https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions

Leveraging a Novel Hybrid Ensemble and Optimal Interpolation Approach for Enhanced Streamflow and Flood Prediction

Mohamad El Gharamti[1], Arezoo RafieeiNasab[2], and James L. McCreight[3]

[1]NCAR, Computational and Information Systems Laboratory (CISL), Boulder CO, USA
[2]NCAR, Research Applications Laboratory (RAL), Boulder CO, USA
[3]UCAR, CPAESS Cooperative Programs for the Advancement of Earth System Science, Boulder CO, USA

Correspondence: Moha Gharamti (gharamti@ucar.edu)

Abstract. In the face of escalating instances of inland and flash flooding spurred by intense rainfall and hurricanes, accurately predicting rapid streamflow variations has become imperative. Traditional data assimilation methods face challenges during extreme rainfall events due to numerous sources of error, including model deficiencies, forcing biases and observational uncertainties. This study introduces a cutting-edge hybrid ensemble and optimal interpolation data assimilation scheme tailored to precisely and efficiently estimate streamflow during such critical events. Our hybrid scheme builds upon the ensemble-based framework of El Gharamti et al., integrating the flow-dependent background streamflow covariance with a climatological error covariance derived from historical model simulations. The dynamic interplay (weight) between the static background covariance and the evolving ensemble is adaptively computed both spatially and temporally. By coupling the National Water Model (NWM) configuration of the WRF-Hydro modeling system with the Data Assimilation Research Testbed (DART), we validate the performance of our hybrid prediction system using two impactful test cases: 1. West Virginia's flash flooding event in June 2016, and 2. Florida's inland flooding during Hurricane Ian in September 2022. Our findings reveal that the hybrid scheme significantly outperforms its ensemble counterpart, delivering enhanced streamflow estimates for both low and high flow scenarios, with an improvement of up to 50%. This heightened accuracy is attributed to the climatological background covariance, mitigating bias and augmenting ensemble variability. The adaptive nature of the hybrid algorithm ensures reliability even with a very small time-varying ensemble. Moreover, this innovative hybrid data assimilation system propels streamflow forecasts up to 18 hours in advance of flood peaks, marking a substantial advancement in flood prediction capabilities.

1 Introduction

Flooding can stem from various causes, including prolonged rainfall events like tropical storms or hurricanes, as well as intense rainfall over short periods or complications such as debris and ice jams. When examining events causing at least a billion dollars in damage, river and urban flooding alone account for 7.4% of US natural disasters from 1980 to 2023. Tropical cyclones top the list, contributing to 52% of the damage (Smith, 2020).

Tropical storms and hurricanes are characterized by destructive winds, storm surge, and catastrophic flooding. Hurricanes can unleash 2.4 trillion gallons of rainwater in a day (Nunez and Yang, 2023). According to the National Weather Service, torrential rain from hurricanes can flood the neighborhoods of coastal communities within minutes. This phenomenon, known

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024


Hydrology and Earth System Sciences
Discussions

as freshwater or inland flooding, can damage infrastructure, cause landslides, destroy crops, and take lives. Approximately, 25% of US hurricane deaths from 1963 to 2012 were related to freshwater flooding (Rappaport, 2014). Predicting floods has the potential to save lives, protect infrastructure, and minimize socio-economic impacts. Such predictions are challenging and remain an area of active research and operational development.

Streamflow predictions commonly integrate real-world observations with hydrologic model simulations, a practice known as data assimilation (DA) (Yucel et al., 2015). This approach, prevalent across diverse fields including numerical weather prediction (NWP) (Lorenc, 1986), has witnessed substantial growth in streamflow prediction applications over the last two decades (e.g., Kim et al., 2021; Chao et al., 2022). Ensemble filtering techniques, based on the Kalman filter (Kalman, 1960), are widely adopted due to their ease of implementation and high portability. The ensemble Kalman filter (EnKF) (Evensen, 2003) stands out among these methods, utilizing model forecast realizations during the analysis. The EnKF employs a minimum variance estimator, utilizing observations to compute ensemble increments that are subsequently linearly combined with the predicted ensemble. Extensively used in real-time and operational settings, the EnKF has been applied to enhance streamflow simulations. For instance, Rakovec et al. (2012) explored its performance in updating streamflow through synthetic and real-world experiments in the Meuse River basin, Belgium. In an operational hydrological context, McMillan et al. (2013) introduced a recursive EnKF variant stabilizing streamflow estimates across different catchments in New Zealand. Rafieeinasab et al. (2014) compared the performance of the maximum likelihood ensemble filter with the EnKF for real-time streamflow assimilation in a Southern Texas headwater basin. Meanwhile, Huang et al. (2017) utilized the EnKF to update streamflow using snow water equivalent data in basins like the Pacific Northwest, Rocky Mountains, and California in the western US. Additionally, Lee et al. (2019) developed a bias-aware version of the EnKF to enhance flood forecasting for a subset of 10 headwater basins in the Southern USA. Beyond the EnKF, various methods, including particle filters (e.g., Weerts and El Serafy, 2006; Noh et al., 2013; Abbaszadeh et al., 2020), ensemble smoothers (e.g., Meng et al., 2017), variational methods (e.g., Seo et al., 2009; Mazrooei et al., 2021), and machine learning (e.g., Boucher et al., 2020), have been employed for streamflow and flood prediction across diverse spatial and temporal scales. A comprehensive review of methods used in streamflow prediction over the past four decades is available in Troin et al. (2021).

Despite its achievements in research, real-time applications, and operational frameworks, the EnKF operates as a sequential estimation tool with inherent limitations, grappling with biases and sampling errors. The EnKF approximates the true prior (or forecast) covariance by employing a sample covariance from the ensemble, assuming unbiased background errors. In scenarios with small ensemble sizes, the estimated sample covariance may be plagued by substantial sampling errors, leading to suboptimal analysis (or posterior) estimates. These errors often result in the underestimation of the true ensemble variance, potentially causing filter divergence in severe situations (Furrer and Bengtsson, 2007). Moreover, the use of a restricted ensemble size can introduce spurious correlations in space, proving detrimental over successive assimilation cycles (Anderson, 2012). Model biases, typically unaccounted for in the EnKF, tend to dominate other error sources, potentially causing catastrophic consequences such as complete ensemble collapse (Sacher and Bartello, 2008). Localization and inflation are commonly employed methods to address sampling errors. Localization (Houtekamer and Mitchell, 2001) restricts the impact of observations to nearby state variables, mitigating nonphysical correlations, especially between spatially distant observations and state vari-

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024


Hydrology and Earth System Sciences
Discussions

ables. Inflation (Anderson and Anderson, 1999) increases the ensemble spread around its mean, countering sampling errors and enhancing the fit to observations. Both inflation and localization have become integral tools in the majority of streamflow ensemble DA studies in the literature (e.g., Emery et al., 2020). Additional techniques addressing sampling errors encompass relaxation of prior spread and perturbations (Zhang et al., 2004; Whitaker and Hamill, 2012), the utilization of stochastic perturbations and multi-physics ensembles (e.g., Berner et al., 2011), and covariance hybridization (Hamill and Snyder, 2000). Beyond sampling errors and biases, issues such as non-linearity and non-Gaussianity frequently compromise the optimality of the EnKF update step (Anderson, 2010).

Integrating the EnKF with other DA methods can mitigate some of its shortcomings. For instance, optimal interpolation (OI) or three-dimensional variational (3D-Var) systems[1] avoid sampling errors by depending on a time-invariant background error covariance, typically estimated from the model's climatology. This approach has been extensively explored in atmospheric and ocean DA literature (Counillon et al., 2009; Bannister, 2017, and references therein). However, in the realm of streamflow applications, the utilization of hybrid ensemble and variational DA techniques remains limited and is actively researched. In a study on high-resolution hydrologic forecasting, Hernández and Liang (2018) investigated streamflow predictive accuracy using a hybrid scheme that combines particle filtering (PF) with four-dimensional variational (4D-Var) data assimilation. The authors reported the hybrid algorithm's ability to provide skillful streamflow predictions, accommodating non-Gaussian, non-linear, and high-resolution estimation. In another exploration, Abbaszadeh et al. (2019) developed a similar hybrid PF and 4D-Var system, assessing its performance across several river basins in the US. Their hybrid scheme, adept at handling various sources of uncertainties, proved efficient and robust to the number of particles. In this work, we delve into the functionality, performance, and efficiency of a hybrid EnKF and OI scheme, hereafter referred to as EnKF-OI, in the context of flash flooding and freshwater events. We meticulously examine the weighting between the ensemble-based flow-dependent covariance and the climatological static background covariance. Furthermore, we implement the adaptive hybrid weighting scheme proposed by El Gharamti (2021) and investigate the temporal and spatial variations of the weighting factor across the stream network. Our analysis extends to assessing the impacts of the hybrid scheme on streamflow biases and sampling errors. To our knowledge, this marks the inaugural application of an adaptive hybrid EnKF and 3D-Var DA scheme to real-world streamflow flooding problems.

The data assimilation framework employed in this study is based on the integrated HydroDART system, as detailed by El Gharamti et al. (2021). This system utilizes NOAA's National Water Model (NWM) configuration within the WRF-Hydro hydrological framework (Gochis et al., 2020). Our implementation involves a sub-model configuration of the NWM, specifically designed to incorporate both streamflow and conceptual groundwater storage. To facilitate the assimilation process, the model is interfaced with the Data Assimilation Research Testbed (DART, Anderson et al., 2009). The HydroDART system, as highlighted in El Gharamti et al. (2021), incorporates Along-The-Stream (ATS) localization, along with adaptive prior and posterior inflation, to enhance ensemble performance. To extend the capabilities of HydroDART, we introduce a hybrid method

[1]Throughout this study, the terms OI and 3D-Var are used interchangeably, both solving the same DA problem and relying on the same static background error covariance matrix.

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024


Hydrology and Earth System Sciences
Discussions

using a 42-year retrospective run of WRF-Hydro covering the entire Contiguous United States (CONUS). This extensive model run is leveraged to construct climatological error covariances for both streamflow and groundwater bucket storage.

Our EnKF-OI prediction system undergoes testing in two regional flooding scenarios. The initial case, spanning June 2016, addresses an 11-day flash flooding event in West Virginia. The second case study, from September 15th to October 15th, 2022, focuses on inland flooding caused by Hurricane Ian in central Florida. In both instances, streamflow observations obtained from United States Geological Survey (USGS) gauges are assimilated hourly. Our evaluation encompasses an assessment of the performance of both hybrid (EnKF-OI) and non-hybrid (EnKF) DA methodologies in improving simulated hydrographs under diverse flooding conditions. To gain insights into the behavior of the hybrid EnKF-OI algorithm, we conduct sensitivity experiments concerning the hybrid weighting coefficient and the ensemble size. Furthermore, the analyses derived from the hybrid scheme's posterior states contribute to generating short-range streamflow forecasts, allowing us to evaluate the impact of data assimilation on these predictions.

The subsequent sections delineate the paper's organization. Section 2 delves into the detailed description of the integrated HydroDART prediction system, shedding light on both the model and the DA tool. Particular attention is paid to elucidating the methodology employed in generating the static-background covariance. Section 3 expounds on the specifics of the test cases, providing insights into the hydrologic domains' extent and elaborating on the USGS observations. Moving forward, Section 4 unfolds the results obtained from the EnKF and the hybrid EnKF-OI methodologies across the two hydrologic domains. Distinct spatial and temporal evaluations underscore the hybrid algorithm's performance. The concluding insights and broader discussions emanate in Section 5, encapsulating a comprehensive summary of the findings.

2 Model and Methods

2.1 WRF-Hydro

The Weather Research and Forecasting, Hydrological modeling system (WRF-Hydro) is used widely across the hydrology community both in coupled and uncoupled modes with atmospheric models (e.g., Senatore et al., 2015; Kerandi et al., 2018; Wang et al., 2022; Son et al., 2023). A prominent application of WRF-Hydro is the National Water Model (NWM) which became operational in August 2016 and has gone through several version upgrades since then. The National Water Model is a particular configuration of the uncoupled WRF-Hydro system which is operational over CONUS, Hawaii, Puerto Rico and Virgin Islands (NWMv2.1) and Alaska (NWMv3.0). The modules and physics in this paper are equivalent to the NWM version 2.1 standard analysis and assimilation cycle without the streamflow nudging used in the NWM. The NWM v2.1 configuration consists of the Noah-MP land surface model, subsurface and surface flow routing, baseflow/groundwater routing, channel and reservoir routing. Because the channel, reservoir, and conceptual groundwater components are one-way coupled to the other model components, following El Gharamti et al. (2021), a channel, reservoir, and conceptual groundwater submodel of the NWM is used. This configuration is computationally cheaper compared to the full model and therefore appealing for running an ensemble system. As detailed in El Gharamti et al. (2021), the channel routing module is based on the Muskingum-Cunge river routing with variable parameters in a compound channel. Inflow fluxes to lakes and reservoir objects embedded into

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions

the NWM routing network are routed using level pool scheme. The groundwater routing is a highly conceptual model which discharge the base flow following an exponential equation and spills over if the conceptual depth of water in bucket exceed the maximum head. The prognostic variables that are updated in this study are the streamflow discharge and groundwater bucket head. It should be noted that the lake/reservoir objects are defined on the stream reaches, however they are not considered in the state updating. Figure 1 shows the chain of modeling components in our system and how the deterministic fluxes from the full NWM model arrive as boundary conditions to the HydroDART system.

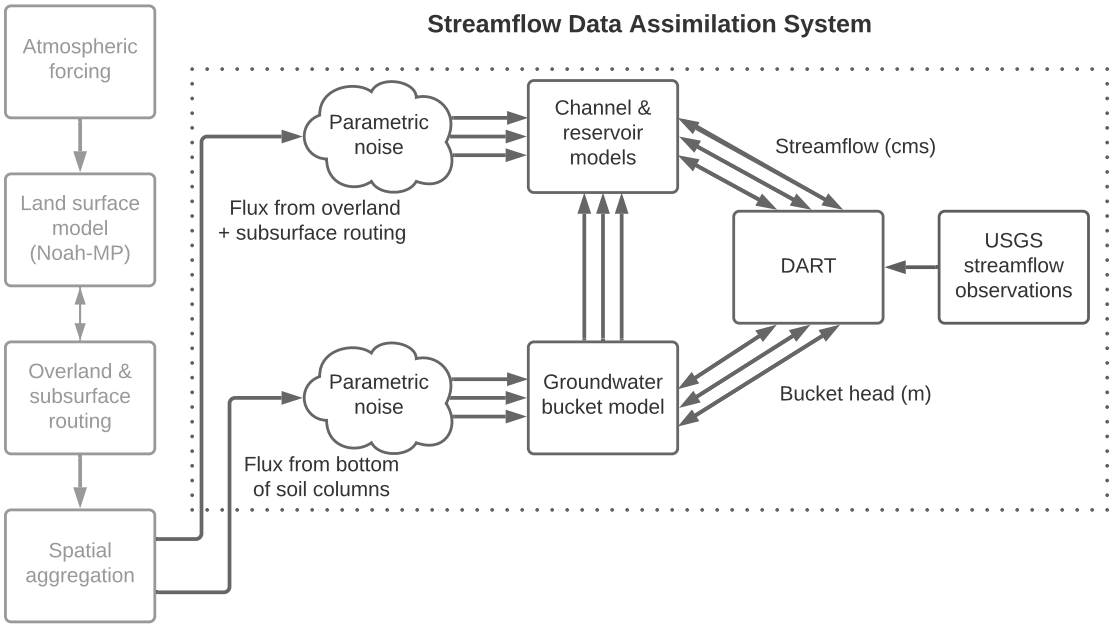


Figure 1. Streamflow data assimilation workflow adapted from El Gharamti et al. (2021). The vertical boxes on the left depict the WRF-Hydro components that are executed once and produce the deterministic fluxes into the channels and groundwater buckets. The arrows represent the order of the physics component execution and a two way arrow represents two way coupling. Dotted box represents the HydroDART components. Random noise is applied to the deterministic input fluxes to HydroDART generating an ensemble of forcing for both channel and conceptual groundwater models. Three arrows denote the presence of ensembles, however, the ensemble size is larger than three. As shown in this workflow, DART assimilates USGS streamflow observations and updates the streamflow (in cubic meters per second) and bucket head (in meters) state variables in the channel and reservoir model, and groundwater bucket model, respectively.

Random noise is applied to these deterministic boundary fluxes to generate an ensemble of forcings for the streamflow, reservoir, and conceptual groundwater system. In addition to this time-varying uncertainty, we also generate an invariant ensemble of channel parameters which provides a "multiphysics" streamflow ensemble (El Gharamti et al., 2021). These two levels of perturbations were found necessary to generate larger variability in the predicted ensemble. Because ensemble DA relies on

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024


Hydrology and Earth System Sciences
Discussions

probabilistic forecasts, increasing the variability in the ensemble can help fit the observations and obtain accurate hydrologic states.

To perturb the boundary fluxes to the streamflow and bucket models, we use Gaussian samples with mean of zero and standard deviation equal to 40% of the flux value at each location. We perturb the channel parameters using uniform noise distributions. For detailed description of the WRF-Hydro configuration used in this paper as well as the creation of the forcing ensemble and channel parameter ensemble, please refer to El Gharamti et al. (2021). Here, the model code (https://github.com/NCAR/wrf_hydro_nwm_public/releases/tag/nwm-v2.1-beta3), domain data and parameter sets (https://www.nco.ncep.noaa.gov/pmb/codes/nwprod/nwm.v2.2.3/parm/domain/) are based on the NWMv2.1. This means that there is a level of model calibration that reduces some of the model background biases, however, there still exists some biases in the model.

2.2 DART

2.2.1 Ensemble Filtering

The Data Assimilation Research Testbed (DART, Anderson et al., 2009) is a sequential ensemble DA tool developed and maintained at the National Center for Atmospheric Research. DART employs different variants of the ensemble Kalman filter for linear and nonlinear estimation problems. The filtering schemes are based on Bayes rule such that the prior distribution is recursively updated using the observation likelihood to obtain a posterior probability density function (pdf). For HydroDART, several streamflow realizations are first integrated forward in time using the hydrologic model, WRF-Hydro, denoted by $\mathcal{M}$, until the next observations become available.

$$\mathbf{x}_k^{f,i} = \mathcal{M}\left(\mathbf{x}_{k-1}^{a,i}, \boldsymbol{\theta}^i, \boldsymbol{\gamma}_k^i\right), \qquad i = 1, 2, \ldots, N_e \tag{1}$$

Here, $\mathbf{x}$ denotes the hydrologic state which consists of streamflow and the conceptual groundwater storage. The superscripts f, a, i denote forecast, analysis and ensemble member, respectively. The subscript k denotes time and N_e is the ensemble size. $\boldsymbol{\theta}$ is a set of 6 physical parameters that describe the geometry of the streamflow compound channel. The parameters are: top channel width, bottom width, side slope, Manning's N, width of the compound channel and Manning's N of the compound channel. Notice that for each member, a different configuration of these channel parameters is used following the multiphysics approach outlined in El Gharamti et al. (2021). Perturbed boundary fluxes to the streamflow and the groundwater models are given in $\boldsymbol{\gamma}$. Using equation (1), the first and second moments of the prior distribution are approximated as follows:

$$\overline{\mathbf{x}}_k^f = \frac{1}{N_e}\sum_{i=1}^{N_e}\mathbf{x}_k^{f,i}, \tag{2}$$

$$\mathbf{P}_k^f = \lambda \cdot \overline{\mathbf{P}}_k^f, \tag{3}$$

$$= \lambda \cdot \frac{1}{N_e - 1}\sum_{i=1}^{N_e}\left(\mathbf{x}_k^{f,i} - \overline{\mathbf{x}}_k^f\right)\left(\mathbf{x}_k^{f,i} - \overline{\mathbf{x}}_k^f\right)^T, \tag{4}$$

where $\overline{\mathbf{x}}_k^f$ and $\overline{\mathbf{P}}_k^f$ are the prior ensemble mean and the prior sample covariance, respectively. The coefficient λ in equation (4) is an inflation factor used to restore the ensemble variance after the forecast. In this study, the inflation factor is selected

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024


Hydrology and Earth System Sciences
Discussions
Open Access
EGU

adaptively in space and time using the adaptive inflation scheme of El Gharamti (2018). We also apply adaptive posterior inflation to the analysis ensemble according to El Gharamti et al. (2019). The estimated (inflated) prior covariance at time t_k is given by $\mathbf{P}_k^f$.

At the time of the analysis, DART assimilates the observations serially according to Anderson (2003):

$$\mathbf{x}_{j,k}^{a,i} = \mathbf{x}_{j,k}^{f,i} + \rho \frac{\sigma_{xy}}{\sigma_y^2} \Delta y^i, \qquad j = 1, 2, \ldots, N_x. \tag{5}$$

The subscript j is an index to the variables in the state and the total number of state variables is denoted by N_x. σ_{xy} is the prior covariance between the observation y and j^{th} state element while σ_y^2 is the sample variance of the observed variable. As can be shown, the EnKF solution is obtained as a linear regression of the observation increments Δy on the entire state vector. ρ is a localization coefficient typically between 0 and 1 and is computed using the common Gaspari-Cohn correlation function (Gaspari and Cohn, 1999). The localization strategy employed here follows the ATS localization scheme (El Gharamti et al., 2021) such that only reaches upstream and downstream from a particular gauge are updated during the analysis.

2.2.2 Hybridized Covariance

Hybridizing the prior ensemble covariance matrix is performed linearly right before the update:

$$\mathbf{P}_k^h = \alpha \mathbf{P}_k^f + (1 - \alpha)\mathbf{B}, \tag{6}$$

where h denotes the hybrid form of the prior covariance and $\mathbf{B}$ is a static background covariance matrix. α is a weighting factor chosen between 0 and 1. Notice that $\mathbf{B}$ is a stationary covariance and is typically used in 3-4D variational systems. $\mathbf{B}$ can be estimated from the model's climatology using a large inventory of historical forecasts. In NWP systems, for instance, the National Meteorological Center (NMC, Parrish and Derber, 1992) technique is often used to compute $\mathbf{B}$. More details on the computation of $\mathbf{B}$ for the current streamflow work can be found in the next section. From equation (6), one may argue that $\mathbf{P}_k^h$ is a better estimate of the true covariance compared to the ensemble one $\mathbf{P}_k^f$. This is because blending in climatological information in the ensemble covariance usually presents the ensemble with new correction directions (Hamill and Snyder, 2000). When α is set to 1, equation (6) results in a purely flow-dependent covariance reducing the algorithm to the EnKF. In contrast, when $\alpha = 0$ the system morphs into an ensemble optimal interpolation (EnOI) scheme. The hybrid EnKF-OI scheme is activated for $0 < \alpha < 1$.

Rather than manually tuning α, one can adaptively estimate it (El Gharamti, 2021) using Bayes rule:

$$p(\alpha|\mathbf{d}_k) \approx p(\alpha) \cdot p(\mathbf{d}_k|\alpha), \tag{7}$$

where $\mathbf{d}_k$ is the background innovation (i.e., observation minus forecast). Equation (7) assumes that α is a random variable with prior distribution $p(\alpha)$ which is considered Gaussian in this work. The likelihood term on the right hand side of (7) is also assumed Gaussian with zero mean and covariance:

$$\mathbb{E}\left[\mathbf{d}_k \mathbf{d}_k^T\right] = \mathbf{R}_k + \mathbf{H}_k \mathbf{P}_k^t \mathbf{H}_k^T, \tag{8}$$

$$\approx \mathbf{R}_k + \alpha \mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + (1 - \alpha)\mathbf{H}_k \mathbf{B} \mathbf{H}_k^T. \tag{9}$$

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024


Hydrology and Earth System Sciences
Discussions
Open Access
EGU

This result holds as long as the background and observation errors are uncorrelated (Desroziers et al., 2005). $\mathbf{H}_k$ is the observation operator and $\mathbf{R}_k$ is the observation error covariance matrix. The hybridized covariance $\mathbf{P}_k^h$ is assumed to be equivalent to the true background covariance $\mathbf{P}_k^t$ allowing the insertion of equation (6) in (8). Using equation (9), the likelihood of the weighting coefficient can be written as:

$$p(\mathbf{d}_k|\alpha) = \left(\sqrt{2\pi}\beta\right)^{-1} \exp\left[-\frac{\mathbf{d}_k^T\mathbf{d}_k}{2\beta^2}\right], \text{ where} \tag{10}$$

$$\beta = \sqrt{\sum_i (\mathbf{R}_k)_{ii} + \alpha \sum_i (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T)_{ii} + (1-\alpha) \sum_i (\mathbf{H}_k \mathbf{B} \mathbf{H}_k^T)_{ii}}. \tag{11}$$

The notation $\sum_i \mathbf{A}_{ii}$ is equivalent to the trace of matrix $\mathbf{A}$. Taking the product of the likelihood and the prior of α results in a Gaussian posterior which can be used in subsequent DA cycles. The updated value of α at time t_k is obtained by maximizing its posterior $p(\alpha|\mathbf{d}_k)$ pdf. Following this formulation, a spatially and temporally adaptive weighting coefficient can be computed. This algorithm has been incorporated within the coupled streamflow prediction system, HydroDART. More details on the adaptive hybrid EnKF-OI scheme can be found in El Gharamti (2021).

3 Test Cases

3.1 Domains

Two extreme flooding events are selected for this study. One is a flash flood resulting from a thunderstorm. The other is a long lasting flooding event resulting from a hurricane landfall. Figures 2 and 3 show the location and extent of the two domains and they are explained in more detail below.

3.1.1 Florida's Flooding Case (2022)

Hurricane Ian became a tropical storm on September 24, 2022. Ian made landfall on western Cuba as a high-end category 3 hurricane on September 27. On September 28, Ian made landfall on southwestern Florida as a category 4 hurricane, producing catastrophic storm surge and historic freshwater flooding across much of central and northern Florida. Ian was responsible for more than 156 fatalities in the United States of which 66 were directly caused by the storm. Ian caused over $112 billion in damages; the costliest hurricane in Florida's history and the third costliest in United States history (National Hurricane Center Tropical Cyclone Report; Bucci et al., 2023).

Figure 2 shows the WRF-Hydro modeling domain for hurricane Ian. The colored map background depicts accumulated rainfall drawn for September 28 through 30, 2022, as modeled by the NWMv2.1 Analysis and Assimilation atmospheric forcing (described below) used to drive this case. The domain is a subset of the NWMv2.1 CONUS domain and includes 18,190 reaches and 151 lakes and reservoirs. Stream reaches are color coded based on their 10-year flood magnitudes as calculated from the 42 year NWMv2.1 respective model simulations described in more detail below. There are 171 USGS gauges with their drainage area fully contained in this domain and are assimilated and also used in performance assessment (dark red circles in Figure 2). Twenty-two of these gauges are GAGES-II reference gauges (Falcone, 2011) which have little or

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024


Hydrology and Earth System Sciences
Discussions

no anthropogenic alterations to their natural streamflows (green squares). Since WRF-Hydro does not have an active reservoir model and is performing a simple level pool routing, GAGES-II reference gauges are better suited for verification because they avoid accounting for heavy flow regulation at many reservoirs. Figure 2 also shows the location of four labeled verification gauges (black triangles) for which streamflow time-series will be provided in the results (section 4).

The full WRF-Hydro/NWMv2.1 model was run (without nudging data assimilation) for the FL's Ian test case using the NWMv2.1 analysis and assimilation cycle forcing dataset (https://water.noaa.gov/about/nwm). This meteorological forcing set is drawn from the Multi-Radar Multi-Sensor (MRMS) Gauge-adjusted and Radar-only observed precipitation products along with short-range Rapid Refresh (RAP) and High-Resolution Rapid Refresh (HRRR) . These atmospheric forcings drive the single model run depicted vertically on the left side of Fig. 1 and produce the output fluxes used by HydroDART to generate ensemble forcing for the channel+conceptual groundwater submodel. The simulation period is from August 15 to October 15, 2022. The model is initialized based on the operational NWMv2.1 model states on August 15, then the first month is used as a spin up period and streamflow assimilation begins on September 15, 2022.

3.1.2 West Virginia's Flooding Case (2016)

Several rounds of thunderstorms on June 23rd, 2016 produced torrential rainfall in West Virginia (WV) and western-central Virginia. Record rainfall accumulations of 200-250 mm were observed over a 24 hour period ending on 1200 UTC of June 24th (Martinaitis et al., 2020). This resulted in a rapid rise of water (in some places less than 6 hours) and extensive flooding across the domain. This was one of the deadliest flooding events in West Virginia's history with 23 fatalities. The event was classified as a billion-dollar disaster which damaged thousands of structures and over 1500 roads and bridges (Martinaitis et al., 2020). Many USGS gauges in the domain reported some level of flooding. Five of those gauges had their record flood stage measured in this event. The data assimilation and verification period for this test case is from June 20, 2016 through June 30, 2016; an 11 day period which encompasses the flash flooding event and the recession period completely at all verification gauges.

Figure 3 shows the WV flooding domain. Just like FL's Ian, the domain is a cutout of the NWMv2.1 CONUS domain including 47,046 reaches and 25 lakes and reservoirs. There are 121 USGS gauges with their drainage area fully contained in the modeling domain. These gauges are assimilated and used in performance assessment. Figure 3 shows the location of all those gauges as well as the subset of GAGES-II reference gauges. Figure 3 also shows the location of a few verification gauges for which streamflow time series are provided and analyzed in section 4.

Analysis of Record for Calibration (AORC, Fall et al., 2023) is used as the atmospheric forcing for the West Virginia test case. This is a new dataset developed to support NWM calibration and long term retrospective model simulations. AORC data has been cut out to the West Virginia modeling domain and used to force the WRF-Hydro model. The inflows to the channel and conceptual groundwater from this simulation are then used as forcing in all experiments (both data assimilation and forecasts). The simulation period is from June 1 through June 31, 2016. The model is initialized based on the NWMv2.1 retrospective model states on June 1, then the first 20 days are used as a spin up period and streamflow assimilation begins on June 20, 2016.

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions

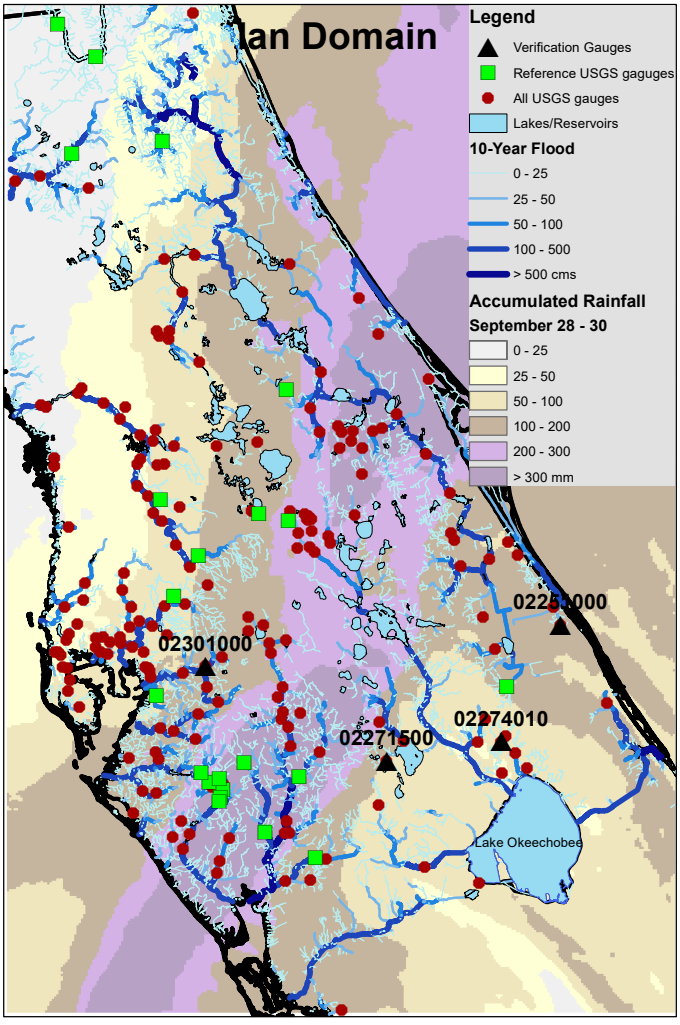


Figure 2. Study domain for Florida's flood case due to hurricane Ian. The map background depicts the accumulated rainfall drawn from NWMv2.1 Analysis and Assimilation atmospheric forcing during September 28 through 30, 2022. Stream reaches in the domain are color coded based on their 10-year flood magnitudes calculated from the 42-year NWMv2.1 retrospective model simulation. USGS streamflow observation gauges are shown in red circles. GAGES-II reference streamflow gauges are shown by green rectangles, and four verification gauges are denoted by black triangles each with their 15-digit gauge identifiers. Lakes and reservoir bodies are shown as blue polygons.

3.2 Observations

Hourly or sub-hourly streamflow observations, with frequencies as high as 5 minutes, are gathered from the USGS streamflow measurement sites. Our data collection involves accessing observations well after the events, and we do not rely on near-real-time data acquisition. Consequently, discharge estimates have been updated through subsequent quality controls, including

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions

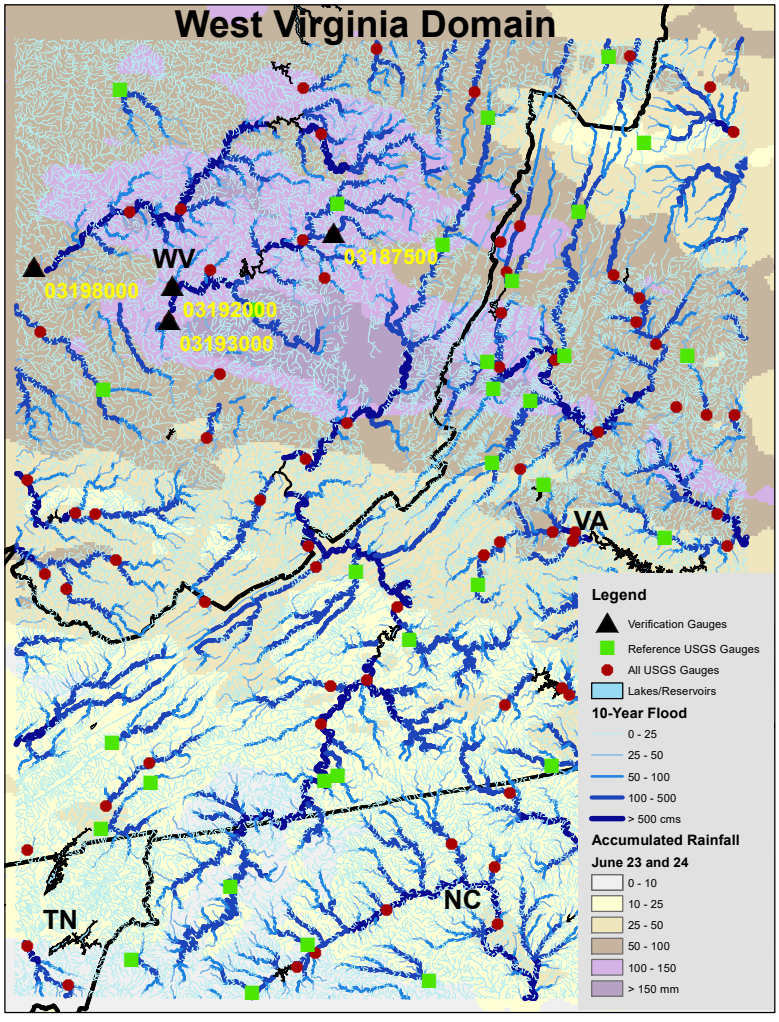


Figure 3. Study domain for West Virginia's flood case. The colored map in the background depicts accumulated rainfall drawn from Analysis of Record for Calibration (AORC) atmospheric forcing during June 23 and 24, 2016. Stream reaches in the domain are color coded based on the 10-year flood magnitudes, calculated based on the 42-year NWMv2.1 retrospective model simulation. USGS streamflow observation gauges are shown in red circles. GAGES-II references streamflow gauges are shown by green rectangles, and four verification gauges are denoted by black triangles each with their 15-digit gauge identifiers.

potential revisions to rating curves to better align with extreme flooding events and out-of-bank flows. For our West Virginia and Florida's Hurricane Ian test cases, streamflow observations were downloaded on 01/11/2022 and 01/30/2023, respectively. In the case of FL's Hurricane Ian, some gauges still lack event information in the observations, and the data remains provisional even months after the occurrence. The assimilation is conducted hourly in both domains, with sub-hourly USGS data averaged before the assimilation process begins.

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024


Hydrology and Earth System Sciences
Discussions

To model streamflow observation errors, we adopt a heteroscedastic Gaussian error model with zero mean and a standard deviation set to 25% of the observed flow, following Abbaszadeh et al. (2019). Given that the gauges are situated on the stream network, this results in a linear observation operator $\mathbf{H}$ which simplifies the Kalman update, excluding nonlinear issues from the analysis step of HydroDART. Additionally, we assume that observations and their associated errors are uncorrelated in both space and time, yielding a diagonal observation error covariance matrix, $\mathbf{R}$.

Within the HydroDART framework, observations falling beyond 3 total standard deviations (equivalent to the square root of the sum of the prior ensemble and observation error variance) from the prior ensemble mean are rejected. This threshold, determined by the outlier threshold parameter in DART, serves multiple purposes. It helps avoid assimilating inaccurate observations, and it prevents the inclusion of observations where the mean of the ensemble members is significantly distant from the observation value. Such assimilation could lead to unacceptably large increments, potentially destabilizing the model run.

3.3 Retrospective Model Simulations

The NOAA National Water Model version 2.1 offers a publicly accessible multi-decade retrospective simulation covering the Contiguous United States (CONUS). This simulation is based on the NWMv2.1 model code and static files, driven by AORC atmospheric forcings. Notably, the retrospective simulation lacks data assimilation within the hydrologic model. While serving as a valuable resource for historical modeling context and real-time operations, it also facilitates the assessment of flow frequencies and model verification over an extended period. The 42-year retrospective simulation for NWMv2.1, spanning from February 1979 to December 2020, is openly available in two formats: Network Common Data Form (NetCDF) and Zarr on Amazon Web Services (AWS). Interested users can access the data at the following link: https://registry.opendata.aws/nwm-archive/.

In our approach, we leverage the 42-year NWMv2.1 retrospective simulations to construct the static background covariance matrix, denoted as $\mathbf{B}$, for our hybrid ensemble-variational filter. To achieve this, we assemble a 1,000-member ensemble from the retrospective model simulation. While the climatological ensemble could potentially be larger given the extensive data volume in the retrospective simulation, we opt for 1,000 samples to manage storage and computational costs. Previous research has explored the climatological ensemble size in the context of hybrid ensemble-variational data assimilation, indicating that an ensemble on the order of hundreds to a few thousands is generally sufficient to match the mean correlation length scales of $\mathbf{B}$ (e.g., Lei et al., 2021). It's important to note that the climatological ensemble is not advanced forward in time.

The climatological ensemble is subjected to an offline Empirical Orthogonal Function (EOF) analysis. The purpose of this analysis is to determine the spatial patterns and correlations present in the ensemble. This exploration aids in understanding the variability captured by the climatology and its potential impact on the hybrid ensemble-variational filter. The EOF analysis reveals that employing 1,000 static members results in a covariance matrix free from noise, where large-scale patterns dominate the initial EOFs, and smaller-scale patterns are encapsulated in the latter EOFs. This information is instrumental in shaping our understanding of the climatological ensemble's composition and behavior.

In terms of implementation, for Hurricane Ian, the process involves extracting two state variables, streamflow and water depth, from the dataset encompassing all 42 years for all reaches and buckets in the domain. A temporal subset is then created,

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions

focusing exclusively on the month of September, given Ian's occurrence during that period. Every model simulation within the month of September from these 42-year model simulations is considered a member of the climatology ensemble. Subsequently, 1,000 members are randomly selected from this dataset and preserved for use by HydroDART, as outlined in equation (6). For West Virginia, a parallel approach is adopted, with the exception that the model simulations from the month of June are utilized in constructing the static/climatology members. It's important to note that, for the purpose of constructing the climatology, we exclude the year 2016 in West Virginia to ensure that the flooding scenario under observation is not part of the climatology. In the case of Hurricane Ian, the year 2022 is not part of the retrospective run, eliminating the need for its exclusion.

4 Experiments

To test the performance of the hybrid scheme against the EnKF, we perform different assimilation runs based on the choice of the hybrid weighting coefficient and the ensemble size. For all DA runs, we set the ATS localization cutoff distance to 100 km and turn on adaptive prior and posterior inflation following El Gharamti et al. (2021). To assess the quality of the estimated streamflow, we use many ensemble and hydrological (deterministic) metrics.

$$\text{RMSE} = \frac{1}{N_t}\sum_{k=1}^{N_t}\sqrt{\frac{1}{N_e}\sum_{i=1}^{N_e}\left(Q_k^{f|a,i}-Q_k^o\right)^2}, \tag{12}$$

$$\text{ES} = \frac{1}{N_t}\sum_{k=1}^{N_t}\sqrt{\frac{1}{N_e}\sum_{i=1}^{N_e}\left(Q_k^{f|a,i}-\overline{Q}_k^{f|a}\right)^2}, \qquad \overline{Q}_k^{f|a}=\frac{1}{N_e}\sum_{i=1}^{N_e}Q_k^{f|a,i} \tag{13}$$

$$\text{Bias} = \frac{1}{N_t N_e}\sum_{k=1}^{N_t}\sum_{i=1}^{N_e}\left|Q_k^{f|a,i}-Q_k^o\right|, \tag{14}$$

$$\text{CRPSS} = 1-\frac{\text{CRPS}^f}{\text{CRPS}^{\text{OL}}}, \qquad \text{CRPS}=\int_{-\infty}^{+\infty}\left(F(Q)-H\left(Q\geq Q^o\right)\right)^2 dQ, \tag{15}$$

$$\text{NSE} = 1-\frac{\sum_{k=1}^{N_t}\left(\overline{Q}_k^f-Q_k^o\right)^2}{\sum_{k=1}^{N_t}\left(Q_k^o-\widehat{Q}^o\right)^2}, \qquad \widehat{Q}^o=\frac{1}{N_t}\sum_{k=1}^{N_t}Q_k^o \tag{16}$$

$$\text{KGE} = 1-\sqrt{(r-1)^2+(\xi-1)^2+(\delta-1)^2}, \qquad \xi=\frac{\sum_{k=1}^{N_t}\overline{Q}_k^f}{\sum_{k=1}^{N_t}Q_k^o}, \qquad \delta=\frac{\sqrt{\sum_{k=1}^{N_t}\left(\overline{Q}_k^f-N_t^{-1}\sum_{k=1}^{N_t}\overline{Q}_k^f\right)^2}}{\sqrt{\sum_{k=1}^{N_t}\left(Q_k^o-\widehat{Q}^o\right)^2}}, \tag{17}$$

$$\text{CV} = \frac{\sqrt{N_t\sum_{k=1}^{N_t}\left(Q_k^o-\widehat{Q}^o\right)^2}}{\sum_{k=1}^{N_t}Q_k^o}, \tag{18}$$

$$\text{C-RMSE} = \sqrt{\frac{1}{N_t}\sum_{k=1}^{N_t}\left[\left(\overline{Q}_k^f-\frac{1}{N_t}\sum_{k=1}^{N_t}\overline{Q}_k^f\right)-\left(Q_k^o-\widehat{Q}^o\right)\right]^2}, \tag{19}$$

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024


Hydrology and Earth System Sciences
Discussions

where RMSE and ES denote the time-averaged root mean squared error and ensemble spread, respectively. Q refers to discharge or streamflow, measured in cubic meters per second (cms). The superscript $f|a$ means either prior or posterior discharge and o denotes the observed discharge. N_t is the total time steps or DA cycles. The $\bar{\cdot}$ and $\hat{\cdot}$ notations refer to averaging over the ensemble and time, respectively. Accordingly, streamflow ensemble mean at time t_k is given by $\overline{Q}_k$ and the time-averaged observed flow is $\widehat{Q}^o$. OL is a term used to describe the open loop or the unconstrained ensemble model run (without assimilation). The continuous ranked probability score, CRPS (CRPS, Matheson and Winkler, 1976), is calculated by comparing the simulated (forecast, analysis, or open loop) cumulative distribution function (F) against the observations over a given period. H is the Heaviside function which is 0 for predicted values smaller than the observed streamflow and 1 otherwise. To compute the added skill by DA over the OL results, we use the continuous ranked probability skill score, CRPSS (CRPSS, Hersbach, 2000). A CRPSS of zero means that DA didn't improve the prediction skill of the model. As CRPSS moves closer to 1, this indicates that DA improves streamflow skill compared to the OL. Negative CRPSS values indicate that DA yields worse predictions than the OL. For deterministic evaluation of hydrographs, we compute the Nash-Sutcliffe Efficiency (NSE, Nash and Sutcliffe, 1970), the Kling-Gupta Efficiency (KGE, Gupta et al., 2009) and the coefficient of variation of the observed flow, CV. In equation (17), r denotes the Pearson correlation coefficient between the estimated flow and the observed one. ξ is a bias term and δ is a measure of flow variability. Generally, high quality streamflow estimates are obtained when NSE and KGE are close to their optimal values of one. C-RMSE is the centered root mean squared error used to aggregate estimates from different gauges in a single metric.

4.1 EnKF Runs

In this section, we execute an EnKF run within the HydroDART framework, employing an ensemble of 80 members. This approach aligns with the experiments outlined in the prior HydroDART study which focused on hurricane Florence in North Carolina (El Gharamti et al., 2021). The objective is twofold: to assess the prediction system's performance in distinct basins characterized by diverse modeling and precipitation complexities, and to establish a baseline performance, both qualitatively and quantitatively, for the EnKF. This baseline serves as a reference point from which we intend to enhance predictive capabilities through the implementation of the hybrid approach.

Figure 4 illustrates the estimated streamflow for Sebastian River in FL and Kanawha River in WV. The hydrographs depict the evolution of observed and OL discharge, overlaid with the prior and posterior estimates. The South Prong St. Sebastian River, located on the eastern side of FL, witnessed a rapid surge in streamflow on September 28, 2022, coinciding with the landfall of Hurricane Ian. This flooding persisted for about a week before returning to normal flow conditions. In comparison to the OL, both the prior and posterior ensemble estimates exhibit significantly improved accuracy, offering a better representation of observed flows, particularly during the flooding event. On average, the prior and posterior streamflow estimates are 45% and 51% more accurate, respectively (measured in terms of RMSE), compared to the OL. Additionally, the NSE and KGE values, derived from the EnKF ensemble mean, are notably higher and closer to 1 than those associated with the OL. This underscores the substantial improvement in prediction achieved through the assimilation of streamflow data with the EnKF.

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions

Adaptive inflation plays a crucial role in this success, dynamically growing during the flooding event to enhance ensemble spread and refine its conformity to observed data.

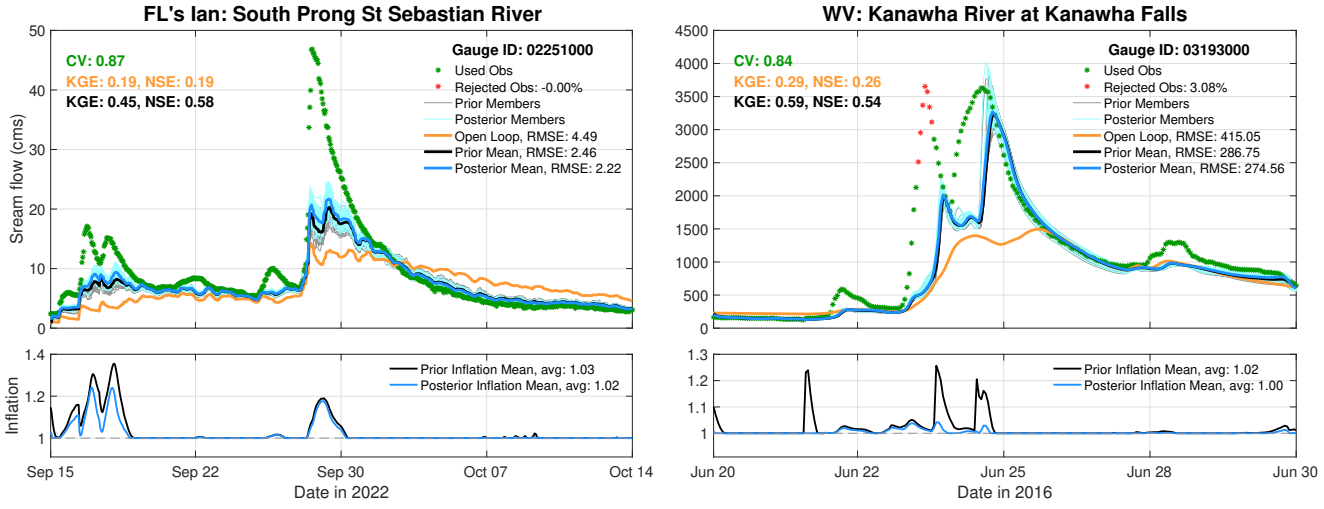


Figure 4. Left: Streamflow time-series (hydrograph) during hurricane Ian at South Prong Street Sebastian River. Observed streamflow is given by hourly asterisks. Green asterisks indicate that the observation was assimilated. The rejected observations due to outlier threshold are shown in red asterisks. OL streamflow estimates are given by the orange curve. Prior and Posterior ensemble means are denoted by the thick black and blue curves, respectively. The light gray and cyan curves show the prior and posterior ensemble members, respectively. Time-averaged streamflow RMSE values are reported in the legend. Other metrics such as CV, NSE and KGE are also annotated. The evolution of prior and posterior inflation over time is shown in the bottom panel. The right panels are similar to the left ones but for the Kanawha River at Kanawha Falls in WV.

Kanawha River in WV underwent severe flooding in our case study, with discharge values soaring to around 4000 cms. Similar to FL's Ian case, the EnKF demonstrated superior alignment with streamflow observations compared to the OL. Notably, the EnKF estimates displayed a precise correspondence with the falling limb of the hydrograph. While excelling on the rising limb, the EnKF fell short of assimilating all data due to a pronounced underestimation of streamflow. In this instance, the EnKF's prior ensemble yielded an approximately 31% lower RMSE than the OL.

Figure 5 presents comprehensive summary diagnostics for both the Ian and WV flooding cases, utilizing data from both 'all' and 'reference' USGS gauges within the domains. The 'reference' designation pertains to the GAGES-II reference gauges. Clearly, estimates derived from the OL and the EnKF at the reference gauges exhibit enhanced accuracy compared to those from all gauges. This improvement is attributed to the fact that reference gauges are subject to little to no regulation and constitute only a small subset of all gauges, resulting in fewer extreme outliers.

The CRPSS boxplots highlight the added value of the DA system over the OL, evident in positive scores. The majority of gauges achieve scores surpassing 0.5. The NSE and KGE scores from the EnKF notably outperform those of the OL. For the OL run, numerous gauges yielded efficiency estimates less than -10 (omitted for figure clarity). The summary boxplots for the

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

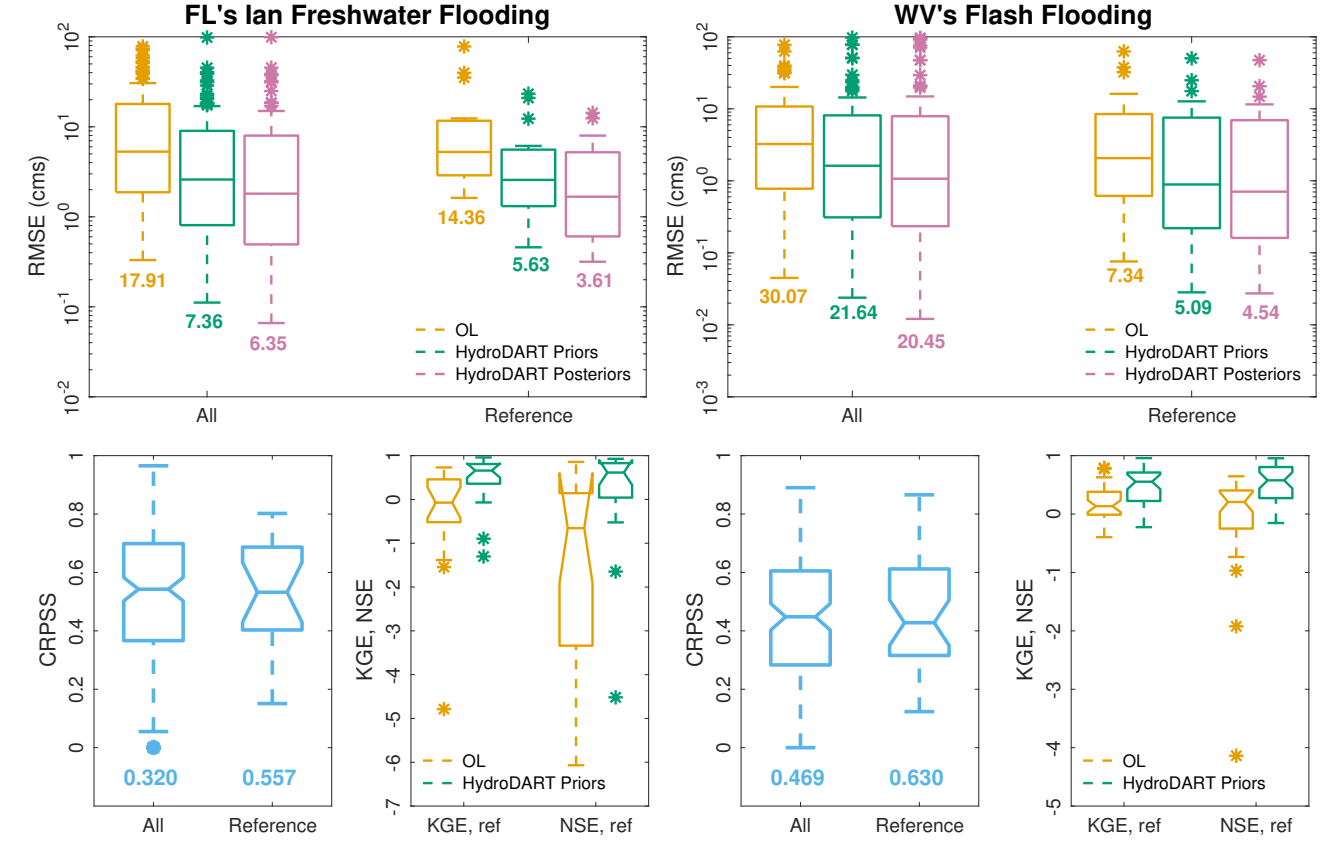


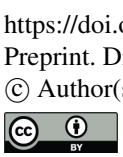
Figure 5. Top left: RMSE boxplots for OL and EnKF prior and posterior estimates for all gauges and for GAGES-II reference USGS gauges in the FL domain. Note that the y-axis is in log-scale. Averaged RMSE are reported underneath the individual boxplots. Bottom Left: CRPSS boxplots (blue) for all gauges and for GAGES-II reference gauges. NSE (green) and KGE (yellow) boxplots are also shown for the reference gauges. The right panels are similar to the left ones but for the WV flash flood case.

RMSE demonstrate the superior assimilation performance of the EnKF in both flooding cases. When averaging over all gauges and across time, EnKF predictions exhibit at least 50% and 30% greater skill than the OL for the Ian and WV flooding cases, respectively. In the subsequent sections, we will delve into the hybrid results to explore how they build upon the achievements of the EnKF presented in this section.

4.2 Hybrid Runs

The hybrid approach integrates static climatological background covariance statistics into the dynamic ensemble. In this section, our focus is on comprehending the performance of the hybrid EnKF-OI scheme with respect to α, where the background covariance is a blend of time-varying dynamic statistics and climatology. We examine five cases where we fix the hybrid weighting coefficient both in space and time, setting α to values of 0.1, 0.3, 0.5, 0.7, and 0.9. It's crucial to note that, from

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024



Hydrology and Earth System Sciences
Discussions

eq. (6), α represents the weight on the dynamic background error covariances, while $(1-\alpha)$ signifies the weight on the static background. Therefore, the chosen values of α progressively emphasize the dynamic background more in order. Experiments involving adaptive hybrid weighting are discussed further in section 4.4.

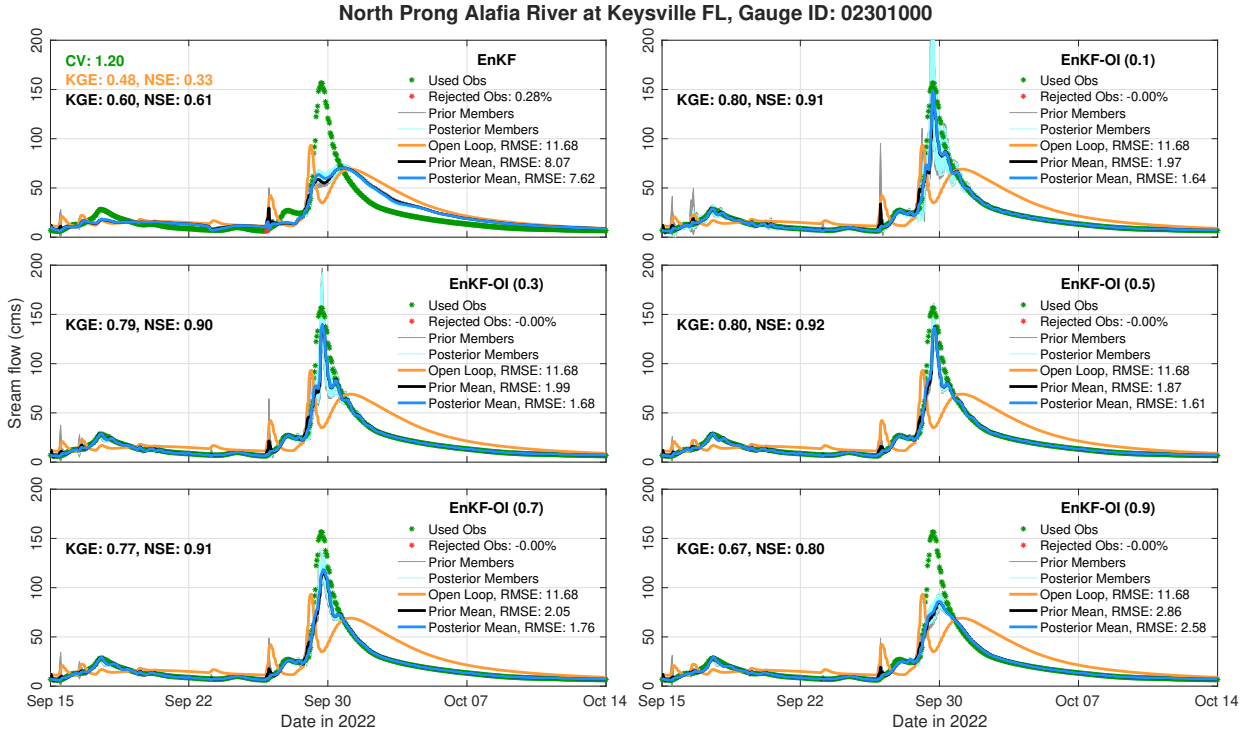


Figure 6. Hydrographs for North Prong Alafia River in FL using the EnKF and the 5 hybrid EnKF-OI runs. The hybrid runs use different weighting coefficients α reported in the title of the individual legends. Similar to Fig. 4, CV, NSE, KGE and RMSE values are reported on all panels.

Figure 6 displays hydrographs for the North Prong Alafia River at Keysville, FL. Situated just outside Tampa on the western corridor of the state, this gauge provides valuable insights. All EnKF-OI runs exhibit an improved fit to the hydrograph peak on September 29. Both the standard EnKF and the OL underestimate the flooding event, predicting a very delayed recession. Beyond the primary flooding event, the hybrid runs precisely capture the rising and falling limbs of the hydrographs. The EnKF-OI with $\alpha = 0.5$ stands out as the most accurate, yielding a KGE value of 0.92. However, as α approaches 1, the performance starts to degrade due to the heavier weight placed on the dynamic sample covariance. Conversely, when α is close to 0, time-dependent information in the background covariance is limited. For instance, in the case of $\alpha = 0.1$, both the prior and posterior ensemble spread become quite large. This unusual increase in ensemble uncertainty is a result of the dominance of the climatological ensemble over the dynamic one.

Figure 7 provides a comparison between the hybrid EnKF-OI scheme and the standard EnKF at the Kanawha River gauge near Charleston, WV. This area was significantly affected by the flooding event under consideration. On June 23rd, the river's

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

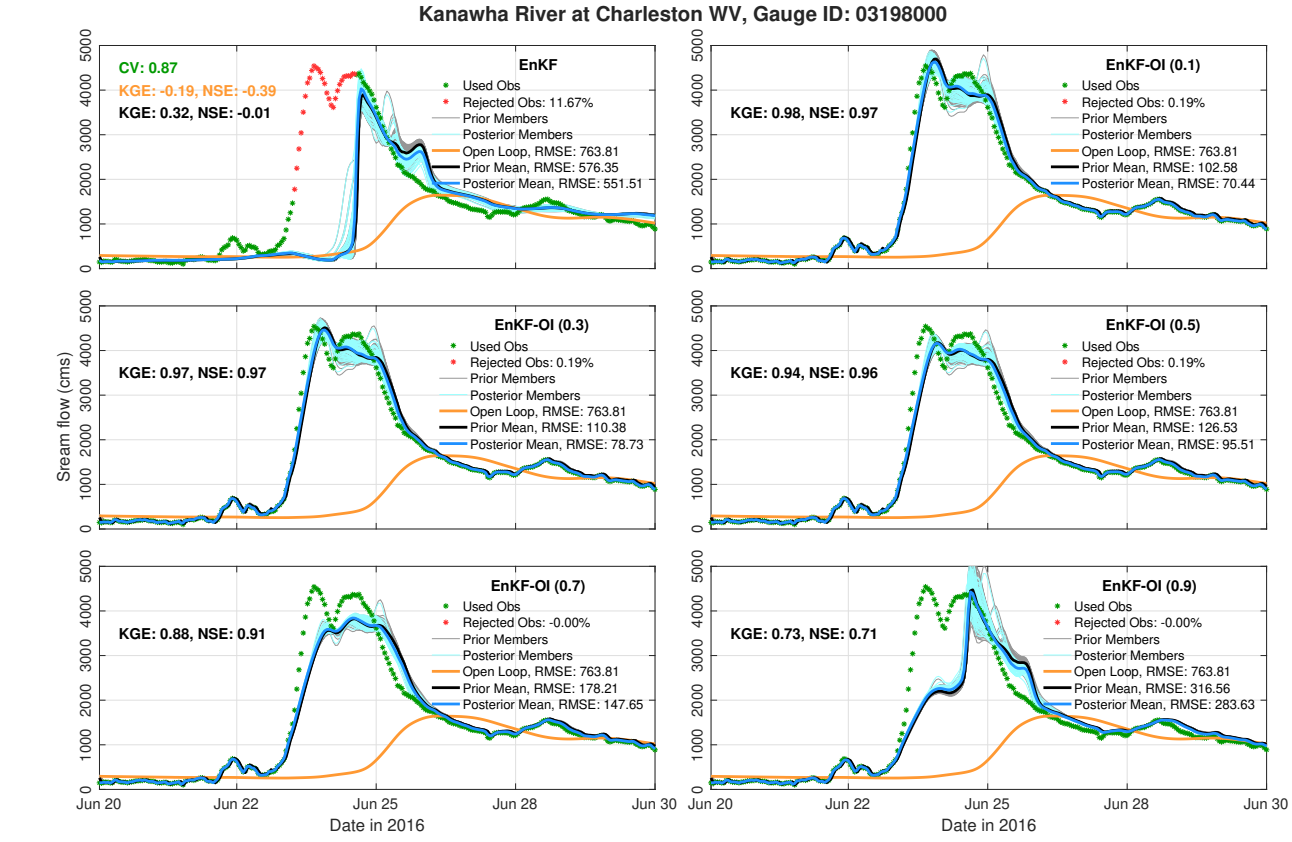


Figure 7. Similar to Fig. 6 but for Kanawha River at Charleston in WV.

discharge rose to almost 5000 cms. The OL, representing the hydrologic model, fails to capture the flooding event entirely. While the EnKF struggles to accurately predict the initial day, it improves toward the 24th by simulating the end of the flood peak and its recession. However, given the intensity of the flood peak, the EnKF estimates, while surpassing the OL, are not satisfactory (NSE ≈ 0). In contrast, the hybrid EnKF-OI solution, particularly with $\alpha \leq 0.5$, more accurately simulates the observed streamflow, resulting in an NSE exceeding 0.95. Increasing the hybrid weight beyond 0.5 diminishes the EnKF-OI's skill, although it still outperforms the EnKF significantly. Remarkably, such outstanding performance can be achieved by incorporating only 10% of the hybrid covariance from the climatology. Since the climatological covariance is not susceptible to sampling errors, its partial integration into the EnKF helps mitigate significant model biases, especially during flooding events.

4.2.1 Ensemble Spread and Inflation

The hybrid covariance approach provides the dynamic ensemble of the EnKF with more diverse correction directions allowing for a better analysis. A crucial aspect of this hybridization is the augmentation of ensemble variability, a factor anticipated to improve performance in the presence of model uncertainties and forcing biases. To illustrate this, Figure 8 depicts the Average

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

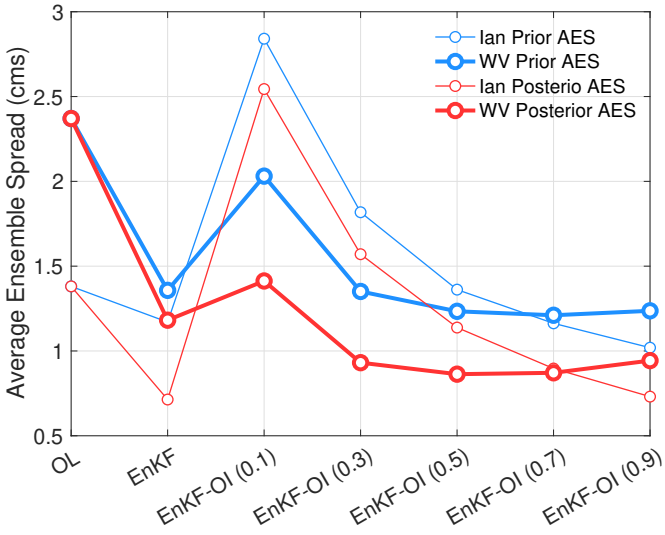


Figure 8. Average Ensemble Spread (AES) across all gauges for the OL, EnKF and the hybrid EnKF-OI with 5 distinct weights. Results are shown for FL (thin) and WV (thick) for the prior and the posterior ensembles.

Ensemble Spread (AES) across all gauges for both study domains. Given the unconstrained nature of the OL, it is reasonable for its AES to be larger than that of the EnKF. The EnKF, due to its hourly analyses, experiences substantial shrinkage of the ensemble spread. This becomes problematic when the shrinkage occurs far from the observations, leading to poor streamflow estimates, especially in the presence of model biases. While inflation is often employed to address this issue, it may not always effectively counter biases. The introduction of climatological information with $\alpha = 0.1$ results in a notable increase in the ensemble spread. This is because the climatological covariance $\mathbf{B}$ comprises a substantial inventory (1,000 instances over 42 years) of historical streamflow distributions, contributing to its relatively large variance. As α increases, the ensemble spread decreases until it aligns with the EnKF spread for $\alpha = 0.9$.

Considering the enlarged ensemble spread resulting from the hybrid approach, the question arises: Is inflation still necessary? Encouragingly, HydroDART incorporates adaptive inflation in both space and time. This adaptive inflation procedure considers the current bias (or innovation), ensemble spread, and observational uncertainty to update inflation over time. Figure 9 illustrates the time evolution of prior inflation for the same gauges examined in Figs. 6 and 7. At North Prong Alafia River in FL, the hybrid EnKF-OI runs utilize inflation solely during flooding events, turning it off completely before and after the rainfall events. Early inflation spikes around Sep. 16 and 17 may address modeling bias and limited initial variability. For Kanawha River in WV, inflation is predominantly unused by the hybrid runs, except for a brief initial period on June 20th. In contrast, the EnKF employs inflation not only during flooding periods but also in non-flooding intervals. Averaging across all gauges, the hybrid EnKF-OI with $\alpha = 0.5$ requires approximately 5 to 10% less inflation than the EnKF, a trend consistent with posterior inflation. Consequently, we opted to retain the inflation algorithm for the hybrid runs. In summary, we posit

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions

that inflation primarily addresses instantaneous bias, while the inclusion of static background information serves to alleviate long-term biases and uncertainties.

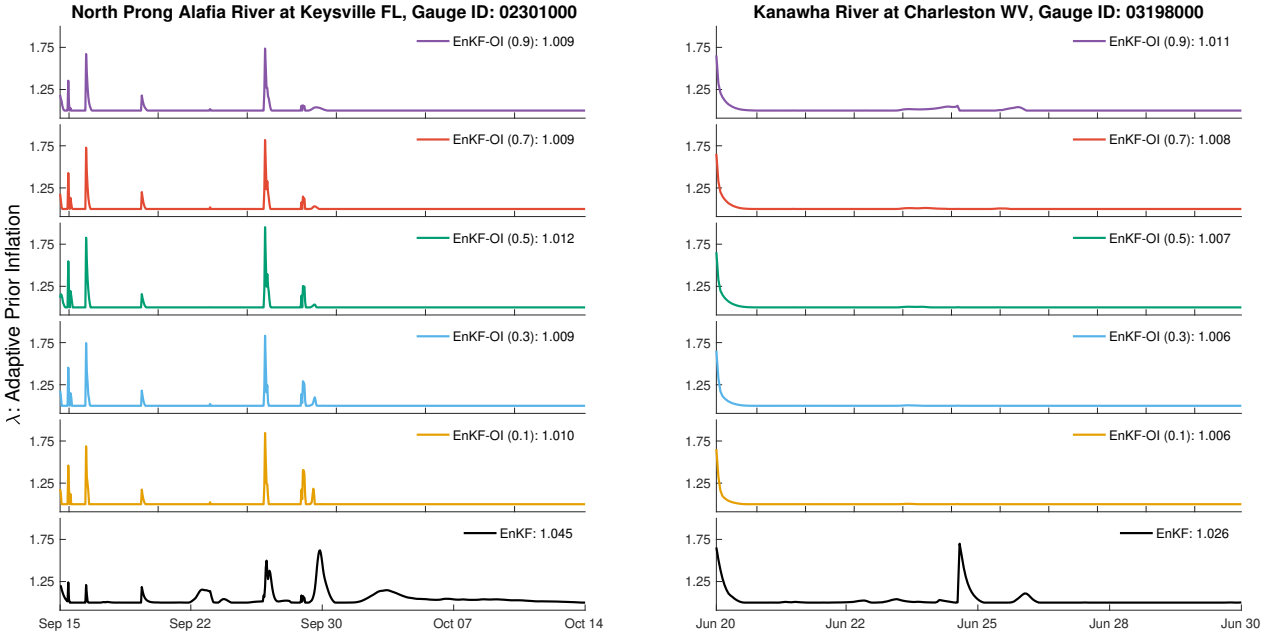


Figure 9. Time-series of prior inflation resulting from the EnKF and the hybrid EnKF-OI with constant weights. The inflation, denoted by λ, in equation (4) is adaptive in space and time. North Prong Alafia River in FL is represented by the left panels. The right panels show the results for Kanawha River at Charleston in WV.

4.2.2 Optimal Hybrid Weight

To synthesize the evaluations of the hybrid EnKF-OI with constant weights, Taylor diagrams (Taylor, 2001) and RMSE boxplots in Figure 10 offer a comprehensive view for all gauges within the FL and WV domains. A Taylor diagram simultaneously encapsulates three metrics-standard deviation, Pearson correlation, and centered RMSE from Eq. (19)-in a single representation, providing a nuanced understanding of the correspondence between observed and estimated flows. Given the multitude of gauges in each domain, the Taylor diagram exhibits a cloud of points, one for each streamflow gauge. Optimal performance is characterized by a correlation of 1, with both C-RMSE and standard deviation equal to 0. Visually, the cloud of points corresponding to the hybrid runs tends to cluster closer to the ideal performance point than the cloud for the EnKF. For instance, in the WV case, the EnKF cloud is perceptibly closer to the left side of the diagram, indicating lower accuracy compared to other schemes. Aggregating results across all gauges, the resulting correlations for each scheme are tabulated in Table 1. In both domains, correlations from the hybrid runs surpass those of the EnKF. The EnKF-OI schemes yield comparable correlations, with $\alpha = 0.5$ consistently offering the best performance in both domains. The boxplots in Figure 10 echo a similar narrative-the hybrid EnKF-OI enhances prior RMSE by over 50% compared to the EnKF. Among the tested hybrid runs, $\alpha = 0.1$ contributes

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024


Hydrology and Earth System Sciences
Discussions

to the least accurate results, underscoring the significance of maintaining a dynamic ensemble in the filtering framework for skillful streamflow estimation. The findings prompt a critical question: How large of an ensemble is necessary to ensure precise and efficient predictions? We will delve into answering this question in the following section.

Table 1. Pearson correlation between the observed discharge and the prior estimates suggested by the EnKF and the hybrid EnKF-OI with constant weights. The correlation is computed for all gauges in each domain and averaged over time.

Domain	EnKF	EnKF-OI (0.1)	EnKF-OI (0.3)	EnKF-OI (0.5)	EnKF-OI (0.7)	EnKF-OI (0.9)
FL	0.62501	0.72253	0.75732	0.75777	0.76555	0.75646
WV	0.66641	0.85900	0.85783	0.86092	0.85192	0.85037

4.3 Dynamic Ensemble Size

In this section, we conduct sensitivity experiments concerning the ensemble size using the hybrid scheme, testing five different sizes $N_e = (10, 20, 40, 60, 80)$ while maintaining a fixed hybrid weight of 0.5. It's important to note that all previous experiments utilized a dynamic ensemble size of 80. The reduction in ensemble size is aimed at enhancing the computational efficiency of the hybrid scheme while retaining a satisfactory prediction skill. Figure 11 illustrates the distribution of the absolute prior bias for all USGS gauges and reference gauges separately. Additionally, we compare the results with our baseline 80-member EnKF runs (without hybrid) in both domains.

The observed prior bias for the WV flash flooding is larger than that for FL flooding, attributed to the shorter duration of the WV event, making it inherently flashier. In comparison to the EnKF, the hybrid runs consistently exhibit significantly improved prior bias, evident for both all and reference gauges. For instance, using only 10 members results in an averaged prior bias of 4.79 cms and 12.41 cms for FL and WV cases, respectively. This represents an average reduction of 29% (FL) and 40% (WV) in prior bias compared to the 80-member EnKF runs.

Among the hybrid runs, a dynamic ensemble of 80 members achieves the best performance in both cases. However, the difference in prediction accuracy between the EnKF-OI with 80 and 20 members is minimal ($\leq 6\%$). Notably, the computational cost of running the EnKF-OI with 20 members is approximately four times smaller than the run with 80 members. While reducing the ensemble size seems to slightly compromise prediction skill in our case studies, the trade-off of sacrificing 6% in accuracy, particularly evident in the 20-member EnKF-OI, is considered favorable for improving computational efficiency in operational settings. This conclusion is consistent across various metrics, including CRPSS, NSE, KGE, and RMSE.

4.4 Adaptive Hybrid Scheme

Building upon the ensemble analysis discussed in the preceding section, we opted for an ensemble size of 20 in our adaptive hybrid runs. The initial distribution for the hybrid weight (α) was strategically chosen as a Gaussian random variable centered at 0.5, with a standard deviation of 0.005. This decision stems from insights gained in section 4.2, where our analysis revealed

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

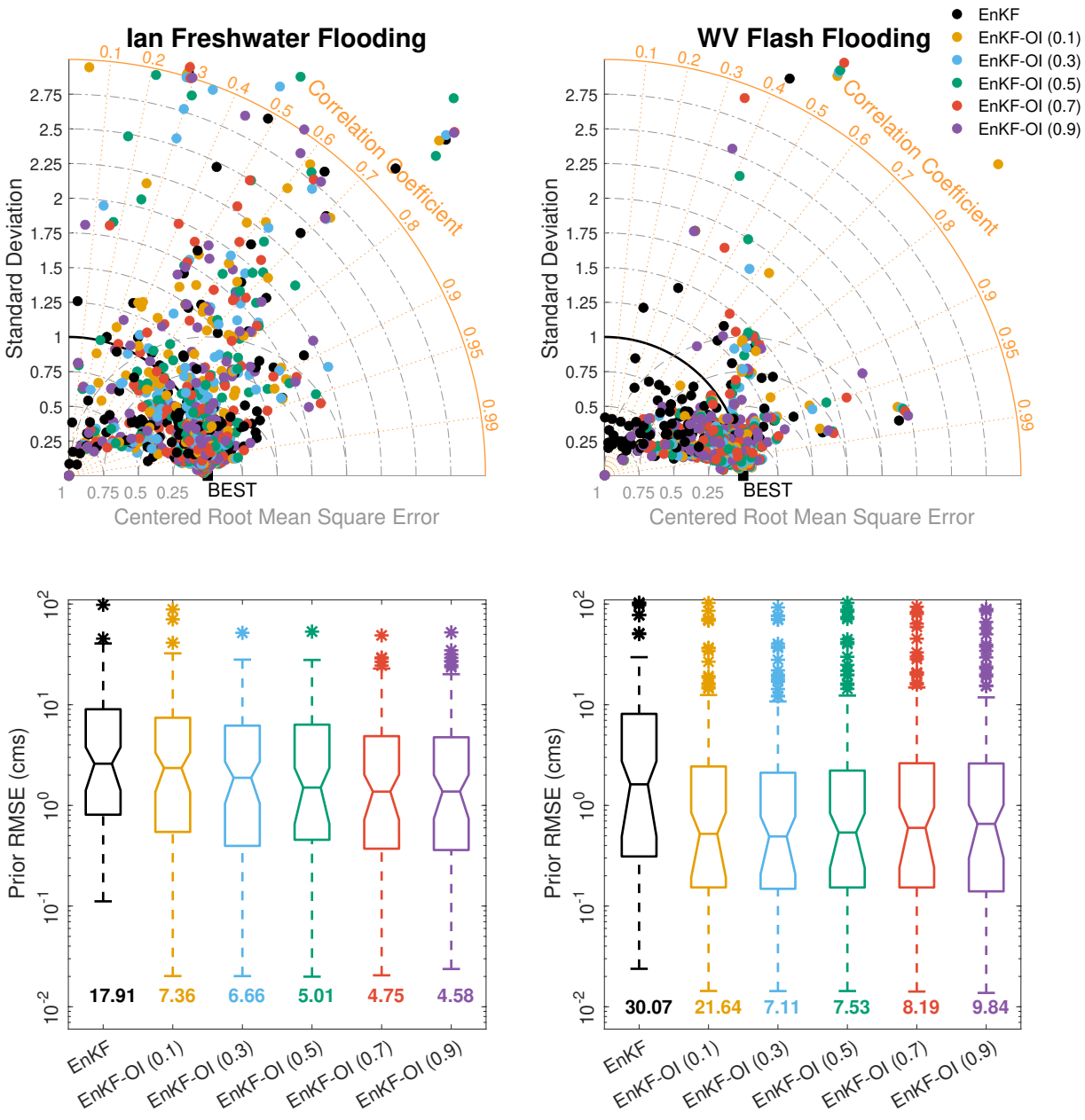


Figure 10. Top panels: Taylor diagrams for Ian's flood (left) and WV's flash flood (right). Each cloud of points on the diagram denotes all gauges available in the domain. "BEST" is the point of ideal performance, having perfect correlation with the observed flow and zero error and standard deviation. Gauges with standard deviations larger than 3 are omitted for visual purposes. Bottom panels: Prior RMSE boxplots for all gauges in each domain, obtained using the EnKF and the hybrid scheme with fixed weights. Overall averaged prior RMSE in each case is reported beneath the individual boxplots.

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions

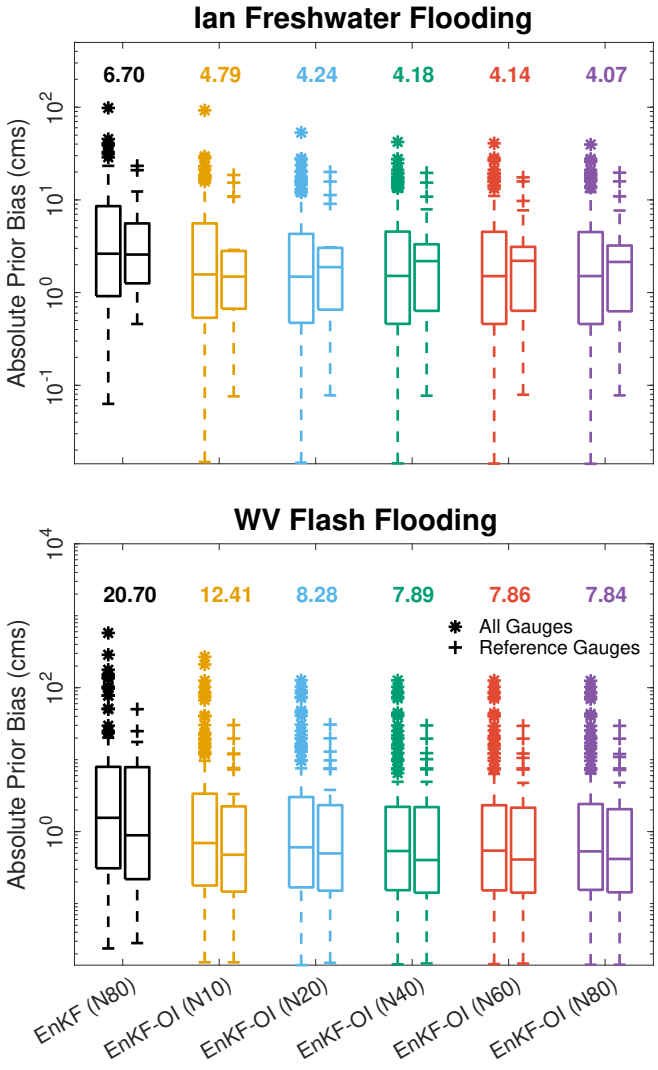


Figure 11. Prior absolute bias boxplots for all/reference gauges at Ian (top) and WV (bottom). The boxplots are obtained using the EnKF (with 80 members) and the hybrid scheme with different dynamic ensemble sizes. Time-averaged prior absolute bias for all gauges in each case is reported above the individual boxplots. Note that the y-axis is in log-scale.

that a weight of 0.5 produced the optimal representation of prior streamflow in both domains. Given the adaptive nature of the algorithm, initiating DA with an equal weighting of dynamic and static covariances at 0.5 was deemed an intuitive starting point. In terms of the variance, we conducted several sensitivity experiments to assess the impact of the initial standard deviation of α on the accuracy and performance of the hybrid filter. Our findings indicated that the standard deviation primarily influences the speed at which the weight gets updated. Consequently, a standard deviation value of 0.005 was selected, as it yielded the most favorable overall behavior in our experiments.

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

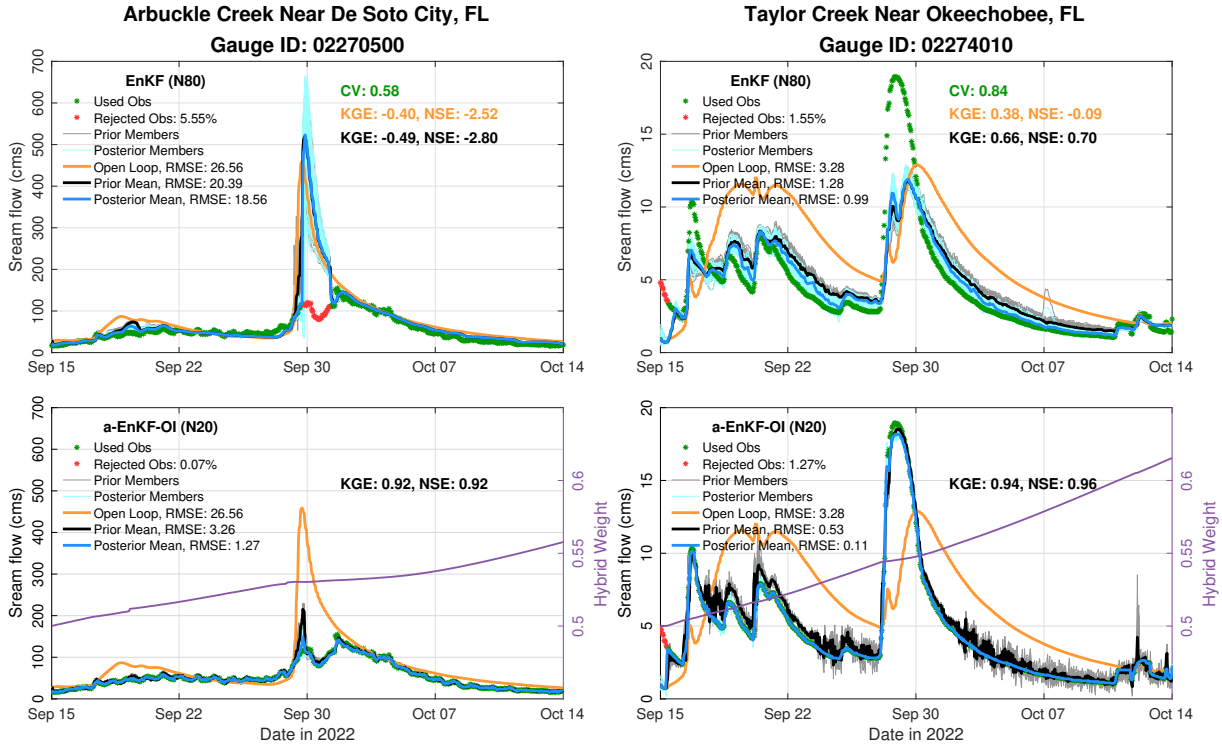
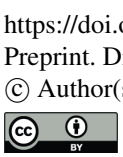


Figure 12. Hydrographs for Arbuckle Creek near De Soto City, FL (left) and Taylor Creek near Okeeechobee, FL (right). Top panels show the results from an 80-member EnKF together with the OL. The results from the a-EnKF-OI with 20 members are given in the bottom panels. For the a-EnKF-OI, the change in the hybrid weight over time is displayed in purple according to the right y-axis. Time-averaged metrics such as RMSE, NSE and KGE are annotated on the individual panels.

Figure 12 displays hydrographs at Arbuckle and Taylor Creeks in FL, offering a comparative assessment of prediction performance between the EnKF with 80 members and the adaptive hybrid EnKF-OI (a-EnKF-OI) with a reduced ensemble size of 20 members. At Arbuckle Creek, the OL tends to overestimate the flooding event in late September by nearly 400 cms. While the EnKF excels before the main event, it mirrors the trajectory of the OL during the flood, rejecting 2 days of data. In stark contrast, the a-EnKF-OI assimilates all observations, achieving a near-perfect fit to the observed discharge. Notably, the a-EnKF-OI attains high KGE and NSE values of 0.92, while the EnKF lags with scores of -0.49 and -2.80, respectively. The adaptive hybrid weight steadily increases from 0.5 to 0.55 during the assimilation period. Similarly, at Taylor Creek, the a-EnKF-OI performs exceptionally well, yielding KGE and NSE values of 0.94 and 0.96, respectively. Most improvements over the EnKF are evident during the last 2 days of September, coinciding with the main flood peak. The lower discharge at Taylor Creek underscores the adaptive hybrid filter's consistency in varying hydrological conditions. The hybrid weight demonstrates an early increase, followed by a sharper rise on September 28th, reaching approximately 0.62 by the end of the simulation. The adaptive scheme's increased weighting on the dynamic ensemble, as reflected by the rising α, signifies reduced bias and improved ensemble statistics. This shift renders climatological information less crucial, indicating the hybrid scheme's

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024



Hydrology and Earth System Sciences
Discussions

adeptness at leveraging climatology to enhance ensemble bias and subsequently placing greater emphasis on the dynamic ensemble. In terms of computational efficiency, the 20-member a-EnKF-OI proves to be roughly 4 times more efficient than the 80-member EnKF, further highlighting the advantages of the adaptive hybrid filter.

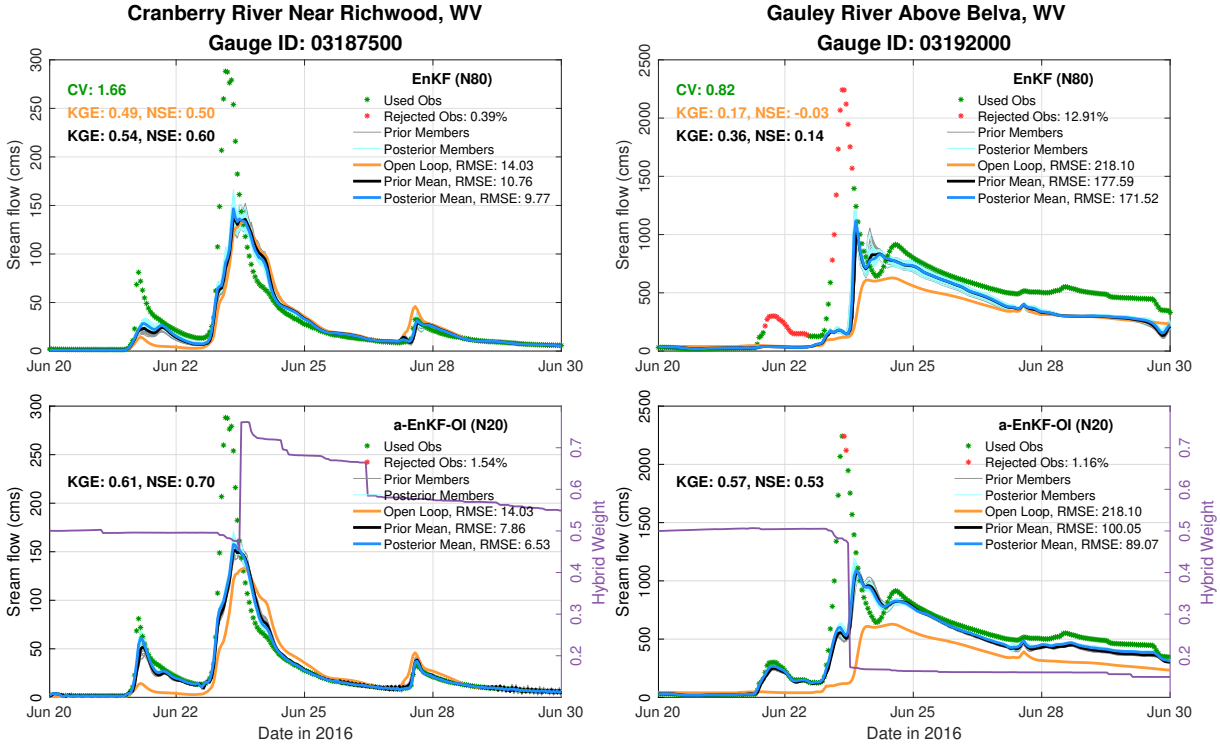


Figure 13. Similar to Fig. 12 but for Cranberry River near Richwood and Gauley River above Belva in WV.

The effectiveness of the 20-member a-EnKF-OI algorithm is further explored through its application to the flash flood events at Cranberry and Gauley Rivers in WV, as illustrated in Figure 13. At Cranberry River, both the EnKF and a-EnKF-OI schemes exhibit challenges in accurately capturing the observed discharge during the main flood peak. However, the a-EnKF-OI offers an improved prediction of the earlier event on June 21st, 2016. Notably, the rising and falling limbs of the main event are more distinctly delineated using the hybrid filter. Both filtering algorithms demonstrate greater skill compared to the OL. During the recession period, the hybrid weight is elevated to 0.75 and subsequently decreases back to 0.55 on June 30th. At Gauley River, the a-EnKF-OI notably outperforms the EnKF, particularly during the early hours of the flooding event and the subsequent recession period, resulting in an improved overall NSE of 0.53 compared to the EnKF's score of 0.14. The hybrid weight at this gauge undergoes a significant drop from 0.5 to almost 0.1 during the heavy rainfall event. This adjustment seems to be linked to the pronounced underestimation of streamflow. The adaptive scheme strategically places more weight on the climatology in an effort to alleviate the observed discrepancy between priors and observations. It's noteworthy that for both

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

gauges, the algorithm dynamically adjusts the hybrid weight, particularly around June 23rd, showcasing its responsiveness to underlying prior ensemble biases.

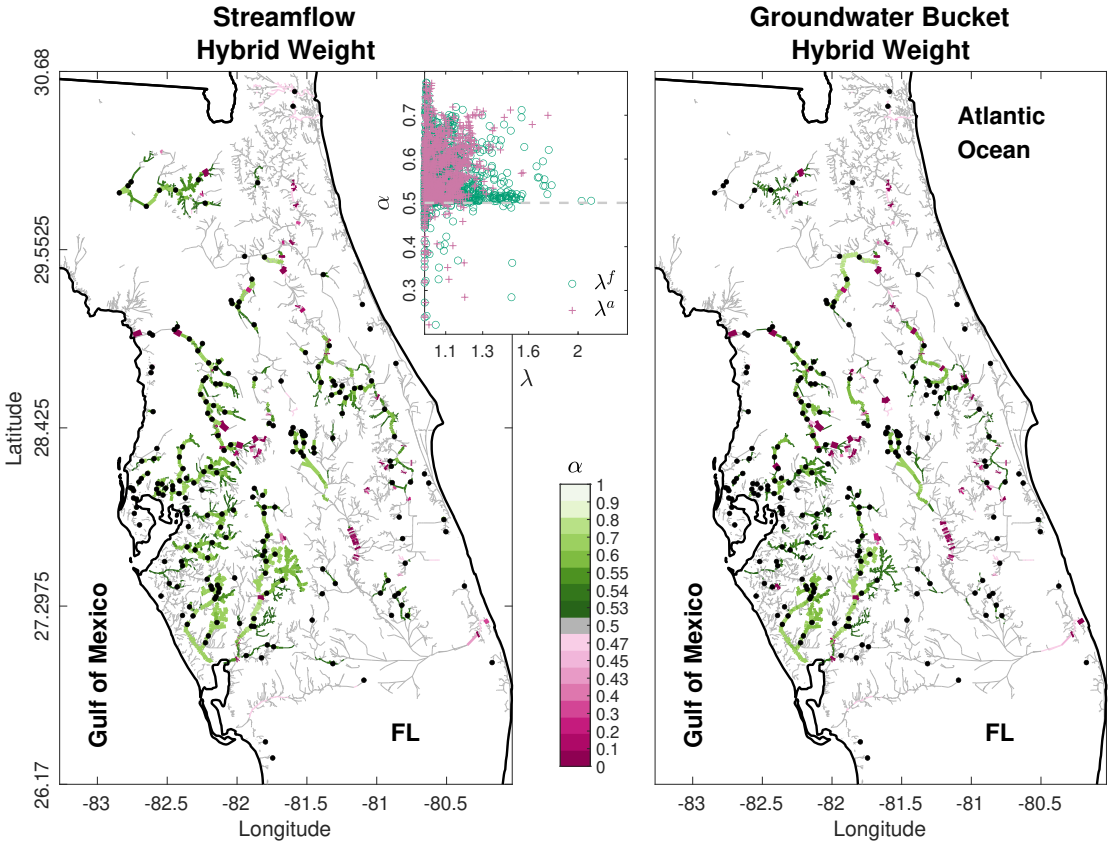


Figure 14. Spatial maps for streamflow (left) and groundwater bucket (right) hybrid weight α. The maps are obtained on the 14th of October, 2022 at 11pm for Ian using the adaptive hybrid scheme. Black dots on the maps show the assimilated USGS gauges. The inset panel (top-middle) plots the time-averaged prior (λ^f) and posterior (λ^a) streamflow inflation on the x-axis and the streamflow hybrid weight on the y-axis (the same weight shown on the map). The initial starting point for the weight; i.e. 0.5, is highlighted by the dashed gray line.

Figure 14 illustrates the spatial variations in hybrid weights for streamflow and groundwater storage on the last DA cycle, October 14, in the FL case. The most significant changes to the hybrid weights are concentrated in the proximity of observation points, aligning with the ATS localization strategy. In reaches far from observations, the hybrid weighting coefficients remain relatively stable, falling within the range of $0.48 < \alpha < 0.52$. It's crucial to recall that ATS localization selectively influences reaches upstream and downstream from a given gauge, using a predetermined cutoff distance, in this case, set at 100 km. Streams with smaller hybrid weights generally indicate limited ensemble variability. A considerable number of streams exhibit increased α, particularly on the western side of the state where Hurricane Ian made landfall. For example, Peace River near Fort Myers and its tributaries predominantly showcase weights exceeding 0.6. The augmentation of α corresponds to an expectation

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions

that streamflow realizations derived from the hydrologic model will align more closely with observed flow, as demonstrated in Figure 12. The adaptive hybrid scheme extends its spatial weight mapping to groundwater storage, an active participant in the assimilation process. The distribution of α for the buckets, as depicted in Figure 14, mirrors that of streamflow. This suggests a non-zero correlation between discharge observations and groundwater storage, leading to multivariate updates in the hydrological state.

Examining streamflow, the most significant deviations from the initial weights (set at 0.5) are observed along streams where inflation was minimal ($\lambda < 1.3$), as highlighted in the inset panel of Figure 14. This observation implies that α undergoes substantial updates in locations where inflation alone couldn't adequately address sampling errors and biases in the ensemble. The synergy between covariance hybridization and ensemble inflation techniques emerges as crucial for enhancing the quality of streamflow predictions. As demonstrated in section 4.2, while inflation tackles current streamflow conditions, long-term biases are effectively mitigated through the incorporation of climatological information-a facet where inflation alone falls short.

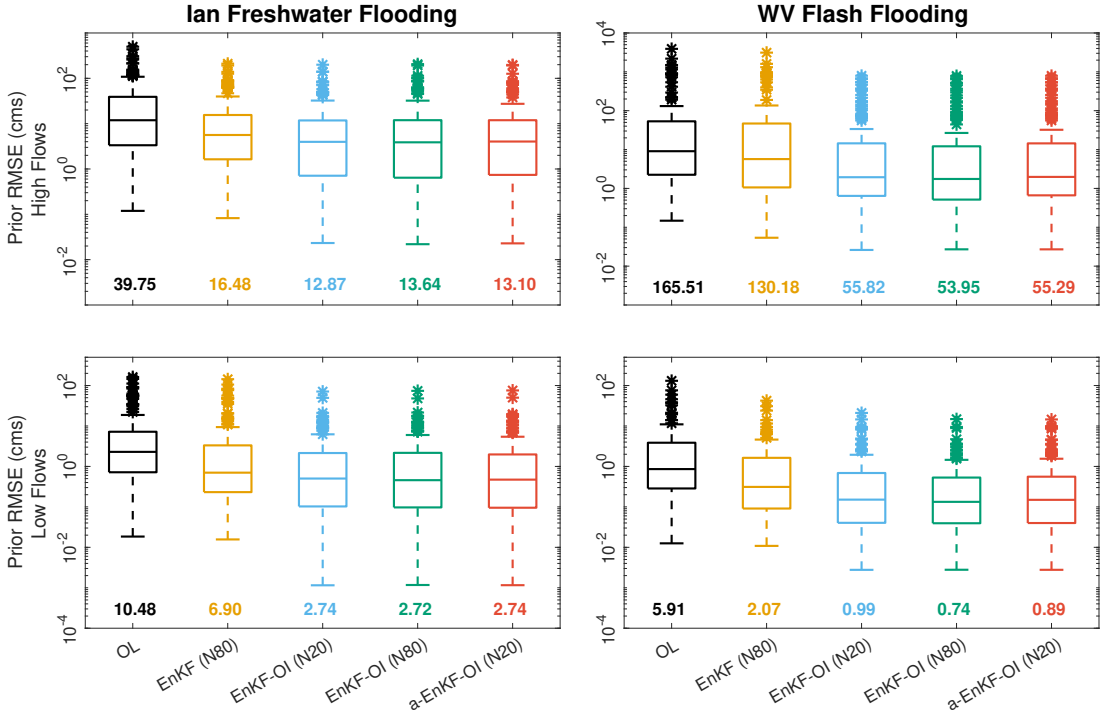


Figure 15. Boxplots for prior RMSE of upper decile (top panels) and lower decile (bottom panels) stream flows. 5 simulation runs are compared: OL, 80-member EnKF, 20-member EnKF-OI ($\alpha = 0.5$), 80-member EnKF-OI ($\alpha = 0.5$) and a 20-member a-EnKF-OI. Ian's flooding results are shown to the left while WV's flash flood estimates are shown to the right. Note that the y-axis is in log-scale. Averaged RMSE values are annotated underneath the individual boxplots.

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024


Hydrology and Earth System Sciences
Discussions

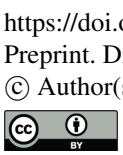
The evaluation of the adaptive hybrid algorithm's performance is succinctly summarized for both low and high stream flows in Figure 15. Low flows are characterized by computing the 10^{th} percentile of observed streamflow at each gauge over the entire assimilation period, while anything above the 90^{th} percentile is deemed a high flow. The a-EnKF-OI with 20 members is compared against the OL, the original EnKF with 80 members, and the fixed weight EnKF-OI ($\alpha = 0.5$) with both 20 and 80 members. In FL, the most substantial improvements offered by the various hybrid schemes over the EnKF are observed for the lowest decile of flows. For instance, the average prior prediction of the a-EnKF-OI (2.74 cms) is 60% more accurate than the EnKF's score (6.90) when considering the lowest flow decile. Conversely, for the highest flow decile, the a-EnKF-OI (13.1 cms) exhibits a 21% gain in prediction skill over the EnKF (16.48), on average. In WV, the advantages of the adaptive hybrid scheme over the EnKF are comparable for both the highest and lowest decile flows. The intriguing behavior observed in the FL case can be attributed to two key factors. Firstly, the Ian simulation in FL spans a full month of streamflow analysis, wherein low-flow periods are more frequent than the main flooding event. Consequently, the performance differences between the schemes are expected to be more pronounced during low-flow periods, highlighting the OI-based scheme's benefits for both floods and non-peak flows. Secondly, unlike WV, where the model exhibits a strong positive bias compared to observed flow, in FL, the model generally suggests a negative bias. Compared to the 20-member EnKF-OI with a fixed α, the adaptive variant demonstrates slightly greater accuracy, particularly evident in Ian's high-flow diagnostics. Overall, the fixed weight 80-member EnKF-OI consistently delivers outstanding RMSE scores for both low and high flows across the two hydrologic domains. The hybrid approach, encompassing various flavors, consistently yields substantial gains compared to the EnKF and remarkable improvements compared to the OL. The 20-member a-EnKF-OI emerges as highly competitive, showcasing exceptional computational efficiency suitable for both flooding and non-flooding applications on larger domains.

4.5 Short-Range Reforecasts

The hybrid scheme and its adaptive variant have demonstrated substantial enhancements in streamflow simulations when contrasted with the OL and the EnKF. Until now, our focus has been solely on verifying analyses (posteriors) and forecasts (priors) within the scope of hourly DA cycling. Put differently, we have yet to explore the influence of analyses beyond the one-hour forecast (prior) timescale. In this section, we delve into assessing the lasting impact of analyses on forecasts spanning up to 18 hours. We aim to unravel questions pertaining to the temporal persistence of state corrections in the model. Furthermore, we seek to identify forecast lead times at which the hybrid streamflow predictions exhibit enhancements over the OL.

To address these inquiries, we conducted a reforecast employing identical forcing data as utilized in the analysis cycles above. A comprehensive reforecast of the NWM would typically involve utilizing its forecast forcing datasets. It is acknowledged that these real-time forecasts might entail larger uncertainties compared to the analysis forcing datasets employed here. However, our primary focus lies in exploring the DA system's capacity to mitigate uncertainties in initial conditions and model biases. The decision to employ retrospective atmospheric forcing allows us to assess the impact of DA on enhancing initial conditions for forecast cycles, without being overshadowed by substantial uncertainties in the forcing dataset, as outlined by Rafieeinasab et al. (2014). Specifically, we employ the AORC forcings in the WV case and the NWMv2.1 analysis and assimilation forcing in the FL case. Given its high skillfulness and computational efficiency, we use the 20-member a-EnKF-OI experiment for both

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024



Hydrology and Earth System Sciences
Discussions

domains. In the context of hourly cycling, the ensemble-mean posterior (analysis) at each hour furnishes initial conditions for predictions spanning up to 18 hours without additional DA. This time frame aligns with the NWMv2.1 short-range forecasts. Given that the same forcings are utilized in the forecasts, our baseline comparison against the OL forecast involves directly assessing the OL run itself.

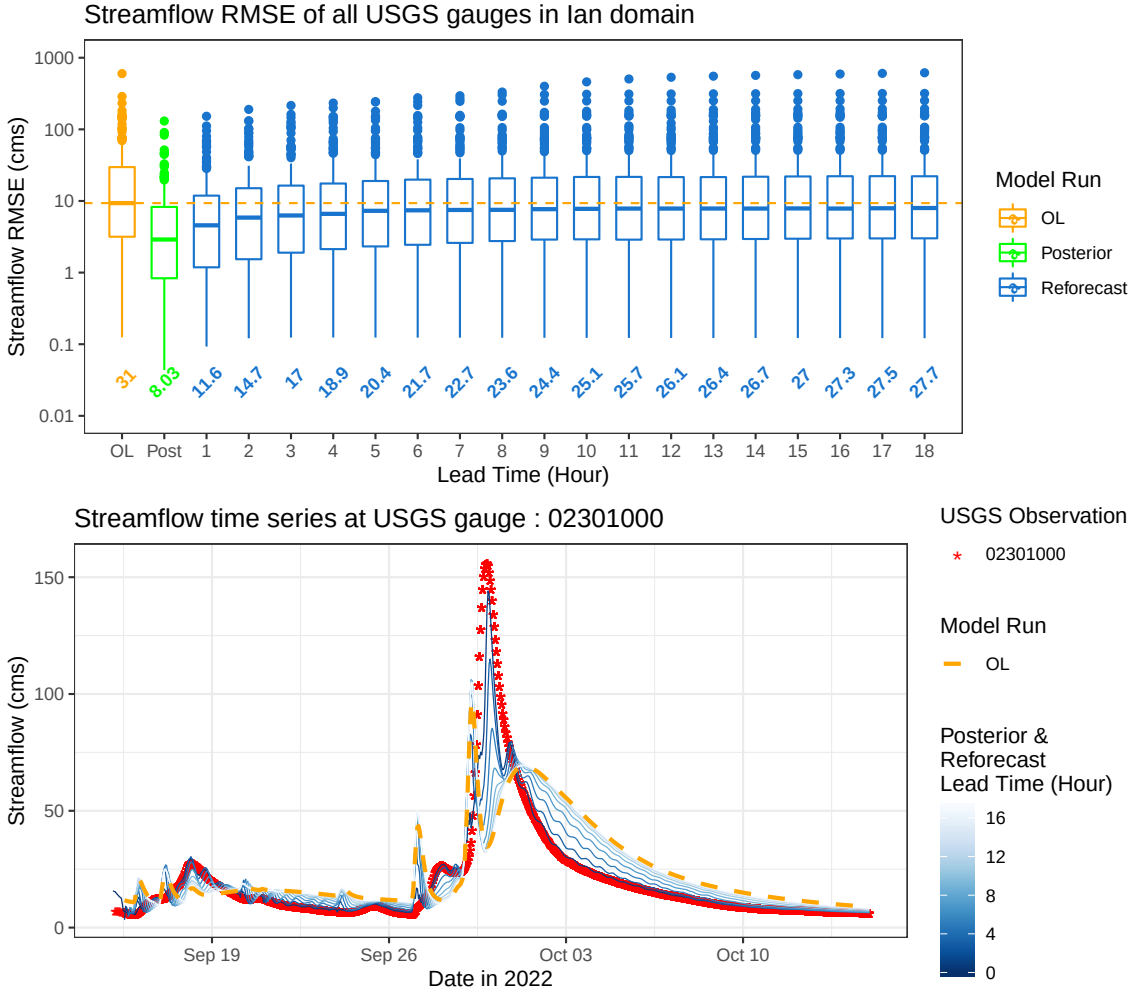


Figure 16. Top Panel: Boxplots summarizing the streamflow RMSE of mean ensemble over time for open loop (orange box), posterior (green box), and reforecasts at different lead times (blue boxes). The horizontal dashed line depicts the median of streamflow RMSE for open loop. Note that the y-axis is in log-scale. Averaged RMSE is annotated underneath the individual boxplots. Note that the y-axis is in log-scale. Bottom Panel: Hydrographs for North Prong Alafia River at Keysville FL (Gauge ID: 02301000) during hurricane Ian. Red dots represent the observed streamflow in cms. The orange dashed line depicts the open loop streamflow simulation. The blue lines show the model posterior (lead time of zero) and reforecasts at different lead times.

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

555 The skill assessment of reforecasts for the FL flooding case is summarized in the top panel of Figure 16. Boxplots illustrate the streamflow RMSE across all gauges in the domain for the OL in orange, the posterior in green, and the reforecasts in blue, considering various lead times (in hours). The horizontal line within each box denotes the median RMSE, while the mean RMSE is provided beneath the boxes. The dashed orange line signifies the median RMSE for the Open Loop. Across all lead times, the reforecasts consistently demonstrate enhanced performance compared to the open loop. However, this improvement 560 diminishes with increasing lead time. For example, at an 18-hour lead time, the reforecasted streamflow (averaged over all gauges) initialized by the a-EnKF-OI estimates is 11% more accurate than the OL. To underscore the significance of DA in achieving more accurate forecasts through improved initial states, hydrographs for the North Prong Alafia River at Keysville, FL (Gauge ID: 02301000) are provided in the bottom panel of Figure 16. This gauge, with a drainage area of 135 square miles and experiencing relatively brief flooding, exemplifies notable improvements in the shorter lead times (< 6 hours) compared 565 to the open loop simulation. However, as the lead time extends to 18 hours, the model's response tends to converge toward the OL solution.

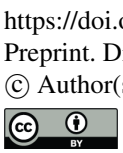
Figure 17 encapsulates the reforecast performance for the WV test case. Similar to Fig. 16, the top panel features boxplots illustrating streamflow RMSE across all USGS gauges in the domain for the OL, the posterior, and various reforecast lead times. Just as observed in the FL case, noteworthy enhancements in streamflow predictions during shorter lead times gradually 570 diminish within the first 6 hours. In a median statistical sense, these improvements align closely with the OL forecasts after approximately 10 hours. Interestingly, when comparing FL's Ian test case to the WV test case, forecasts for the Ian case are superior in quality particularly in higher forecast lead times. This distinction arises from the nature of the WV event, characterized as a short-lived flash flood in contrast to the more prolonged event in FL's Ian test case. Consequently, there is relatively less memory of the DA correction in many streams for the WV event compared to the Ian event in FL.

575 Nevertheless, on average, the reforecasts for the WV test case consistently outperform the OL up to hour 18. This trend is attributed to the fact that the mean RMSE is generally dominated by RMSEs of large rivers. Large rivers, characterized by an enduring memory, preserve the impact of DA for many hours, resulting in a considerable reduction in their error metrics. Consequently, the mean RMSE across the domain remains lower than that of the OL, even for forecast lead times exceeding 10 hours. The bottom panel in Figure 17 exemplifies this phenomenon at the Kanawha River. In the OL simulation, there 580 is a substantial underestimation of streamflow compared to the observations (depicted by red stars). With the assimilation of streamflow into the a-EnKF-OI, predictions at short lead times significantly align with the observations. However, as the forecast lead time extends, the prediction gradually converges towards the OL solution. Given the nature of this relatively large river and the multi-day duration of the event at this location, the model's estimates remain notably more accurate than the OL even after 18 hours.

585 5 Conclusions

In this study, we have delved into the innovative application of hybrid ensemble and variational data assimilation techniques for streamflow and flood prediction within the WRF-Hydro National Water Model (v2.1) configuration and the Data As-

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024



Hydrology and Earth System Sciences
Discussions
Open Access
EGU

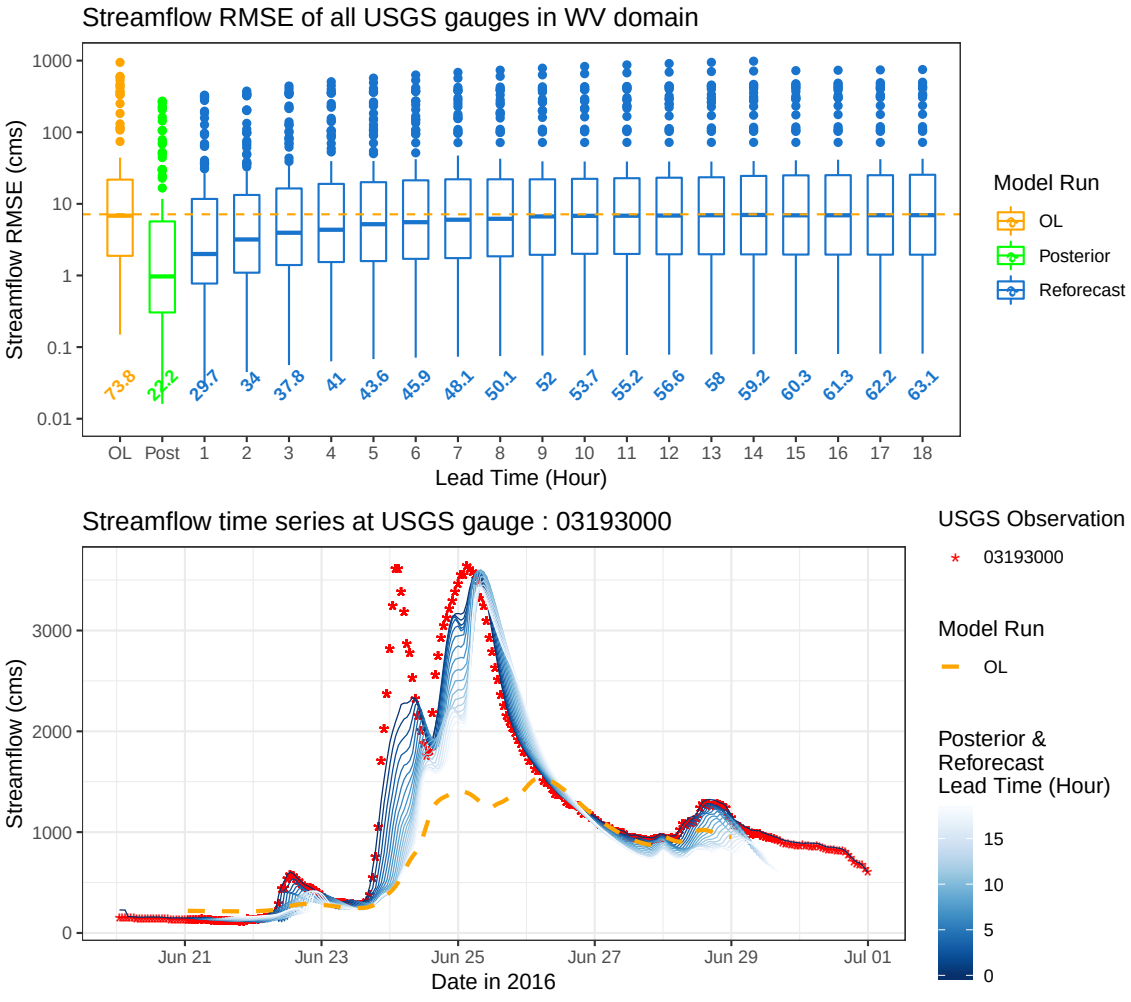


Figure 17. Similar to Fig. 16 but for Kanawha River at Kanawha Falls WV (Gauge ID: 03193000).

simulation Research Testbed (DART). The resulting "HydroDART" system is specifically tailored to offer precise ensemble streamflow predictions during challenging flood events, such as intense rainfall and hurricanes. HydroDART leverages the ensemble Kalman filter, incorporating adaptive covariance inflation and along-the-stream localization to address issues like sampling errors and bias (e.g., El Gharamti et al., 2021). Our system delivers hourly streamflow analyses utilizing data from the extensive USGS gauging network across the United States.

The hybrid ensemble-variational scheme presented in this paper seamlessly combines the time-varying sample error covariance derived from the ensemble with a static climatological error covariance commonly employed in systems like optimal interpolation and 3/4D-Var. Our comprehensive testing across two distinct basins-West Virginia's flash flooding in June 2016

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024


Hydrology and Earth System Sciences
Discussions

and Florida's inland flooding caused by Hurricane Ian in August 2022-demonstrates that the hybrid algorithm outperforms the EnKF, significantly enhancing prediction accuracy.

Our findings reveal that the judicious blending of the static background covariance with the EnKF effectively improves the ensemble spread, successfully mitigating pronounced model biases observed during flooding events. Optimal results from the hybrid filter were achieved when assigning equal weight to the ensemble and climatology (i.e., a hybrid weight of 0.5). Notably, relying solely on climatological information while disregarding the dynamic ensemble led to degraded results and yielded poor-quality discharge estimates. Across various gauges in both hydrologic basins, the hybrid scheme exhibited near-perfect alignment with observations, boasting efficiency metrics such as NSE and KGE very close to 1.

Crucially, the hybridized covariance not only heightened prediction skill but also demonstrated exceptional efficiency by operating effectively with a time-varying ensemble that utilized only 25% of the members employed by the EnKF alone. Our results indicate that employing 20 realizations in the dynamic ensemble is sufficient to maintain high-accuracy streamflow predictions. This, coupled with the NWM submodel design of HydroDART (utilizing only the NWM's streamflow, conceptual groundwater storage, and reservoir models), suggests that a system like HydroDART may be computationally efficient enough for operational use.

Additionally, we explored an adaptive variant of the hybrid scheme that automatically adjusts the hybrid weight at each stream and bucket in the domain based on ensemble statistics. The adaptive scheme, employing 20 members, demonstrated robustness and high skillfulness. Finally, we conducted short-range streamflow forecasts initiated from the hybrid scheme analyses and compared the results to the open loop, revealing consistent improvements in model forecasts for up to 18 hours.

Despite its successful application, the examined framework warrants further research and sensitivity studies. For instance, the ATS localization was specifically tuned for the EnKF and not its hybrid counterparts. It remains conceivable that, with the integration of climatological information, a reduction in localization (i.e., broader cutoff radii) might be possible. The potential for mitigating spurious correlations arising from limited ensemble sizes through the application of the static background covariance was evident in the observations made using the hybrid EnKF-OI scheme, particularly with a minimal number of ensemble realizations in Section 4.3.

In the case of the adaptive hybrid variant, our study focused on scenarios where the hybrid weight initiates at 0.5. Acknowledging that smaller weights may compromise performance, an exploration of commencing the algorithm with larger weights could be a plausible avenue. This strategy might vary across different domains, considering the dynamic nature of hydrologic basins and water conditions. It is noteworthy that the weight coefficients across the stream network in FL and WV did not converge to specific values. This lack of convergence was attributed to the changing ensemble conditions, such as spread and bias, over our relatively short simulation periods. Extending the simulation duration could potentially lead to the convergence of hybrid weights as streamflow conditions stabilize. In our case, the adaptive algorithm predominantly assigned a balanced weight to the majority of gauges, i.e., $\alpha \in [0.48, 0.52]$. This aligns with the demonstrated optimal performance of a constant homogeneous weight of 0.5 (Section 4.2). For extended simulations, one might consider implementing a reset mechanism for the hybrid weights or exploring the utilization of seasonal climatological covariances rather than a single one.

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

A potential extension of the present study involves assessing the efficacy of the hybrid DA approach in medium and long-range forecasts. This expansion could encompass a broader array of hydrologic variables, such as stream temperature, and involve additional modeling components like soil moisture and snow. Furthermore, instead of focusing on specific flooding events, our forthcoming investigations will delve into evaluating the performance of the hybrid DA methodology in a comprehensive simulation covering the entire CONUS. This strategic shift aims to ascertain the robustness of the latest HydroDART version across diverse hydrologic conditions and to analyze its computational complexity within a large-scale domain.

Code availability. The data assimilation code used in this study is openly available as part of the DART repository (directory path: DART/-models/wrf_hydro) on GitHub; https://doi.org/10.5065/D6WQ0202 (last access: 15 September 2022, DART team, 2022). The model code is also freely available and can be accessed at https://ral.ucar.edu/projects/wrf_hydro (last access: 15 September 2022, WRF-Hydro team, 2022).

Data availability. The datasets, scripts and routines used to generate the results of this work can be accessed through the following public Zenodo repository: https://doi.org/10.5281/zenodo.5532569 (El Gharamti, 2023).

Author contributions. MEG developed the hybrid algorithm and implemented its parallel version in DART, ran some DA experiments and wrote more than 70% of the paper. AR prepared the hydrologic domains, ran the experiments and wrote sections 2.1, 3 and 4.5. JLM maintained the Python configuration framework, developed the retro run procedure in section 3.3 and reviewed the work.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Acknowledgements. This work was performed as part of NCAR's Short Term Explicit Prediction (STEP) Program, which is supported by the National Science Foundation funds for the United States Weather Research Program (USWRP). The authors would like to acknowledge high-performance computing support from Cheyenne (https://doi.org/10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation.

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Hydrology and Earth System Sciences
Discussions

References

Abbaszadeh, P., Moradkhani, H., and Daescu, D. N.: The quest for model uncertainty quantification: A hybrid ensemble and variational data assimilation framework, Water resources research, 55, 2407–2431, 2019.

Abbaszadeh, P., Gavahi, K., and Moradkhani, H.: Multivariate remotely sensed and in-situ data assimilation for enhancing community WRF-Hydro model forecasting, Advances in Water Resources, 145, 103 721, 2020.

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., and Avellano, A.: The data assimilation research testbed: A community facility, Bulletin of the American Meteorological Society, 90, 1283–1296, 2009.

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

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

Anderson, J. L.: Localization and sampling error correction in ensemble Kalman filter data assimilation, Monthly Weather Review, 140, 2359–2371, 2012.

Anderson, J. L. and Anderson, S. L.: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Monthly weather review, 127, 2741–2758, 1999.

Bannister, R. N.: A review of operational methods of variational and ensemble-variational data assimilation, Quarterly Journal of the Royal Meteorological Society, 143, 607–633, 2017.

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

Boucher, M.-A., Quilty, J., and Adamowski, J.: Data assimilation for streamflow forecasting using extreme learning machines and multilayer perceptrons, Water Resources Research, 56, e2019WR026 226, 2020.

Bucci, L., Alaka, L., Hagen, A., Delgado, S., and Beven, J.: Hurricane Ian, https://www.nhc.noaa.gov/data/tcr/AL092022_Ian.pdf, 2023.

Chao, L., Zhang, K., Wang, S., Gu, Z., Xu, J., and Bao, H.: Assimilation of surface soil moisture jointly retrieved by multiple microwave satellites into the WRF-Hydro model in ungauged regions: Towards a robust flood simulation and forecasting, Environmental Modelling & Software, 154, 105 421, 2022.

Counillon, F., Sakov, P., and Bertino, L.: Application of a hybrid EnKF-OI to ocean forecasting, Ocean Science, 5, 389–401, 2009.

DART team: DART, Data Assimilation Research Section (DAReS), DAReS/CISL/NCAR [code], National Center for Atmospheric Research (NCAR), https://doi.org/10.5065/D6WQ0202, 2022.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 131, 3385–3396, 2005.

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

El Gharamti, M.: Hybrid Ensemble–Variational Filter: A Spatially and Temporally Varying Adaptive Algorithm to Estimate Relative Weighting, Monthly Weather Review, 149, 65–76, 2021.

El Gharamti, M.: Hybrid HydroDART Data, https://doi.org/10.5281/zenodo.8309815, 2023.

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, Monthly Weather Review, 147, 2535–2553, 2019.

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

El Gharamti, M., McCreight, J. L., Noh, S. J., Hoar, T. J., RafieeiNasab, A., and Johnson, B. K.: Ensemble streamflow data assimilation using WRF-Hydro and DART: novel localization and inflation techniques applied to Hurricane Florence flooding, Hydrology and Earth System Sciences, 25, 5315–5336, 2021.

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, Journal of Hydrometeorology, 21, 453–474, 2020.

Evensen, G.: The ensemble Kalman filter: Theoretical formulation and practical implementation, Ocean dynamics, 53, 343–367, 2003.

Falcone, J. A.: GAGES-II: Geospatial attributes of gages for evaluating streamflow, Tech. rep., US Geological Survey, 2011.

Fall, G., Kitzmiller, D., Pavlovic, S., Zhang, Z., Patrick, N., St. Laurent, M., Trypaluk, C., Wu, W., and Miller, D.: The Office of Water Prediction's Analysis of Record for Calibration, version 1.1: Dataset description and precipitation evaluation, JAWRA Journal of the American Water Resources Association, 2023.

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

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Quarterly Journal of the Royal Meteorological Society, 125, 723–757, 1999.

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, https://doi.org/10.5281/zenodo.3625238, 2020.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of hydrology, 377, 80–91, 2009.

Hamill, T. M. and Snyder, C.: A hybrid ensemble Kalman filter–3D variational analysis scheme, Monthly Weather Review, 128, 2905–2919, 2000.

Hernández, F. and Liang, X.: Hybridizing Bayesian and variational data assimilation for high-resolution hydrologic forecasting, Hydrology and Earth System Sciences, 22, 5759–5779, 2018.

Hersbach, H.: Decomposition of the continuous ranked probability score for ensemble prediction systems, Weather and Forecasting, 15, 559–570, 2000.

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

Huang, C., Newman, A. J., Clark, M. P., Wood, A. W., and Zheng, X.: Evaluation of snow data assimilation using the ensemble Kalman filter for seasonal streamflow prediction in the western United States, Hydrology and Earth System Sciences, 21, 635–650, 2017.

Kalman, R. E.: A new approach to linear filtering and prediction problems, 1960.

Kerandi, N., Arnault, J., Laux, P., Wagner, S., Kitheka, J., and Kunstmann, H.: Joint atmospheric-terrestrial water balances for East Africa: a WRF-Hydro case study for the upper Tana River basin, Theoretical and Applied Climatology, 131, 1337–1355, 2018.

Kim, S., Shen, H., Noh, S., Seo, D.-J., Welles, E., Pelgrim, E., Weerts, A., Lyons, E., and Philips, B.: High-resolution modeling and prediction of urban floods using WRF-Hydro and data assimilation, Journal of Hydrology, 598, 126 236, 2021.

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, Journal of Hydrology, 575, 596–611, 2019.

Lei, L., Wang, Z., and Tan, Z.-M.: Integrated hybrid data assimilation for an ensemble Kalman filter, Monthly Weather Review, 149, 4091–4105, 2021.

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Lorenc, A. C.: Analysis methods for numerical weather prediction, Quarterly Journal of the Royal Meteorological Society, 112, 1177–1194, 1986.

Martinaitis, S. M., Albright, B., Gourley, J. J., Perfater, S., Meyer, T., Flamig, Z. L., Clark, R. A., Vergara, H., and Klein, M.: The 23 June 2016 West Virginia flash flood event as observed through two hydrometeorology testbed experiments, Weather and Forecasting, 35, 2099–2126, 2020.

Matheson, J. E. and Winkler, R. L.: Scoring rules for continuous probability distributions, Management science, 22, 1087–1096, 1976.

Mazrooei, A., Sankarasubramanian, A., and Wood, A. W.: Potential in improving monthly streamflow forecasting through variational assimilation of observed streamflow, Journal of Hydrology, 600, 126 559, 2021.

McMillan, H., Hreinsson, E., Clark, M., Singh, S., Zammit, C., and Uddstrom, M.: Operational hydrological data assimilation with the recursive ensemble Kalman filter, Hydrology and Earth System Sciences, 17, 21, 2013.

Meng, S., Xie, X., and Liang, S.: Assimilation of soil moisture and streamflow observations to improve flood forecasting with considering runoff routing lags, Journal of hydrology, 550, 568–579, 2017.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I?A discussion of principles, Journal of hydrology, 10, 282–290, 1970.

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, Journal of Hydrologic Engineering, 18, 1684–1696, 2013.

Nunez, C. and Yang, A.: Here's how Hurricanes Form-and why they're so destructive, https://www.nationalgeographic.com/environment/article/hurricanes-typhoons-cyclones, 2023.

Parrish, D. F. and Derber, J. C.: The National Meteorological Center's spectral statistical-interpolation analysis system, Monthly Weather Review, 120, 1747–1763, 1992.

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, Journal of hydrology, 519, 2663–2675, 2014.

Rakovec, O., Weerts, A., Hazenberg, P., Torfs, P., and Uijlenhoet, R.: State updating of a distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency and observation network density on forecast accuracy, Hydrology and Earth System Sciences, 16, 3435–3449, 2012.

Rappaport, E. N.: Fatalities in the United States from Atlantic tropical cyclones: New data and interpretation, Bulletin of the American Meteorological Society, 95, 341–346, 2014.

Sacher, W. and Bartello, P.: Sampling errors in ensemble Kalman filtering. Part I: Theory, Monthly Weather Review, 136, 3035–3049, 2008.

Senatore, A., Mendicino, G., Gochis, D. J., Yu, W., Yates, D. N., and Kunstmann, H.: Fully coupled atmosphere-hydrology simulations for the central M editerranean: Impact of enhanced hydrological parameterization for short and long time scales, Journal of Advances in Modeling Earth Systems, 7, 1693–1715, 2015.

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

Smith, A. B.: U.S. billion-dollar weather and climate disasters, 1980 - present (NCEI accession 0209268), 2020.

Son, Y., Di Lorenzo, E., and Luo, J.: WRF-Hydro-CUFA: A scalable and adaptable coastal-urban flood model based on the WRF-Hydro and SWMM models, Environmental Modelling & Software, 167, 105 770, 2023.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, Journal of geophysical research: atmospheres, 106, 7183–7192, 2001.

https://doi.org/10.5194/hess-2023-269
Preprint. Discussion started: 25 January 2024

Troin, M., Arsenault, R., Wood, A. W., Brissette, F., and Martel, J.-L.: Generating ensemble streamflow forecasts: A review of methods and approaches over the past 40 years, 2021.

765 Wang, W., Liu, J., Xu, B., Li, C., Liu, Y., and Yu, F.: A WRF/WRF-Hydro coupling system with an improved structure for rainfall-runoff simulation with mixed runoff generation mechanism, Journal of Hydrology, 612, 128 049, 2022.

Weerts, A. H. and El Serafy, G. Y.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water resources research, 42, 2006.

Whitaker, J. S. and Hamill, T. M.: Evaluating methods to account for system errors in ensemble data assimilation, Monthly Weather Review,
770 140, 3078–3089, 2012.

WRF-Hydro team: WRF-Hydro, RAL/NCAR [code], National Center for Atmospheric Research (NCAR), https://ral.ucar.edu/projects/wrf_hydro, 2022.

Yucel, I., Onen, A., Yilmaz, K., and Gochis, D.: Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction model, data assimilation and satellite-based rainfall, Journal of Hydrology, 523, 49–66, 2015.

775 Zhang, F., Snyder, C., and Sun, J.: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter, Monthly Weather Review, 132, 1238–1253, 2004.