaa.gov/projects/mrms/, last access: 27 September 2019; Zhang et al., 2016) gauge-adjusted and radar-only observed precipitation products along with short-range Rapid Refresh (RAP; https://rapidrefresh.noaa.gov/, last access: 27 September 2019) and high-resolution Rapid Refresh (HRRR; https://rapidrefresh.noaa.gov/hrrr/, last access: 27 September 2019) products (Benjamin et al., 2016; see https://water.noaa.gov/about/nwm, last access: 27 September 2019). For the period 1 August through 15 October 2018, fluxes were saved for forcing the channel+bucket submodel in the data assimilation experiments. Figure 2 shows these fluxes as inputs to the data assimilation system. Initial states for the data assimilation experiments were also taken from the full model run on 1 August 2018.

2.4 Muskingum–Cunge streamflow model

The NWM implements Muskingum–Cunge (M–C) streamflow routing with variable parameters (e.g., Ponce and Yevjevich, 1978) in a compound channel (Garbrecht and Brunner, 1991). M–C with variable parameters is a common approach to streamflow routing over large watersheds and has been successfully applied in many instances. The compound channel (Fig. 3) provides a lower trapezoidal channel and an upper rectangular channel section to simulate overbank flows. M–C is applied to the stream channel network derived from NHDPlus version 2, shown in Fig. 1, with trapezoidal channel geometry and Manning's N, n (roughness) parameter values for each reach.

The one-dimensional storage (S) relationship between inflow (I) and outflow (O) on a reach (spatial segment) is given by the following:

$$S = K\left[XI + (1 - X)O\right], \tag{1}$$

with storage coefficient K and weighting factor X. Formulated as a finite difference over a reach, this yields the following:

$$O_k = C_1 I_{k-1} + C_2 I_k + C_3 O_{k-1} + C_4 L. \tag{2}$$

This is an explicit solution for the current outflow (O_k, with k denoting the time step index) as a function of previous and current inflows (I_{k-1} and I_k, respectively), previous outflow (O_{k-1}), and the lateral inflows (combined overland and subsurface, L) to the reach. The coefficients in Eq. (2) can be expressed as combinations of the Courant and Reynolds numbers (e.g., Ponce and Lugo, 2001), respectively, as follows:

$$C = \frac{c\Delta t}{\Delta x} = \frac{\Delta t}{K}, \tag{3}$$

$$D = \frac{q}{sc\Delta x} = 1 - 2X, \tag{4}$$

where Δt is the time step, Δx is the reach length, s is the reach slope, c is the celerity, and q is the unit discharge (discharge divided by the top width of the flow). Also shown are the relationships to K and X parameters.

The assumptions of the M–C approach do not allow for backwater effects in the solution. However, the M–C variable parameter approach allows nonlinear flood wave dynamics by accounting for the interdependence of the time-varying flow rate and its geometry. Specifically, the celerity and unit discharge as follows:

$$c = \frac{dQ}{dA},\tag{5}$$

$$q = \frac{Q}{b},\tag{6}$$

which is used for calculating the coefficients in Eq. (2), depending on the flow (Q), its area (A), and top width (b), which are mediated by the channel geometry and roughness parameters in each reach. These parameters, some of which are shown in Fig. 3, are described in more detail in Sect. 2.6. The equation for celerity can be solved from Manning's equation for uniform flow. Garbrecht and Brunner (1991) solve the celerity equation for the case of the compound channel shown in Fig. 3. The variable parameter approach is an iterative solution, updating the flow and its geometry in alternate steps, to converge on a physically consistent discharge–geometry solution. The implementation in the NWM follows the secant method, which takes a high and a low departure from an initial water depth and iterates through the equations of geometry and flow until the calculated flow rates converge within some threshold. Before the flow rates converge, their differences are used to reduce the discrepancy in the estimated water depths.

We note that the NWM makes a short time step approximation, $I_k = I_{k-1}$, to eliminate the spatial/topological dependence at the current time and render the solution of the M–C method embarrassingly parallel.

Reservoir objects embedded into the NWM routing network accept fluxes from the streamflow network and from the overland and subsurface routing model on adjacent grid cells. Water is discharged to the stream network via equations for both weir and orifice flow in the NWM level pool scheme. Because we do not include the reservoir level in the assimilation state vector, the reader is referred to Gochis et al. (2020) for further details.

2.5 Groundwater bucket model

Even when lateral routing processes are included in hydrologic modeling, deficiencies in soil and aquifer data and model process representations commonly lead to underestimation of the baseflow component of streamflow. The NWM employs a groundwater bucket model as a simple aquifer representation to mitigate this baseflow problem. This model accepts water fluxes from the bottom of the land model's soil columns. The spatial representation of the buckets is derived from the NHDPlus (McKay et al., 2012) catchments. These map roughly on to the stream reaches (with certain exceptions). The buckets have an average areal extent of ~ 3 km

across the NWM CONUS domain and, therefore, accept fractions of discharge from multiple land model columns. The mapping from the land surface model to the buckets is performed by user-defined mappings capability of the model.

The bucket scheme is simple and highly conceptual. For this reason, calibration of its parameters is critical for reasonable model simulations. The groundwater bucket model and its parameters are expressed by the following set of equations, which are the only model components taken from NWM v2.1 (instead of v2.0). The current bucket head, z_k, is solved from the previous bucket head, z_{k-1}, plus the change in head due to the bucket inflow $\widetilde{I}_k$, as follows:

$$z_k = z_{k-1} + \frac{\widetilde{I}_k \Delta t}{\widetilde{A}},\tag{7}$$

where $\widetilde{A}$ is the bucket area. The finite capacity of the bucket is expressed in terms of a maximum head, z_{max}, a tunable parameter. When the current head exceeds this threshold, the Q_{spill} term becomes nonzero, discharging all excess head in a single time step.

if　$z_k \leq z_{max}$　**then**
　$Q_{spill} = 0$
else
　$z_{spill} = z_k - z_{max}$　　　　　　　　　　　　　(8)
　　$z_k = z_{max}$　　　　　　　　　　　　　　　(9)
　$Q_{spill} = \dfrac{\widetilde{A} z_{spill}}{\Delta t}.$　　　　　　　　　　　　(10)

Head up to and including the z_{max} results in bucket discharge, following an exponential equation, containing the following two additional tunable parameters, E (unitless) and G (cubic meters per second; hereafter cms):

$$Q_{exp} = G \left[\exp\left(E \frac{z_k}{z_{max}} \right) - 1 \right].\tag{8}$$

The spill discharge and the exponential bucket discharge are finally combined to give the total bucket outflow at the current time step ($\widetilde{O}_k$), and the depth of water corresponding to Q_{exp} is removed from the bucket.
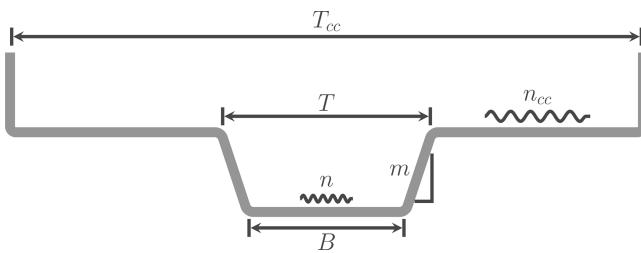
$$\widetilde{O}_k = Q_{spill} + Q_{exp},\tag{9}$$

$$z_{k+1} = z_k - \frac{Q_{exp} \Delta t}{\widetilde{A}}.\tag{10}$$

Calibration of the bucket parameters (in advance of NWM version 2.1 calibration) yielded the following spatially uniform bucket parameters used in this study: $G = 0.005$, $E = 7.1244$, and $z_{max} = 15.6476$ mm.

2.6 Sources of uncertainty: ensemble design

We construct an ensemble of 80 members. This number was selected to balance computational demands and statistical

Figure 3. Schematic of the geometry and roughness parameters of the streamflow compound channel, with the top width (T), bottom width (B), side slope (z), Manning's N (n), width of the compound channel (T_{cc}), and Manning's N of the compound channel (n_{cc}).

performance. A more detailed justification on the choice of the optimal ensemble size can be found in Appendix A. Incorporating different sources of uncertainty into the ensemble is necessary to create variability and to obtain a good estimate of the background error covariance. Background error covariances are considered among and between the spatially distributed streamflow and bucket states (O_k and z_k, respectively). We produce error distributions in these states through a priori error distributions on (1) stream channel parameters, (2) forcing fluxes to the channel reaches, and (3) forcing fluxes to the buckets.

The error distribution imposed on the streamflow channel parameters is time invariant and unaffected by the state update. This kind of error source is termed "multiphysics" (Berner et al., 2011), meaning that each member runs a different physical configuration of the model. The parameters shown in Fig. 3 describe the compound channel geometry used in the NWM v2.0, including the top width (T), bottom width (B), side slope (m), Manning's N (n), width of the compound channel (T_{cc}), and Manning's N of the compound channel (n_{cc}). While the lower part of the compound channel is trapezoidal, the upper part of the channel is assumed to be rectangular and, therefore, has no side slope parameter. These parameters vary in space, and we define a scalar multiplier for each parameter and ensemble member to generate a perturbed parameter vector from the existing NWM parameter vector. The multipliers are sampled from uniform distributions, and in the case of three parameters, we redraw the multiplier until the following physical constraints are satisfied: $n_{cc} > 1.5n$, $T > 1.2B$, and $T_{cc} > 2T$. For the geometric quantities, we draw multipliers from $\mathcal{U}[0.6, 1.4]$. For the Manning's N parameters, we draw multipliers from $\mathcal{U}[0.8, 1.8]$ based on the prior belief that the original values are somewhat too low.

Perturbations of the boundary fluxes to the streamflow and bucket models are applied at the hourly forcing time step. These perturbations are uncorrelated (in space, time, and member) Gaussian samples with zero mean and standard deviation equal to 40 % of the flux value at each location. When the perturbations are added to the fluxes, a minimum of zero flux is ensured. Random noise generators are seeded as a

function of "datetime" (in Python) and ensemble member to ensure that identical forcing distributions are used across all experiments. Finally, perturbations are applied to the model initial states on 1 September 2018. However, these ensemble initial conditions account for very little of the uncertainty in the overall experiment.

2.7 USGS streamflow observations

Streamflow observations served by the USGS's (National Water Information System, NWIS; https://waterdata.usgs.gov/nwis?, last access: 20 August 2019) are used by the NWM in near real time. The observation files ingested by the NWM are provided along with its output in near real time on NOMADS (https://nomads.ncep.noaa.gov/, last access: 20 August 2019). The streamflow observations in these files are always provisional because they are near real time, and they are subject to revision until they have been thoroughly assessed. For this study, we collected NWM observation files and revised values from the USGS's NWIS many months after the time period of this study. As expected, there were significant revisions to the streamflow values in the months following Hurricane Florence. These revisions are for multiple reasons, not least of which is that existing rating curves do not typically extrapolate well to extreme and out-of-bank flows. We note that the difference between these observation sets had a significant impact on our results, and that the provisional data proved more challenging for the assimilation methodology in this paper. It is extremely important to study the differences between such provisional and approved data in order to bridge the gap between the methods offered in this paper and real-time data assimilation applications. Ultimately, one would want to assimilate provisional data and evaluate against revised data. There are multiple issues to consider in this regard, including observation gaps, uncertainty, and quality measures. In our study, we chose to use the revised observations to evaluate the performance of our methodological innovations. This study could be extended to simulate real-time streamflow assimilation.

Figure 1 shows the 107 USGS streamflow gauges used for evaluation in this study in green and red. The names of the gauges in red are given in the caption as these locations are specifically called out in the results. All stream gauges considered in this study have their contributing area entirely contained within this subdomain. Most gauges have a 15 min reporting frequency, and the available observations in the previous hour are used for hourly assimilation. All experiments presented here use the heteroscedastic error model of 20 % of the observed flow for the observation error (cms). This is certainly a simplistic approach, but the magnitude is roughly in line with previous studies (e.g., Coxon et al., 2015). We note that, while important, the observation error plays a somewhat secondary role in the quality of the assimilation, particularly given the application of inflation.

3 Data assimilation framework and results

3.1 DART

This study uses the Data Assimilation Research Testbed (DART; Anderson et al., 2009) to perform ensemble Kalman filtering for streamflow forecasting. Utilizing Bayes' rule, the goal is to sequentially use streamflow gauge data to guide the trajectory of the hydrologic model towards a better flood prediction. The procedure consists of successive forecast and analysis steps. During the forecast, a set of model realizations of the state variables are integrated forward in time using the nonlinear hydrologic model as follows:

$$x_k^{f(i)} = \mathcal{M}\left(x_{k-1}^{a(i)}, \boldsymbol{\theta}^{(i)}, \boldsymbol{\gamma}_k^{(i)}\right), \quad i = 1, 2, \ldots, N_e, \tag{11}$$

where $x_k = [O_k, z_k]^T$ is the DART state consisting of the streamflow and the bucket. The superscript i is the number of the ensemble member, N_e is the ensemble size, f denotes forecast (prior), and a is the analysis (posterior). The function $\mathcal{M}$ refers to the WRF-Hydro submodel. $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ denote a set of physical parameters and input model forcings (described in Sect. 2.6), respectively. As the data become available, DART assimilates the observations serially and applies an EAKF update (ensemble adjustment Kalman filter; Anderson, 2003) as follows:

$$\Delta x_j^{(i)} = \sigma_{xy}\sigma_y^{-2}\Delta y^{(i)}, \quad j = 1, 2, \ldots, N_x \tag{12}$$

$$x_{j,k}^{a(i)} = x_{j,k}^{f(i)} + \alpha\Delta x_j^{(i)}. \tag{13}$$

Given an observation, the prior ensemble members of the observed variable y are first updated. The observation space increments, $\Delta y^{(i)} = y^{a(i)} - y^{f(i)}$, are computed using a scalar ensemble filter (Anderson, 2003). These increments are then used to obtain the state–space increments, $\Delta x_j^{(i)}$, as shown in Eq. (12). σ_{xy} denotes the prior covariance of the observed variable, y, and the jth element in the state vector, x. The total number of elements in the state is denoted by N_x. The sample variance of the observed variable is σ_y^2. A localization coefficient, $0 \leq \alpha \leq 1$, is used to limit the impact of spurious correlations in the update. α is computed as a function of the distance between observation and state variables given a predefined correlation structure (refer to Sect. 3.2).

Streamflow gauges are available at the location of the state variables and assumed representative of the model element to which they are associated. This makes the (forward) observation operator linear and equal to the identity matrix, significantly simplifying the implementation of the update step in DART. Variance underestimation is tackled through covariance inflation such that the ensemble right after the forecast or analysis steps is inflated around its mean, as follows:

$$x_j^{f|a(i)} = \sqrt{\lambda}\left(x_j^{f|a(i)} - \overline{x}_j^{f|a}\right) + \overline{x}_j^{f|a}, \tag{14}$$

where $\overline{x}_j$ is the jth element of the ensemble mean and the notation $f|a$ is used to refer to either forecast or analysis
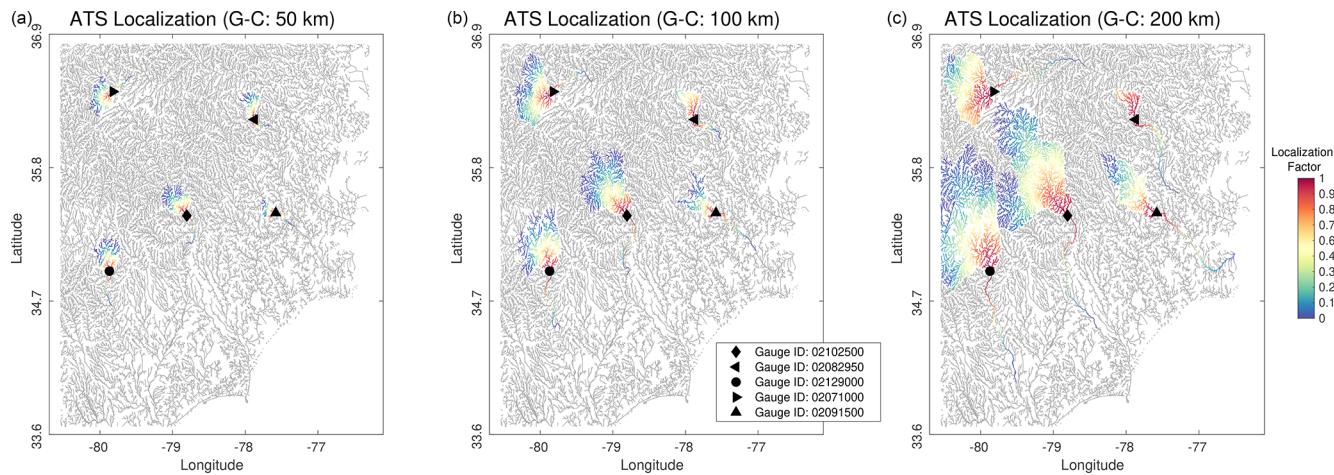
ensemble. The inflation factor $\sqrt{\lambda}$ (typically larger than 1) yields a sample covariance matrix scaled by λ. The effects of prior versus posterior inflation on the filtering performance is explored in details in Sect. 3.3.

3.2 Along-the-stream localization

It is well-recognized that the use of small ensemble sizes produces imperfect sample covariance matrices (e.g., Houtekamer and Mitchell, 2001). In fact, with a small ensemble the probability density function of the state remains only partially explored, which can possibly yield loss of information and even filter divergence. In addition, the sample covariance would generally be contaminated with spurious unrealistic correlations that may degrade the quality of the Kalman update.

To overcome these issues, we resort to using covariance localization. The idea is to taper any spurious correlations between variables that are physically far from each other and are possibly uncorrelated, using α in Eq. (13). Studies have shown that, given the Euclidean distance between different variables, a correlation function could be utilized to compute a localization factor, α. In the present study, a simple Euclidean distance could be inappropriate in many circumstances. For example, reaches from two different watersheds could be physically close but highly unrelated, particularly in terms of their error correlations. To this end, a topologically based localization strategy that adheres to the river network structure is applied. We introduce the along-the-stream localization (ATS localization) strategy. The idea is that only the reaches upstream and downstream from a particular observation are considered during the update (Fig. 4). The localization factor, α, is computed using a selected functional form (e.g., Gaspari–Cohn, boxcar, or ramped boxcar; see Table A1 and the inset in Fig. 6) which depends on the distance between any two reaches and the tunable localization radius, r.

ATS localization highlights some key features. (i) Upstream from each observation, information flows up the network, including through the bifurcations. Downstream from each observation, we assume that the flow of information only travels downstream with the observed flow. As such, we obtain tree-like shapes where the number of close reaches upstream (tree canopy) of the observation is significantly larger than the number of close reaches in the downstream direction (tree trunk). Not allowing information to round the bend or bifurcate back upstream below the gauge, we choose to only update flows which contribute to the observation (upstream) and to which the observation contributes (downstream). This choice was made to be distinct from Euclidean distance-based localization and out of caution, given a modestly sized ensemble, since observations near the confluence of major tributaries might have undue influence on large flows with potentially low (true) error correlations. Allowing upstream bifurcations below the gauge could be a reasonable approach

Figure 4. Illustration of the along-the-stream (ATS) localization strategy in the model domain using the following three different effective localization radii: 50 km **(a)**, 100 km **(b)**, and 200 km **(c)**. The resulting localization factor α is displayed for five different stream gauges. The correlation function used to compute α is based on the Gaspari–Cohn (G–C) fifth-order compactly supported Gaussian-like curve.

as well, pending the choice of ensemble size and understanding of correlated errors at major tributaries. (ii) The total number of close reaches does not necessarily increase as r increases. For instance, as can be seen in Fig. 4, the number of close reaches to gauge ID 0210500 (♦ marker) using $r = 50$, 100, and 200 km is 185, 646, and 1126 reaches, respectively. The same is not true for gauge ID 02082950 (◄ marker) because of the limited number of upstream reaches within the catchment. (iii) Observations in different catchments do not have common close reaches. Gauge IDs 02102500 and 02129000 for $r = 200$ km clearly demonstrate this feature.

The proposed localization method shares a lot of similarities with that of Emery et al. (E20; 2020b). The fundamental difference is that we are tackling sampling errors in the forecast error covariances with each assimilation cycle. Our sample covariances are computed using the evolving ensemble, unlike E20 in which the authors use time-invariant covariances. Given that the hydrologic model used in E20 is linear, their system is technically an optimal interpolation with fixed error statistics. Localization in this context is used to address structural errors of the covariances and to compensate for time-invariant covariances. Another important difference is that our ATS localization approach can ensure that the impact of the observation decreases as the distance from the observation, both upstream and downstream, increases. For example, the fifth-order polynomial function of Gaspari and Cohn (1999) can be used to find the localization coefficients or other functional forms can be used. In E20, on the other hand, all reaches that are close to the observation are assigned the same weight (i.e., $\alpha = 1$).
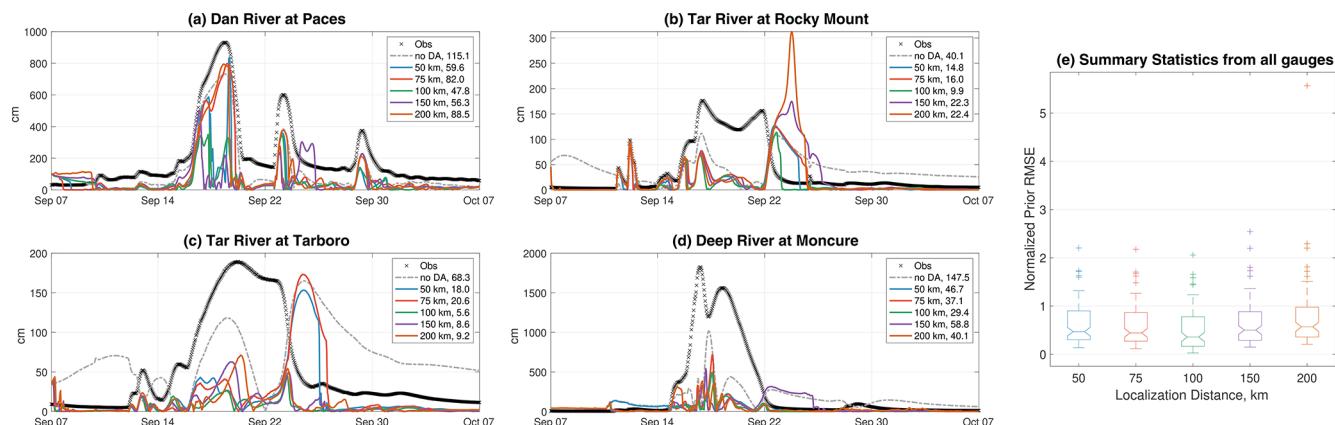
3.2.1 Tuning localization parameters

We conduct five DA experiments to study the sensitivity of the chosen localization radius on the accuracy of the stream-
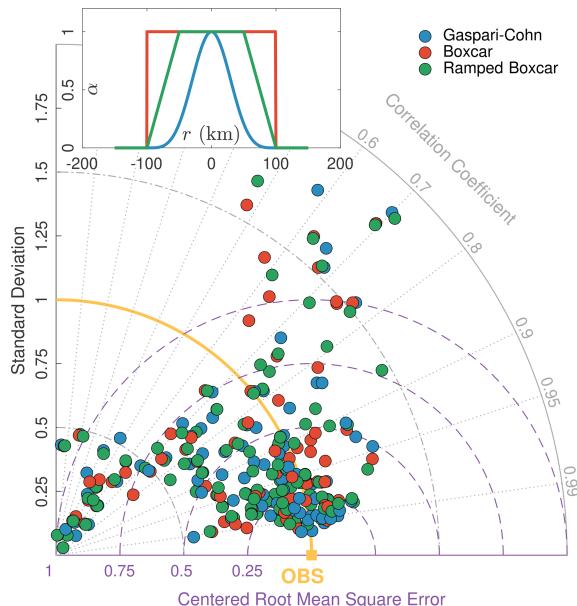
flow estimates. The tested localization radii are 50, 75, 100, 150, and 200 km. The performance of each experiment is assessed at four different locations inside the Hurricane Florence domain (refer Fig. 1). Time series of hourly forecast root mean squared errors (RMSEs) are displayed at each gauge in Fig. 5. As can be seen, all DA runs clearly outperform the open loop (i.e., no DA), especially during the main event at around 17 September. Concerning the localization radius, it is shown that DA runs using $r = 50$ and 75 km produce the least accurate streamflow estimates. At Tar River (Tarboro), for example, the RMSE from these two experiments is almost similar (~ 160 cms) to that of the open loop on 25 September. This suggests that the Kalman update with such a small localization radius may be inadequate. Larger localization radii, $r = 150$ and 200 km, produce on average slightly better estimates. Such a performance, however, is inconsistent in space, as can be seen at Tar River (Rocky Mount).

Overall, the best performance is obtained using $r = 100$ km. This was confirmed not only at the diagnosed gauges but also at the rest of the gauges in the domain, as can be seen from the box plots in Fig. 5e. Our analysis indicates that larger radii generally give rise to spurious correlations which could yield catastrophic streamflow estimates, as the box plot outliers suggest for $r = 200$ km. On the other hand, smaller localization distances limit the amount of useful information, yet they do not severely degrade the quality of the streamflow estimates as in the case of $r > 100$ km.

The effect of the choice of the correlation function used in the ATS localization scheme is also investigated. We compare the following three different functions: Gaspari–Cohn, simple boxcar (similar to E20), and a ramped boxcar. The formulas used to compute α for the different correlation functions are given in Table A1 of Appendix A. The resulting

Figure 5. (a–d) Time series of hourly forecast root mean square errors (RMSEs) of five DA experiments using ATS localization with the Gaspari–Cohn function and radius $r = 50, 75, 00, 150$, and 200 km. For each gauge, the observations are plotted by black crosses. The other curves show the RMSE of each run. For example, the open-loop (ensemble run with no DA) RMSE is given by the dashed gray curve. The time-averaged RMSE for each run is reported in the legend. **(e)** Box plots of the normalized prior RMSEs (time-averaged RMSE divided by the temporal mean of the actual observations at all the gauges), using all gauges as a function of the localization radius.

Figure 6. Taylor diagram for hourly prior streamflow estimates using ATS localization with three different correlation functions, namely Gaspari–Cohn, boxcar, and ramped boxcar. The localization radius is set to 100 km. The shape of the functions is compared at the top of the plot. Ramped boxcar decays linearly to 0, starting at half-width (i.e., 50 km) distance from the observation. Comparisons to all gauges in the domain are performed; however, estimates with high errors and standard deviations, resulting from boxcar and ramped boxcar, are not shown for clarity.

prior ensemble mean from each scenario is evaluated at all 107 streamflow gauges in the domain, and the results for $r = 100$ km are summarized in the Taylor plot (Taylor, 2001) of Fig. 6. The diagram is useful to quantify the degree of cor-

respondence between the gauge observations and the prior streamflow estimates in terms of three statistics, namely the Pearson correlation coefficient, centered root mean squared error, and the standard deviation. We note that boxcar and ramped boxcar estimates produced erroneous results at a few (around seven) gauges, and these results had to be removed from the diagram for visual purposes. Averaging over all gauges, the correlation coefficient resulting from Gaspari–Cohn, boxcar, and ramped boxcar was found to be 0.83, 0.77, and 0.79, respectively. The Gaspari–Cohn function further yielded the lowest values, on average, for the other two statistics of the Taylor diagram.

Boxcar and ramped boxcar functions only outperform Gaspari–Cohn for small localization radii (e.g., $r = 50$ km); however, the results produced with Gaspari–Cohn and $r = 100$ km have the best accuracy. This configuration, as a result, is selected and used in all other results shown in this study.

3.2.2 ATS comparison to Euclidean distance-based localization

This section compares the proposed topologically based ATS localization to the regular Euclidean distance-based localization. Instead of searching for close streams on the river network as in the previous section, the regular approach looks for close-by reaches with a circle, given a prespecified localization radius. In total, five different localization radii, namely $r = 1, 2, 5, 10$, and 20 km, are tested. The resulting streamflow estimates are summarized and compared to the ATS (100 km) run for two gauges in Table 2. The two gauges, shown in Fig. 1, are selected such that the performance is assessed at relatively low (i.e, Tar River at Tarboro) and high (i.e., Deep River at Moncure) streamflow regimes.

Table 2. Comparison of ATS and regular (Reg) localization at Tar River and Deep River. The localization radius used in the ATS approach is 100 km. For the regular localization approach, five different radii are tested, namely 20, 10, 5, 2, and 1 km. The metrics used to compare the schemes are prior and posterior RMSE, prior and posterior bias, and prior and posterior spread. The metrics (in cubic meters per second) are all averaged over the entire simulation period.

		ATS (100 km)	Reg (20 km)	Reg (10 km)	Reg (5 km)	Reg (2 km)	Reg (1 km)
Tar River at Tarboro (NWIS 02083500)	Prior RMSE	5.579	18.541	8.860	33.459	41.607	34.323
	Posterior RMSE	4.930	17.819	6.748	25.106	33.664	26.411
	Prior bias	−1.130	−11.648	−1.706	−20.242	−18.091	−11.068
	Posterior bias	−0.848	−11.410	−0.740	−20.373	−17.163	−10.005
	Prior spread	1.919	3.291	2.803	10.895	10.839	9.535
	Posterior spread	1.551	3.004	2.271	6.283	6.425	5.170
Deep River at Moncure (NWIS 02102000)	Prior RMSE	29.440	48.328	69.206	102.992	130.085	146.354
	Posterior RMSE	22.749	36.163	38.009	38.169	38.159	63.991
	Prior bias	−6.498	−14.880	−21.001	−38.676	−58.086	−75.678
	Posterior bias	−3.328	−11.287	−6.914	−4.558	−4.075	−30.536
	Prior spread	15.598	22.425	36.784	60.877	77.549	70.834
	Posterior spread	11.867	16.812	19.645	21.434	22.038	25.595

Among the five experiments that use regular localization, the best performance is suggested using $r = 10$ km. As r decreases below 10, the quality of the prior and posterior streamflow estimates diminishes. For instance, the time-averaged prior RMSE for $r = 10$ and $r = 1$ km at Tar River is 8.86 and 34.323 cms, respectively. It is also noticeable that smaller localization radii yield large prior and posterior ensemble spread. This happens because the tiny localization radius tends to limit the impact of the data during the update, and hence, the shrinkage of the uncertainty around the ensemble mean becomes restricted. For $r = 20$ km, the performance strongly degrades at Tar River. For example, the posterior bias is shown to grow from −0.74 cms, using $r = 10$ km to −11.41 cms for $r = 20$ km. The reason for such a behavior is that, with a radius of 20 km, streamflow is falsely updated with information from nearby basins. Although these basins are physically close, they are, however, governed by different flow regimes. The same is not true at Deep River, and that is because the basin which Deep River belongs to is much larger, and thus, a localization radius of 20 km cannot contaminate the streamflow as we described at Tar River. Increasing the localization radius beyond 20 km yielded catastrophic updates for the streamflow and the subsurface bucket state. In fact, all DA experiments run with regular localization and $r > 20$ km failed at different stages (typically 2 weeks into the run).

Prior and posterior streamflow results obtained using ATS localization are significantly better than those with the regular localization. Unlike regular localization, using the proposed ATS approach, we are able to increase the effective search radius because the algorithm adheres to the physical aspects of the streamflow problem. Compared to the 10 km regular localization run, ATS produces at least 40 % more accurate (in terms of RMSE) streamflow estimates. This is consistent for all 107 gauge locations. Because the algorithm allows the use of large localization radii, ATS scheme further yields more certain estimates (smaller spread) than those that use regular localization.

3.3 Inflation

Variance underestimation in ensemble Kalman filters is a common issue that usually happens in the presence of large sampling errors and model biases (Furrer and Bengtsson, 2007). Sampling errors are the result of using a limited ensemble size. Model biases are deficiencies in the model, causing predictions to be far from the observations. Other sources of errors that might degrade the performance of the filter include non-Gaussianity (Anderson, 2010), systematic errors in the observational operator, and representativity errors (Hodyss and Nichols, 2015). In practice, studies have shown that when model biases exist they tend to dominate other errors in the system (e.g., El Gharamti, 2018), and thus, treating model errors is often prioritized.

In this section, we consider the following three approaches to dealing with the issue of variance underestimation: prior inflation (PR-inf), posterior inflation (PO-inf), and combined prior and posterior inflation (PP-inf). In PR-inf, the prior ensemble is inflated, while in PO-inf the posterior ensemble is inflated. In PP-inf, the update the prior ensemble is inflated before, and then the posterior ensemble is inflated after the update. In their recent study, El Gharamti et al. (2019) compared the three approaches in an atmospheric application. The authors argued that PR-inf is effective at mitigating model errors, while PO-inf can only tackle sampling errors and other issues associated with the analysis such as non-Gaussianity. Combining both inflation schemes was shown

to produce the best results in application to atmospheric general circulation models.

The algorithm used to compute the inflation is adaptive in time, based on Bayes' theorem as in Eq. (15), and results in spatially varying inflation fields.

$$p\left(\lambda|d^{f|a}\right) \propto p\left(d^{f|a}|\lambda\right) \cdot p\left(\lambda\right). \tag{15}$$

The algorithm assumes the inflation to be a random variable with an inverse-Gamma prior distribution $p\left(\lambda\right)$. A Gaussian likelihood function $p\left(d^{f|a}|\lambda\right)$ is constructed using forecast or analysis innovations, $d^{f|a}$, the observation error variance, $\sigma_{\rm o}^2$, and the variance of the ensemble, $\sigma_{y^{f|a}}^2$, as follows:

$$\begin{aligned} p\left(d^f|\lambda\right) &\approx \mathcal{N}\left[\mathbb{E}\left(d^f\right), {\rm var}\left(d^f\right)\right], \\ &= \frac{1}{\sqrt{2\pi}}\exp\left[-\left(\frac{1}{\sqrt{2}}d^f\right)^2\left(\sigma_{\rm o}^2 + \lambda\sigma_{y^f}^2\right)^{-1}\right] \\ &\quad\left(\sigma_{\rm o}^2 + \lambda\sigma_{y^f}^2\right)^{-\frac{1}{2}}. \end{aligned} \tag{16}$$

Equation (16) assumes the forecast and the observation errors are uncorrelated, i.e., $\mathbb{E}\left(d^f\right) = 0$, and $\mathbb{E}$ here denotes the expected value. This is generally valid for most Earth system models. In certain modeling scenarios, where forecast and observation errors have nonzero correlations, one could decorrelate them before proceeding (e.g., Hoteit et al., 2015). The variance of the forecast innovations, $\sigma_{\rm o}^2 + \lambda\sigma_{y^f}^2$, is estimated using observation–space diagnostics (Desroziers et al., 2005). In the case of posterior inflation, the likelihood function is similar to Eq. (16), except that the variance of the analysis innovations, ${\rm var}(d^a)$, is given as $\sigma_{\rm o}^2 - \lambda\sigma_{y^a}^2$. The posterior distribution of the inflation is obtained by taking the product of the likelihood and prior densities of λ, as shown in Eq. (15). To find the updated value of the inflation, $p\left(\lambda|d^{f|a}\right)$ is maximized, and the resulting value is used, in addition to the updated inflation variance, as the mode of the prior density for the next DA cycle. More details can be found in El Gharamti et al. (2019).

The hydrographs in Figs. 7 and 8 compare the performance of PR-inf, PO-inf, and PP-inf with a no inflation (NO-inf) case at two gauges along the Neuse River. At the upstream gauge (near Clayton; Fig. 7), the open loop shows a phase misalignment with the observations, where the model floods almost a week after the main event on ∼ 15 September. The hydrograph resulting from the NO-inf run is hugely biased, as can be seen on 22 September. Because of the large discrepancies between the model estimates and the observation, the filter rejected almost 60 % of the data[1]. PO-inf estimates

---

[1]Observation rejection (also known as the outlier threshold) in DART is applied when the distance between the ensemble mean and the observation is larger than 3 times the total spread. The total spread is computed as the square root of the sum of the prior variance and the observation error variance.

are slightly better than those of the NO-inf run; however, almost half of the observations are still rejected. Using prior inflation (PR-inf), the majority of the observations are assimilated, producing high-quality streamflow estimates. As can be seen, the large biases between 14 and 22 September are completely removed. Whenever the model prediction starts to deviate from the observations' trajectory, the adaptive inflation algorithm reacts immediately by restoring enough spread to bring the ensemble closer to the data during the update. Once the model predictions become consistent with the observations, the inflation relaxes to smaller values. A value of 1 means no inflation is applied. The best fit to the observations is demonstrated by the PP-inf run. Its overall prior and posterior averaged RMSE values are slightly better than those obtained using the PR-inf run.
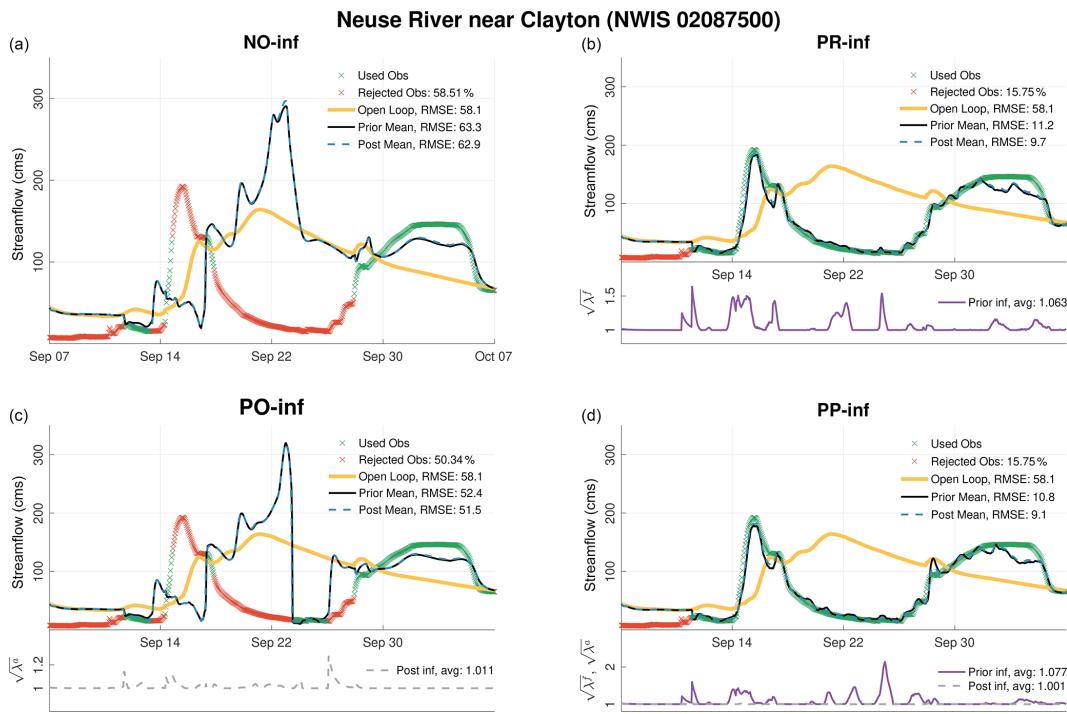
At the downstream gauge (near Goldsboro, as shown in Fig. 8), the discharge is almost 4 times larger than the upstream gauge, and the overall model fit to the data looks better. Towards the end of the flooding event (around 30 September), PO-inf better delineates the data compared to the NO-inf case. However, a false modeled flood wave appears during this time in both simulations. Similar to Fig. 7, PR-inf run clearly outperforms the NO-inf and PO-inf runs, yielding an average prior and posterior RMSE of 45 and 38.1 cms, respectively. On average, the PP-inf prior and posterior estimates are ∼ 5 % more accurate than those of the PR-inf. Both PR-inf and PP-inf runs assimilate almost 80 % of the available observations.

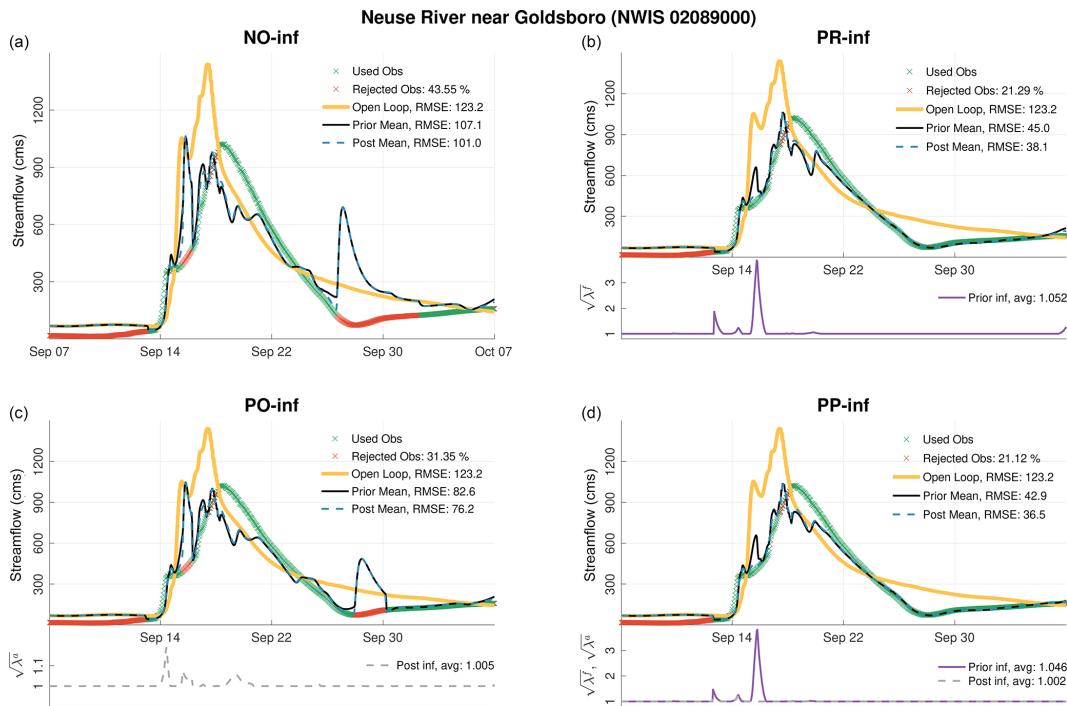## 3.4 Choosing the best inflation

The results shown in Figs. 7 and 8 clearly demonstrate the usefulness of prior inflation at mitigating model biases. The benefits of using posterior inflation are only minimal. To illustrate how important prior inflation is, one could check out the rising limb of the hydrograph at Neuse River near Goldsboro on 15 September. With no inflation, the filter estimates are shown to overestimate the observed discharge and follow the trajectory of the open loop. Although $\sqrt{\lambda^a}$ is shown to increase to almost 1.2 in the PO-inf run, it is insufficient to bring the streamflow closer to the data. Assessing the PR-inf run, one could see that as the prior innovations begin to increase the adaptive scheme counteracts this by increasing $\sqrt{\lambda^f}$ to almost 4. As a result, the posterior mean is kept close to the data, and consequently, the prior estimates improved in the proceeding DA cycles.

Consistent with the findings of El Gharamti et al. (2019), adding posterior inflation on top of prior inflation further increased the accuracy. This suggests that posterior inflation might be resolving other regression issues such as sampling noise and non-Gaussianity. In fact, the gain from using posterior inflation on top of prior inflation is more pronounced at other gauges, as shown in Fig. 9. As can be seen, PR-inf completely misses the falling limb of the hydrograph starting from 22 to 25 September. Prior and posterior means are
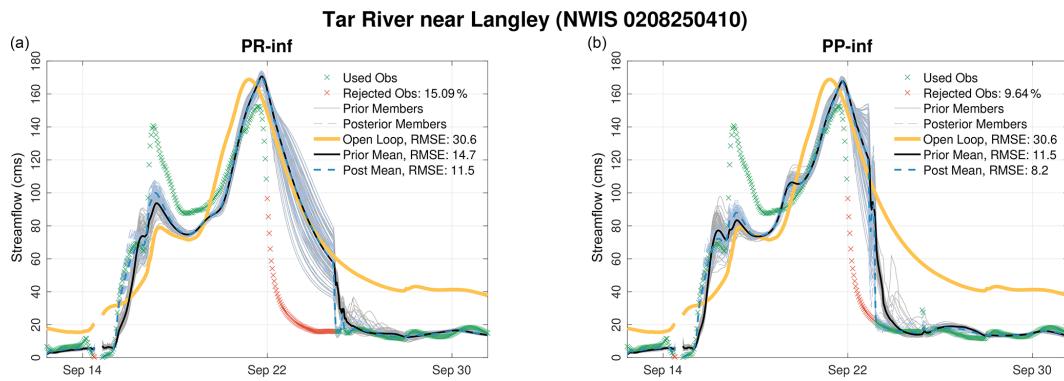
Figure 7. Time series of prior and posterior ensemble means, at the upstream Neuse River near Clayton gauge, resulting from the following four different DA runs: no inflation (**a**), prior inflation (**b**), posterior inflation (**c**), and prior and posterior inflation combined (**d**). The open-loop hydrograph (in cms) is also shown. Assimilated and rejected observations are shown with green and red crosses, respectively. The inflation mean time series is plotted according the right y axes. Time-averaged RMSE for each hydrograph is reported in the legend. The average values of prior, $\sqrt{\lambda^f}$, and posterior, $\sqrt{\lambda^a}$, inflation are also given in the legend.

Figure 8. Similar to Fig. 7 but for the downstream Neuse River near Goldsboro gauge.

Tar River near Langley (NWIS 0208250410)

Figure 9. Similar to Fig. 7 but for Tar River near Langley. Only the PR-inf **(a)** and PP-inf **(b)** results are shown. In addition to the ensemble means, the prior and posterior members are also plotted. Discontinuities in the hydrograph means that there are no observations to compare.

shown to overlap with the open-loop discharge. In the PP-inf run, on the other hand, the falling limbs of the simulated hydrographs are more similar to the data. The recession happens almost 2 d earlier, and this in turn helps the filter reject less data (9.6 % compared to PR-inf run's 15.1 %) and produce higher-quality estimates. The same behavior was observed at a few other gauge locations in the domain (not shown).

Computationally, combining both adaptive prior and posterior inflation schemes is more expensive than running each scheme alone. Our experiments suggest that the extra wall clock time required to perform a full PP-inf run is around 20 % of the total computing time required by PR-inf or PO-inf. In the current framework, the higher complexity is not found to be prohibitive, especially when one takes into account the performance benefits that PP-inf provides. As a future study, it would be interesting to run other PP-inf cases with smaller ensemble size – to match the cost of the PR-inf run – and investigate the performance.

3.5 Inflation in space
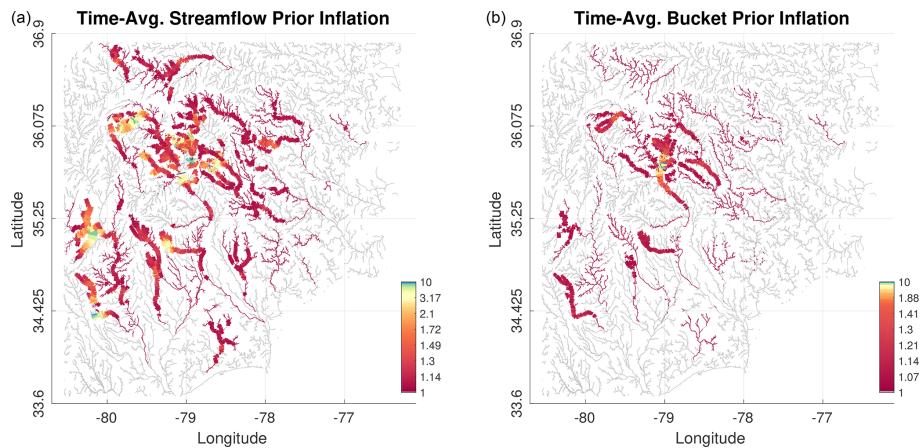
The adaptive inflation varies spatially. With each cycle a different inflation factor is assigned to each value in the state vector. Using cross-correlations in the joint covariance, inflation is therefore computed not only for streamflow but also for the bucket portion of the state. Fig. 10 maps the prior inflation for both streamflow and bucket obtained using PP-inf run. The displayed inflation field is an average over all fields obtained during the flooding period, i.e., between 12 and 18 September. Because of the localized update, the displayed inflation patterns generally follow the tree-like localization shapes (Fig. 4). Inflation values tend to increase near the observation locations and decrease away from the gauges. This is why many reaches, especially in the northeastern part of the domain, have no inflation (i.e., $\sqrt{\lambda^f} = 1$). Given the hourly assimilation of streamflow data, bucket inflation values are relatively smaller than the streamflow ones. Streamflow inflation at more than 90 % of the reaches does not ex-

ceed the value of 2. Reaches with very large inflation values are located in densely observed areas, and the inflation helps restore ensemble spread after multiple, sequential state updates results in loss of spread.
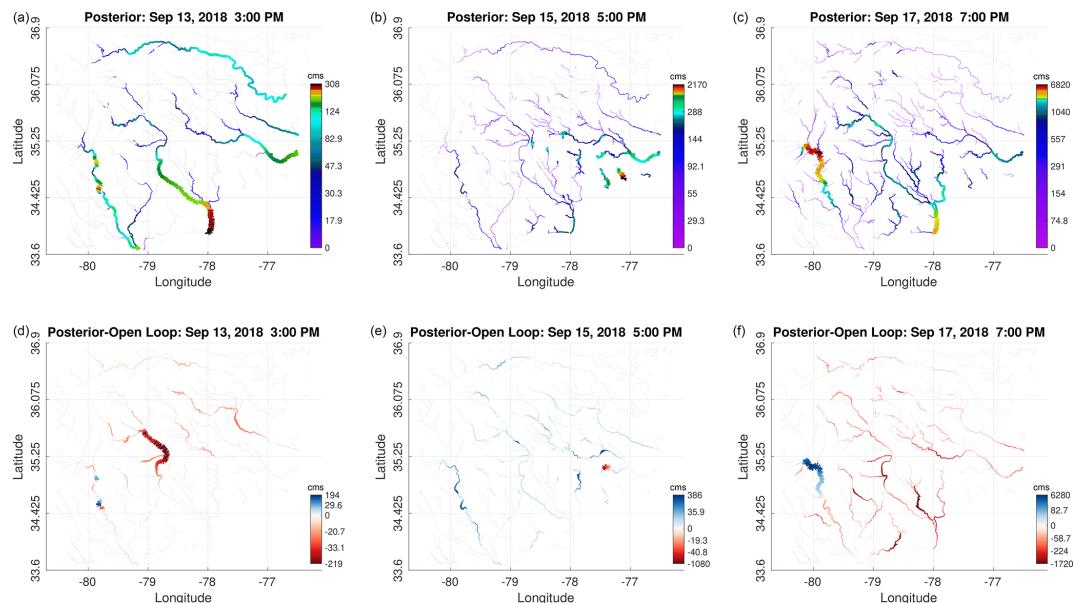
3.6 Overall assessment

Prior to the hurricane landfall on 14 September, streamflow estimates of the model appeared relatively good. The major differences between observed and modeled streamflows resulted from the hurricane. The impact of DA prior to the hurricane is marginal. To investigate this further, we show posterior streamflow maps on 13, 15, and 17 September in Fig. 11a–c. We also show the difference between the posteriors and the open-loop estimates (Fig. 11d–f). Before flooding took place, the highest flow was observed along the Pee Dee, Cape Fear, and Neuse rivers, as shown on 13 September. The difference between the DA result and the open loop is confined to Cape Fear River and is equal to $\sim$ 200 cms. Predicted streamflow on more than 70 % of the reaches in both runs is identical, and hence, the difference is shown to be 0 cms. The differences grow near Neuse River on 15 September to around 1000 cms. The posterior estimate of the streamflow in the rest of the domain on 15 September is generally larger than the open loop (mostly blueish in color). It is notable that streamflow in the domain increased by a factor of 7 before (i.e., maximum of 308 cms on 13 September) and after (i.e., maximum of 2170 cms on 15 September) landfall. On 17 September, the spatial flow distribution changed considerably, especially near the northwestern side of Pee Dee River in which posterior streamflow increased to nearly 7000 cms. Open-loop streamflow estimates are surprisingly small in that area, unlike the rest of the flooded domain.

In order to understand the huge discrepancy between the posteriors and the open-loop results, we study the streamflow evolution at Rocky River (just north of Pee Dee) in Fig. 12. On top of streamflow, we display the mean areal precipitation rates that are used to force the hydrologic model upstream of the gauged streamflow point. As can be seen,
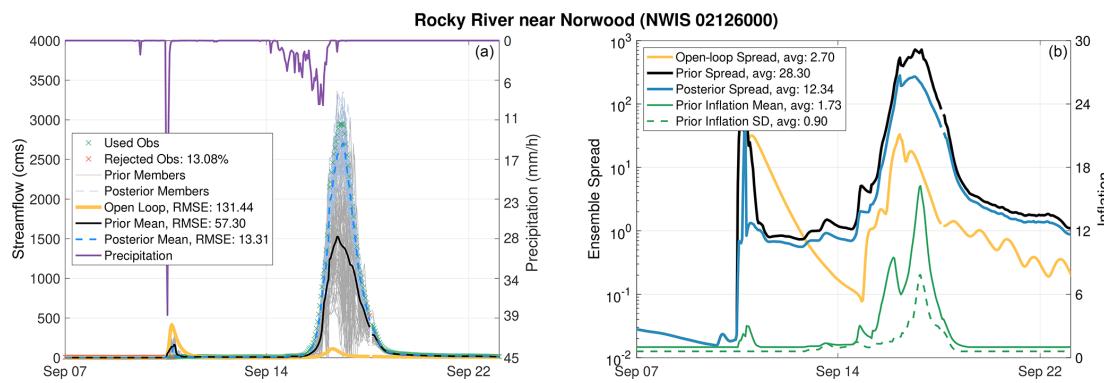
Figure 10. Time-averaged prior inflation for streamflow **(a)** and the bucket **(b)** resulting from PP-inf run. The inflation is averaged over all estimates between 12 and 18 September. Color (log scale) and line thickness both indicate the inflation value. Gray reaches have an inflation value of 1.

Figure 11. (a, b, c) Posterior streamflow ensemble mean maps resulting from a PP-inf run on 13 September at 15:00 UTC (universal coordinated time) **(a, d)**, 15 September at 17:00 **(b, e)**, and 17 September at 19:00 **(c, f)**. **(d, e, f)** Similar maps but for the difference between the posterior means and the open-loop estimates. Reaches with 0 cms flow are shown in gray. Color (log scale) and line thickness both indicate the magnitude value.

the open-loop hydrograph severely underestimates the observed discharge on 17 September. While the observed discharge reaches 3000 cms, the open-loop estimate does not surpass 100 cms. The reason for this huge bias is mainly attributed to the inaccurately specified rainfall rates which do not exceed 10 mm h^{-1} during this time. Before this period, on 10 September, the forcings also falsely simulate heavy rainfall (around 40 mm h^{-1}) prior to the hurricane's landfall. It is surprising how well the DA prior and posterior estimates are given these errors in the precipitation forcing. In fact, by looking at the RMSE values, one finds that the prior and the

posterior estimates are 56 % and 90 % more accurate than the open loop. Such a significant enhancement is obtained due to a massive inflation that is applied to the prior streamflow ensemble. As shown in Fig. 12a, the inflation mean on 17 September grows to 15. This growth was accompanied with a sizable increase in the inflation standard deviation. This further illustrates how powerful the adaptive inflation algorithm is in tackling large biases in the model. It is important to note that if the inflation variance was fixed in time, then the inflation mean will not have the room to grow as much, and hence, the fit to the observed discharge will not
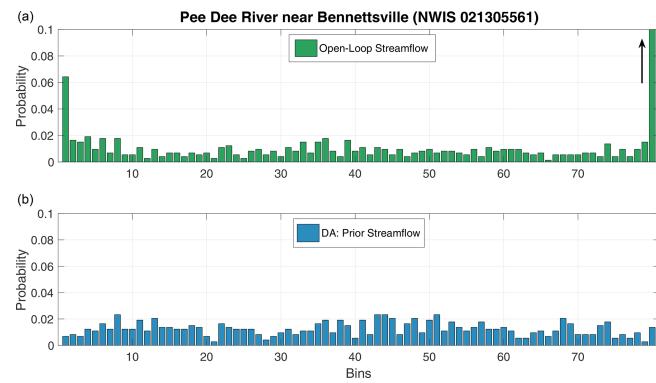
Figure 12. (a) Open-loop, prior, and posterior hydrographs obtained at the Rocky River near Norwood. Precipitation rates computed at that gauge are plotted according the right y axis. **(b)** Open-loop, prior, and posterior ensemble spread (in cms) time series are displayed in log scale. The evolution of the prior inflation mean and standard deviation (SD) is also shown. RMSEs, average ensemble spread and average inflation mean, and standard deviation values are reported in the legends.

be as good. The posterior inflation mean values during the flood (not shown) ranged between 1 and 2. In terms of ensemble spread, due to inflation, the DA estimates are almost 2 orders of magnitude larger than the open loop during the flood. The posterior ensemble spread is consistently smaller than the prior given the continuous hourly shrinkage caused by the Kalman update.
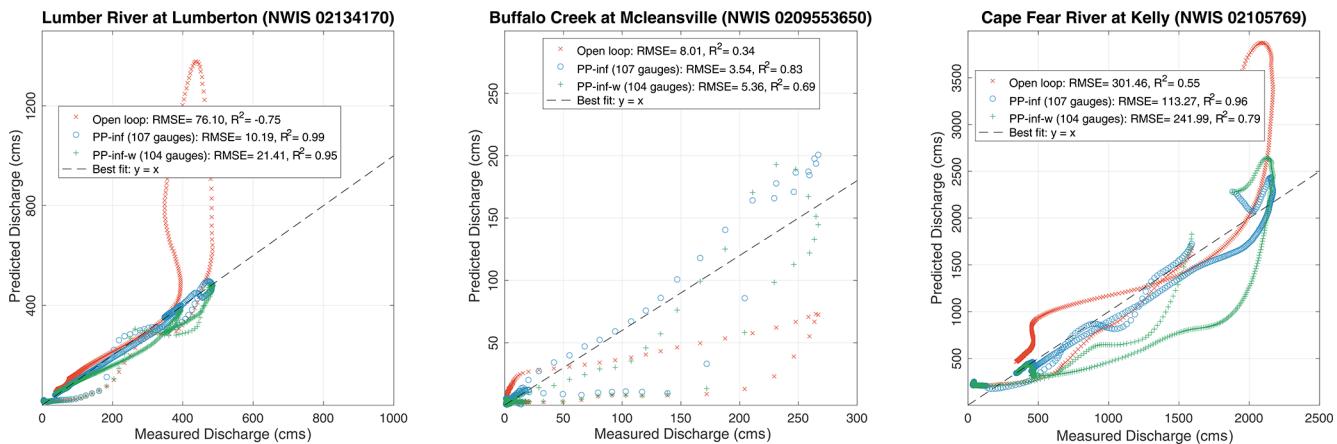
The rank histogram is a useful statistical approach to visualize the behavior of the model and the priors along Pee Dee River. The observed streamflow is binned with respect to the open-loop and prior ensemble members at a single gauge near Bennettsville, South Carolina. The resulting probability bar diagrams are shown in Fig. 13. If the observation falls within the span of the ensemble members at all times, then one would expect to have a flat rank histogram. This is in fact exactly what we obtain for the prior streamflow ensemble (Fig. 13b), making the observed discharge statistically indistinguishable from the ensemble members. As for the open loop, the rank histogram suggests that the probability of the observation falling outside the open-loop ensemble is larger than 50 %. The rank histogram for the open loop is heavily skewed to the right, indicating that the observations are most frequently much larger than the ensemble members, consistent with our previous analysis. The high probability in the first bin of the histogram reflects the open loop's overestimation of the observed streamflow during the no-flood period.

To further assess the performance of the presented DA framework, we run an additional PP-inf experiment, and instead of assimilating all 107 gauges, we withhold three gauges for validation. By withholding gauges, we can infer the impact of the assimilation methods on ungauged points within the domain. The regime at the withheld gauges ranges between relatively low flow at the Buffalo Creek, moderate flow at Lumber River, and high flow at Cape Fear River. Linear regression is performed to validate the streamflow estimates obtained using the open loop, PP-inf (assimilate all

Figure 13. Rank histograms for the open-loop and the prior streamflow obtained at Pee Dee River near Bennettsville. The histograms have been normalized to show probability instead of the observation count. The arrow in the last bin of panel **(a)** indicates that the probability is large (~ 0.5).

107 gauges), and PP-inf-w (withhold three gauges) at the withheld gauges. The resulting analysis is shown in Fig. 14. The PP-inf run is presented as the best case scenario. We check if PP-inf-w can outperform the open loop and how well it approximates PP-inf. Compared to the open loop, the performance of PP-inf-w at the withheld gauges is considerably more accurate. At Lumber River, for instance, the open loop shows a strong overestimation of the observed discharge strongly improved in both assimilation runs. For all three gauges, the estimates from PP-inf-w are able to reasonably mimic those of the PP-inf. Overall, PP-inf-w yields better RMSE and more desirable R^2 (coefficient of determination) values than the open loop. This result indicates that the streamflow at unobserved locations is significantly improved by the assimilation.

Figure 14. Cross-plots of the streamflow at three withheld gauges. The results are shown for the open loop, the PP-inf (where all 107 gauges are assimilated), and PP-inf-w (where only 104 gauges are assimilated). The best-fit line is denoted by a black dashed line. The average RMSE value and the coefficient of determination (R^2) are computed and reported in the legends.

4   Summary and discussion

NOAA's National Water Model (NWM) configuration of the WRF-Hydro framework is coupled to the Data Assimilation Research Testbed (DART) to improve ensemble streamflow forecasts under extreme rainfall conditions during Hurricane Florence in September 2018. Streamflow and bucket head states are simulated using a channel+bucket submodel of the NWM. These states are then updated through data assimilation (DA) using streamflow observations collected from 107 USGS gauges. The system uses 80 ensemble members, incorporating multiphysics uncertainty (each ensemble member assumes different channel model parameters) and time-varying uncertainty in the forcing fluxes to the channel and the bucket models.

This study presents two main contributions within a generalized ensemble DA framework for hydrologic systems, particularly those defined on irregular grids such as a stream network. First, a topologically based along-the-stream (ATS) localization is shown to improve information propagation during the model state. Localizing the impact of the update mitigates sampling errors due to undersampling as well as other analysis errors. Moreover, ATS localization specifically eliminates error covariances between unconnected streams. The algorithm requires tuning of a localization radius, and we do not attempt to diagnose a physical basis for estimating the optimal radius a priori (as such, a discussion should probably include estimation of temporal error covariances not considered in this study). However, ATS localization was found to produce results significantly better than the regular Euclidean distance-based approach. The improved results stem in part from a larger localization radius under the ATS approach, indicating more effective propagation of the observations in the update along the stream than through Euclidean space. While the ATS approach does not further the cause of pre-

dictions in ungauged basins, it indicates that further research into novel localization strategies for streamflow DA may bear additional fruit. On this point, we note that the impact of the ATS localization strategy on the results of this study relative to the impact of adaptive inflation and bias correction is remarkably larger than would be expected in application to atmospheric DA.

The second major contribution of our study is to demonstrate utility of spatially and temporally varying adaptive inflation (El Gharamti, 2018) in hydrologic applications, particularly to help control model bias. Prior and posterior adaptive inflation is shown to mitigate model biases and sampling errors, respectively. Results during major flooding events illustrate that severe model biases can be effectively reduced using adaptive prior inflation. Because the method is spatially varying, different degrees of bias in different parts of the stream network can be efficiently tackled. Posterior inflation was not found to be as effective as prior inflation; however, combining both inflation schemes yielded the highest streamflow accuracy. Overall, inflation plays an indispensable bias correction role, without which the quality of the ensemble streamflow prediction would best be described as poor.

To validate the results of the presented DA system, a variety of diagnostics are presented. Hydrographs at different locations in the domain were investigated. Prior and posterior streamflow estimates were compared to the open-loop result. The largest streamflow improvements were found along Pee Dee River in South Carolina after landfall, during which the observed streamflow was strongly underestimated by the open loop. Improvements due to assimilation were also demonstrated using rank histograms at a gauge along Pee Dee River. Streamflow and inflation spatial maps were also analyzed. It was found that streamflow inflation values are larger than those of the bucket state, given that stream-

flow is directly observed. The overall changes to the bucket state after DA were minimal. To test the impact of DA at non-observed locations, three gauges were withheld from the assimilation and the resulting prior estimates were verified against the data. Linear regression tests revealed that observations at nearby gauges are able to improve the streamflow at the location of the withheld gauges, eventually reducing the systematic biases of the open loop.

The most challenging aspect of the hydrologic DA is the problem of model biases or errors. These biases are usually associated with inaccurate boundary conditions (e.g., precipitation), uncertain parameters (e.g., channel roughness and slope), or model physics deficiencies. This study has shown that adaptive inflation can prove effective at handling biases in the data assimilation. Apart from inflation, a handful of other techniques can be performed to mitigate bias issues. Jointly estimating highly uncertain model parameters alongside the state is an approach commonly found in hydrology (e.g., Vrugt et al., 2006; Gharamti et al., 2015; Abbaszadeh et al., 2018; Ziliani et al., 2019). Updating parameters often increases the complexity of the DA framework (nonlinearity often increases in state–parameters estimation problems), and the computational cost may become prohibitive, especially for spatially varying parameters. Yet, such an approach may yield improvements to the analyses of this study. The multiphysics approach considered here aims to incorporate uncertainty in the fixed boundary condition (geometry and roughness parameters) into the ensemble in order to better model the background error covariance. A combined multiphysics and joint parameter estimation approach might also be pursued. Uncertainty of updated parameters tends to dissipate in time and may be more appropriate for certain kinds of conceptual model parameters instead of those considered in our multiphysics approach. Further up the model chain, not considered in this study, running WRF-Hydro with a land surface model would allow for updating of soil moisture and surface head states. Instead of treating deterministic fluxes with parameterized noise, introducing these prognostic variables would provide the ability to adjust the fluxes coming to the channel. Many studies have tried this and remarked on the problematic updating of soil moisture from streamflow due to the highly nonlinear relationship between the states, particularly for flood forecasting applications (Rakovec et al., 2015). While expanding the prognostic states of the model may potentially improve aspects of the flood prediction problem, such as overland and subsurface fluxes to the channel routing configuration, it is possible that shifting the boundary conditions up the model chain may result in a similar bias issue with more degrees of freedom in the state vector. Coupled atmospheric and hydrologic DA would be a further step towards updating the prognostic states causing hydrologic errors in the state vector. These are ideas to be pursued in future studies.

An essential DA ingredient that this study did not cover is Gaussian anamorphosis (Simon and Bertino, 2009; Gharamti et al., 2017). Streamflow, being strictly nonnegative, is a non-Gaussian variable. Since the Kalman update is linear and assumes Gaussian statistics, it becomes more appropriate to transform streamflow to a Gaussian space where the update is performed, and then it can be pulled back to the physical space. This is well known in hydrological applications (e.g., Clark et al., 2008). Such a transformation guarantees that the updated streamflow does not consist of any unphysical (i.e., negative) values. The transformation is often conducted using empirical functions or analytical ones, such as the natural logarithm. This will be investigated in a follow-up study.
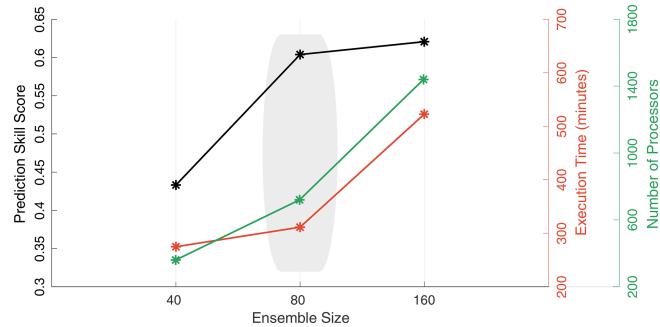
Finally, 1 h ahead (prior) forecasts of flooding event were the focus of this study. Future research will study the impact of DA in the whole forecast time window, up to 18 h in the short-range forecasts, and expand the DA application to medium- and long-range forecasts, including additional hydrologic components and observations. The functionality of the ATS localization and inflation may change in different forecasting modes. For instance, longer localization radii could be found more desirable in a long-range forecast.

Appendix A: Ensemble optimization and localization correlation functions

Figure A1 shows the prediction skill score (PSS) for three different data assimilation experiments using 40, 80, and 160 ensemble members. For a perfect prediction skill, one would expect to obtain a PSS = 1. As can be seen from the plot, the larger the ensemble size, the better the PSS. However, the improvements obtained using 160 members over 80 are relatively marginal (PSS is 0.6 and 0.62 for 80 and 160 members, respectively). Decreasing the ensemble size to 40 degrades the quality of streamflow and yields a PSS of 0.43. On top of the forecast skill, the total runtime and number of processors needed to run the experiments are also plotted. As shown, there is quite an exponential increase in computational time and demand when going from 80 and 160 ensemble members. For instance, the time needed to finish the experiment using 40, 80, and 160 members is 5, 5.5, and 9 h, respectively. Similarly, the number of central processing units (CPUs) required by these respective ensemble runs is 360, 720, and 1440. Based on this analysis, we reckoned that 80 members will give us high enough accuracy without having to utilize a lot of computational power and time. This ensemble size is fixed to 80 members for all experiments described in this study.

Figure A1. Prediction skill score (PSS; in black) obtained for three different WRF-Hydro and DART assimilation experiments where the ensemble size is varied, namely $N_e = 40, 80$, and 160 members. Also plotted as function of the ensemble size is the experiment execution time (right red axis) and the total number of processors needed to run the experiments (right green axis). The shaded region indicates the optimal ensemble size configuration we select in this study. PSS is computed as follows: $1 - \text{MSE}_{\text{DA}}/\text{MSE}_{\text{REF}}$, where MSE denotes mean squared errors, DA means a data assimilation run, and REF is a reference run (the open loop in this case).

Table A1. Localization factor α for different correlation functions is given as function of the absolute distance between the observation gauge and the streamflow reaches, ξ, and the localization radius, r. The distance, ξ is computed along the river network and is equal to the sum of lengths of the reaches separating the observation gauge and the link that is subject to the update. The parameter Ω, used in the Gaspari–Cohn function, is defined as $\frac{2\xi}{r}$.

Type	Correlation function
Gaspari–Cohn	$\alpha(\xi, r) = \begin{cases} 1 - \frac{1}{4}\Omega^5 + \frac{1}{2}\Omega^4 + \frac{5}{8}\Omega^3 - \frac{5}{3}\Omega^2 & 0 \le \xi < \frac{r}{2} \\ \frac{1}{12}\Omega^5 - \frac{1}{2}\Omega^4 + \frac{5}{8}\Omega^3 + \frac{5}{3}\Omega^2 - 5\Omega + 4 - \frac{2}{3}\Omega^{-1} & \frac{r}{2} \le \xi < r \\ 0 & \xi \ge r \end{cases}$
Boxcar	$\alpha(\xi, r) = \begin{cases} 1 & 0 \le \xi < r \\ 0 & \xi \ge r \end{cases}$
Ramped boxcar	$\alpha(\xi, r) = \begin{cases} 1 & 0 \le \xi < \frac{r}{2} \\ 2\frac{r-\xi}{r} & \frac{r}{2} \le \xi < r \\ 0 & \xi \ge r \end{cases}$

References

Abbaszadeh, P., Moradkhani, H., and Yan, H.: Enhancing hydrologic data assimilation by evolutionary particle filter and Markov chain Monte Carlo, Adv. Water Resour., 111, 192–204, 2018.

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., and Avellano, A.: The data assimilation research testbed: A community facility, B. Am. Meteorol. Soc., 90, 1283–1296, 2009.

Anderson, J. L.: A local least squares framework for ensemble filtering, Mon. Weather Rev., 131, 634–642, 2003.

Anderson, J. L.: A non-Gaussian ensemble filter update for data assimilation, Mon. Weather Rev., 138, 4186–4198, 2010.

Ashman, M. S., Socolow, R. S., Zanca, J. L., and Bonito, M. V.: One Hundred Years of Streamflow Measurements in Massachusetts and Rhode Island, US Department of the Interior, US Geological Survey, 2004.

Bauser, H. H., Berg, D., Klein, O., and Roth, K.: Inflation method for ensemble Kalman filter in soil hydrology, Hydrol. Earth Syst. Sci., 22, 4921–4934, https://doi.org/10.5194/hess-22-4921-2018, 2018.

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., Olson, J. B., James, E. P., Dowell, D. C., Grell, G. A., Lin, H., Peckham, S. E., Smith, T. L., Moninger, W. R., Kenyon, J. S., and Manikin, G. S.: A North American hourly assimilation and model forecast cycle: The Rapid Refresh, Mon. Weather Rev., 144, 1669–1694, 2016.

Berner, J., Ha, S.-Y., Hacker, J., Fournier, A., and Snyder, C.: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations, Mon. Weather Rev., 139, 1972–1995, 2011.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J.: Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, Adv. Water Resour., 31, 1309–1324, 2008.

Coxon, G., Freer, J., Westerberg, I., Wagener, T., Woods, R., and Smith, P.: A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations, Adv. Water Resour., 51, 5531–5546, 2015.

DART team: DART, Data Assimilation Research Section (DAReS) National Center for Atmospheric Research (NCAR), DAReS/CISL/NCAR [code], https://doi.org/10.5065/D6WQ0202, 2021.

DeChant, C. M. and Moradkhani, H.: Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting, Water Resour. Res., 48, W04518, https://doi.org/10.1029/2011WR011011, 2012.

Dee, D. P.: Bias and data assimilation, Q. J. Roy. Meteor. Soc., 131, 3323–3343, 2005.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Q. J. Roy. Meteor. Soc., 131, 3385–3396, 2005.

Drécourt, J.-P., Madsen, H., and Rosbjerg, D.: Bias aware Kalman filters: Comparison and improvements, Adv. Water Resour., 29, 707–718, 2006.

El Gharamti, M.: Enhanced adaptive inflation algorithm for ensemble filters, Mon. Weather Rev., 146, 623–640, 2018.

El Gharamti, M.: HydroDART Research Data, Zenodo [data set], https://doi.org/10.5281/zenodo.5532569, 2021.

El Gharamti, M., Raeder, K., Anderson, J., and Wang, X.: Comparing adaptive prior and posterior inflation for ensemble filters using an atmospheric general circulation model, Mon. Weather Rev., 147, 2535–2553, 2019.

Emerton, R. E., Stephens, E. M., Pappenberger, F., Pagano, T. C., Weerts, A. H., Wood, A. W., Salamon, P., Brown, J. D., Hjerdt, N., Donnelly, C., Baugh, C. A., and Cloke, H. L.: Continental and global scale flood forecasting systems, WIREs Water, 3, 391–418, 2016.

Emery, C. M., Biancamaria, S., Boone, A., Ricci, S., Rochoux, M. C., Pedinotti, V., and David, C. H.: Assimilation of wide-swath altimetry water elevation anomalies to correct large-scale river routing model parameters, Hydrol. Earth Syst. Sci., 24, 2207–2233, https://doi.org/10.5194/hess-24-2207-2020, 2020a.

Emery, C. M., David, C. H., Andreadis, K. M., Turmon, M. J., Reager, J. T., Hobbs, J. M., Pan, M., Famiglietti, J. S., Beighley, E., and Rodell, M.: Underlying Fundamentals of Kalman Filtering for River Network Modeling, J. Hydrometeorol., 21, 453–474, 2020b.

Ercolani, G. and Castelli, F.: Variational assimilation of streamflow data in distributed flood forecasting, Water Resour. Res., 53, 158–183, 2017.

Furrer, R. and Bengtsson, T.: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants, J. Multivariate Anal., 98, 227–255, 2007.

Garbrecht, J. and Brunner, G.: Hydrologic channel-flow routing for compound sections, J. Hydraul. Eng., 117, 629–642, 1991.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Q. J. Roy. Meteor. Soc., 125, 723–757, 1999.

Gharamti, M., Ait-El-Fquih, B., and Hoteit, I.: An iterative ensemble Kalman filter with one-step-ahead smoothing for state-parameters estimation of contaminant transport models, J. Hydrol., 527, 442–457, 2015.

Gharamti, M., Samuelsen, A., Bertino, L., Simon, E., Korosov, A., and Daewel, U.: Online tuning of ocean biogeochemical model parameters using ensemble estimation techniques: Application to a one-dimensional model in the North Atlantic, J. Marine Syst., 168, 1–16, 2017.

Gochis, D., Barlage, M., Cabell, R., Dugger, A., Fanfarillo, A., FitzGerald, K., McAllister, M., McCreight, J., RafieeiNasab, A., Read, L., Frazier, N., Johnson, D., Mattern, J. D., Karsten, L., Mills, T. J., and Fersch, B.: WRF-Hydro® v5.1.1, Zenodo [data set], https://doi.org/10.5281/zenodo.3625238, 2020.

Gochis, J. and Chen, F.: Hydrological enhancements to the community Noah land surface model, University Corporation for Atmospheric Research, NCAR Scientific Technical Report No. NCAR/TN-454+STR, https://doi.org/10.5065/D60P0X002003, 2003.

Guha-Sapir, D., Santos, I., Borde, A. (Eds.): The economic impacts of natural disasters, Oxford University Press, New York, NY, ISBN: 978-0-19-984193-6, 2013.

Hodyss, D. and Nichols, N.: The error of representation: Basic understanding, Tellus A, 67, 24822, https://doi.org/10.3402/tellusa.v67.24822, 2015.

Hoteit, I., Pham, D.-T., Gharamti, M., and Luo, X.: Mitigating observation perturbation sampling errors in the stochastic EnKF, Mon. Weather Rev., 143, 2918–2936, 2015.

Houtekamer, P. L. and Mitchell, H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation, Mon. Weather Rev., 129, 123–137, 2001.

Jamal, A. and Linker, R.: Inflation method based on confidence intervals for data assimilation in soil hydrology using the ensemble Kalman filter, Vadose Zone J., 19, e20000, https://doi.org/10.1002/vzj2.20000, 2020.

Jozaghi, A., Nabatian, M., Noh, S., Seo, D.-J., Tang, L., and Zhang, J.: Improving Multisensor Precipitation Estimation via Adaptive Conditional Bias–Penalized Merging of Rain Gauge Data and Remotely Sensed Quantitative Precipitation Estimates, J. Hydrometeorol., 20, 2347–2365, 2019.

Kitanidis, P. K. and Bras, R. L.: Real-time forecasting with a conceptual hydrologic model: 2. Applications and results, Water Resour. Res., 16, 1034–1044, 1980.

Lee, H. and Seo, D.-J.: Assimilation of hydrologic and hydrometeorological data into distributed hydrologic model: Effect of adjusting mean field bias in radar-based precipitation estimates, Adv. Water Resour., 74, 196–211, 2014.

Lee, H., Shen, H., Noh, S. J., Kim, S., Seo, D.-J., and Zhang, Y.: Improving flood forecasting using conditional bias-penalized ensemble Kalman filter, J. Hydrol., 575, 596–611, 2019.

Li, Y., Ryu, D., Western, A. W., and Wang, Q.: Assimilation of stream discharge for flood forecasting: The benefits of accounting for routing time lags, Water Resour. Re., 49, 1887–1900, 2013.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, Hydrol. Earth Syst. Sci., 16, 3863–3887, https://doi.org/10.5194/hess-16-3863-2012, 2012.

McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A.: NHDPlus version 2: User guide, US Environmental Protection Agency, 2012.

McMillan, H. K., Hreinsson, E. Ö., Clark, M. P., Singh, S. K., Zammit, C., and Uddstrom, M. J.: Operational hydrological data assimilation with the recursive ensemble Kalman filter, Hydrol. Earth Syst. Sci., 17, 21–38, https://doi.org/10.5194/hess-17-21-2013, 2013.

Moradkhani, H., Hsu, K.-L., Gupta, H., and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resour. Res., 41, W05012, https://doi.org/10.1029/2004WR003604, 2005.

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

Noh, S. J., Tachikawa, Y., Shiiba, M., and Kim, S.: Ensemble Kalman filtering and particle filtering in a lag-time window for short-term streamflow forecasting with a distributed hydrologic model, J. Hydrol. Eng., 18, 1684–1696, 2013.

Pauwels, V. R. and De Lannoy, G. J.: Improvement of modeled soil wetness conditions and turbulent fluxes through the assimilation of observed discharge, J. Hydrometeorol., 7, 458–477, 2006.

Pauwels, V. R. and De Lannoy, G. J.: Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space, Water Resour. Res., 45, W08428, https://doi.org/10.1029/2008WR007590, 2009.

Ponce, V. M. and Lugo, A.: Modeling looped ratings in muskingum-cunge routing, J. Hydrol. Eng., 6, 119–124, 2001.

Ponce, V. M. and Yevjevich, V.: Muskingum-Cunge method with variable parameters, Journal of the Hydraulics Division, 104, 1663–1667, 1978.

Rafieeinasab, A., Seo, D.-J., Lee, H., and Kim, S.: Comparative evaluation of maximum likelihood ensemble filter and ensemble Kalman filter for real-time assimilation of streamflow data into operational hydrologic models, J. Hydrol., 519, 2663–2675, 2014.

Rakovec, O., Weerts, A. H., Sumihar, J., and Uijlenhoet, R.: Operational aspects of asynchronous filtering for flood forecasting, Hydrol. Earth Syst. Sci., 19, 2911–2924, https://doi.org/10.5194/hess-19-2911-2015, 2015.

Rasmussen, J., Madsen, H., Jensen, K. H., and Refsgaard, J. C.: Data assimilation in integrated hydrological modelling in the presence of observation bias, Hydrol. Earth Syst. Sci., 20, 2103–2118, https://doi.org/10.5194/hess-20-2103-2016, 2016.

Ridler, M.-E., Zhang, D., Madsen, H., Kidmose, J., Refsgaard, J. C., and Jensen, K. H.: Bias-aware data assimilation in integrated hydrological modelling, Hydrol. Res., 49, 989–1004, 2018.

Seo, D.-J., Koren, V., and Cajina, N.: Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting, J. Hydrometeorol., 4, 627–641, 2003.

Seo, D.-J., Cajina, L., Corby, R., and Howieson, T.: Automatic state updating for operational streamflow forecasting via variational data assimilation, J. Hydrol., 367, 255–275, 2009.

Seo, D.-J., Saifuddin, M. M., and Lee, H.: Conditional bias-penalized Kalman filter for improved estimation and prediction of extremes, Stoch. Env. Res. Risk A., 32, 183–201, 2018.

Simon, E. and Bertino, L.: Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: a twin experiment, Ocean Sci., 5, 495–510, https://doi.org/10.5194/os-5-495-2009, 2009.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, Hydrolog. Sci. J., 48, 857–880, 2003.

Sun, L., Nistor, I., and Seidou, O.: Streamflow data assimilation in SWAT model using Extended Kalman Filter, J. Hydrol., 531, 671–684, 2015.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, 2001.

Vrugt, J. A., Gupta, H. V., Nualláin, B., and Bouten, W.: Real-time data assimilation for operational ensemble streamflow forecasting, J. Hydrometeorol., 7, 548–565, 2006.

Weerts, A. H. and El Serafy, G. Y.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resour. Res., 42, W09403, https://doi.org/10.1029/2005WR004093, 2006.

Wood, E. F. and Szöllösi-Nagy, A.: An adaptive algorithm for analyzing short-term structural and parameter changes in hydrologic prediction models, Water Resour. Res., 14, 577–581, 1978.

WRF-Hydro team: WRF-Hydro, RAL/NCAR [code], available at: https://ral.ucar.edu/projects/wrf_hydro, last access: 15 September 2021.

Xia, Y., Sheffield, J., Ek, M. B., Dong, J., Chaney, N., Wei, H., Meng, J., and Wood, E. F.: Evaluation of multi-model simulated soil moisture in NLDAS-2, J. Hydrol., 512, 107–125, 2014.

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., Grams, H., Wang, Y., Cocks, S., Martinaitis, S., Arthur, A., Cooper, K., Brogden, J., and Kitzmiller, D.: Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities, B. Am. Meteorol. Soc., 97, 621–638, 2016.

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