https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


# BARRIERS TO OPERATIONAL FLOOD FORECASTING IN COMPLEX TERRAIN: FROM PRECIPITATION FORECASTS TO PROBABILISTIC FLOOD FORECAST MAPPING AT SHORT LEAD TIMES

Luiz Bacelar[1], Arezoo ReifeeiNasab[3], Nathaniel Chaney[1], and Ana Barros[2]

[1]Department of Civil and Environmental Engineering, Duke University
[2]Department of Civil and Environmental Engineering, University of Illinois Urbana-Champaign
[3]National Center for Atmospheric Research

**Correspondence:** Luiz Bacelar (luiz.bacelar@duke.edu)

**Abstract.** As flood alert systems move towards higher spatial resolutions, there is a continued need to enable approaches that provide robust predictions of flood extent that adequately account for the uncertainties from meteorological forcing, hydrologic and hydraulic model structure, and parameter uncertainty. In flood forecasting, two primary sources of uncertainty are the quantitative precipitation forecasts (QPF) and the representation of the channel and floodplain geometry. This is especially relevant as simple approaches (e.g., HAND) are being used to map floods operationally at field scales (< 10 m). This article investigates the benefits of using a computationally efficient probabilistic precipitation forecast (PPF) approach to generate multiple flood extension scenarios over a region of complex terrain prone to flash floods. First, we assess the limitations of using a calibrated version of the gridded version of the WRF-Hydro model to predict an extreme flash flood event in the Greenbrier River Basin (West Virginia) on 24 June 2016. We investigated an ensemble methodology to combine operational High-Resolution Rapid Refresh (HRRR) QPF with radar-based Quantitative Precipitation Estimates, specifically MRMS QPE products. This approach was most effective to increase the headwaters streamflow accuracy in the first hour lead time, which is still insufficient to issue actionable flood warnings in operational applications. At longer lead-times, success was elusive due to epistemic uncertainties in MRMS rainfall intensity and HRRR rainfall spatial patterns. Furthermore, a QPF ensemble was used to generate an ensemble of flood heights using the HAND flood mapping methodology at different spatial resolutions. Results revealed a scale-dependency with increasing dispersion among the predicted flooded areas with increasing spatial resolution down to 1 meter. We hypothesize the overprediction of flooded areas at higher spatial resolutions reflects the increasing number of river reaches and the need for scale-aware representation of river hydraulics that impacts flood propagation in the river network.

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


## 1 Introduction

Climate projections suggest that in the future there will be an increase in flash flood frequency and severity, mainly due to an increase in extreme rainfall events combined with continued urbanization (Hocini et al. (2021),Li et al. (2022)). This is especially true in watersheds with steep topography and lower surface permeability conditions that are already highly prone to the flashiness of flood waves (Gourley et al. (2017)). As a result, there is a high need to mitigate flash flood risks in urban areas in tropical and subtropical mountainous areas. One solution is to build dams in headwaters to dampen the flood wave that reaches populated areas (Mondal et al. (2021)). However, cost constraints limit the feasibility of this approach in many regions; this is especially true in developing countries. An alternative option is for continued investments in early warning systems to drastically reduce damage while minimizing the associated costs; this approach is also seen as a more sustainable alternative to the environment (Ward et al. (2020)). However, this approach remains limited by the lack of flash flood forecasting skills in early warning systems (Kuller et al. (2021). As High-Performance Computing (HPC) continues to increase the potential modeling capabilities, physically based fully distributed hydrological-hydraulic models are becoming more common in state-of-the-art flood forecast chains (Flack et al. (2019)). Ongoing scientific advancements and the use of advanced modeling techniques have continuously improved flash flood predictability, but uncertainties in outcomes cannot be completely eliminated.

Mapping potential flood areas remains a critical aspect of land use planning and urban expansion (Yu et al. (2021)). Authorities manage to decrease the flood risk levels by avoiding new urbanized areas in pre-determined high-risk locations via zoning regulations. These static flood risk maps are calculated following statistical attributes of local rainfall and, or streamflow time-series, in combination with the geometry description of the river channel and floodplain and physically based hydraulic formulations (Mudashiru et al. (2021)). The advances in computation, digital elevation maps, and flood mapping using state-of-the-art hydraulic models have enabled operational agencies to produce maps of potential flood-delineated areas to define the high-level risk of flood areas according to return periods (Knighton et al. (2020)). However, these static maps provide only climatological bounds and cannot be relied upon to support emergency response for specific flood events(Di Baldassarre et al. (2020)). This is due in part to the need to assess antecedent soil moisture conditions prior to a specific event, which strongly affects runoff generation at small scales and thus impacts flood propagation (Ran et al. (2022)). Furthermore, the scale dependency of rainfall and runoff processes increases the uncertainty of those maps to estimate the potential flash flood hazards.

Maps are indeed a helpful visual product to stakeholders when managing the relocation of human resources during a crisis (Evers et al. (2012)). An alternative is to increase the accuracy of near-real-time flood maps for response to natural disasters (Oddo and Bolten (2019)). In the context of flash floods, real-time hazard mapping quantifies the flood extent not only based on floodplain water height but also the velocity of the flood waves (Mudashiru et al. (2021)). The destructive potential of water can cause damage to roads, buildings, and vehicles (Bocanegra and Francés (2021)). However, developing high-resolution flood maps for existing flood forecast systems remains a persistent challenge (Braud et al. (2018)). This is in part due to the lack of availability of high-resolution digital elevation models (<5m) in headwaters, that are needed to describe the channel

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


cross-sections and floodplain geometries necessary for hydraulic transport simulations. Even with enough data availability, complex 1D and 2D hydraulic models are usually not chosen for operational tasks since they can be computationally expensive at very high resolution. Therefore, operational flood warning systems often adopt simplified methodologies, which derive the flood extent maps based on planar intersections across the DEM (Teng et al. (2017); Hu and Demir (2021)). For instance, the National Water Model (NWM) framework currently uses HAND (height above nearest drainage) (Rennó et al. (2008)) maps along with WRF-Hydro streamflow outputs to generate flood extension maps at 10 meters of spatial resolution over the United States territory (Viterbo et al. (2020)).

The use of non-physically based methodologies for flood mapping can increase epistemic uncertainties along the channel and floodplain for different hydraulic conditions. In other words, the streamflow simulated by the hydrological model is not two-way coupled to the final flood extent but instead, river discharge is mapped into water levels according to synthetic rating curves (Zheng et al. (2018)). The final change in velocity and water height necessary to adjust the flood extent will therefore not feedback into the calculation of the streamflow in the hydrological model. This introduces mass conservation and velocity errors that can lead to underestimation or overestimation of real-time flood extent, especially in complex topography where dynamic flood propagation effects could be more relevant (Teng et al. (2017)). Like any other flood mapping methodology, the roughness coefficient plays an important role to formulate the synthetic rating curves, usually defined per river reaches. This creates a tendency in the HAND methodology to overpredict the flood extent, even when the streamflow used as input is approximately accurate if the roughness coefficient is not well calibrated (Li et al. (2022)). Recently Scriven et al. (2021) tested different roughness coefficients along the floodplain and after an optimum configuration concluded that topographic regions with steep river gradients and relatively long reaches (> 5km) exhibited more accurate flood delineation by the HAND methodology. This shows that the HAND methodology can be suitable for flash flood mapping if local uncertainties are well-defined. Scriven's evaluations were made considering the epistemic uncertainties in the final methodology (channel geometry, flow direction, DEM resolution) but did not address the uncertainties related to input streamflow forecast (i.e., hydrological model input, calibration process) for different watersheds. It is well known that in operational hydrological models, a certain level of parameter calibration is necessary for delivering an accurate streamflow prediction. Uncertainty in hydrological model forcing impacts the calibrated parameters, including the floodplain roughness coefficient. These uncertainties are propagated differently according to basin size, physiography, climate region, and the dominant runoff mechanisms for different flood events (Moges et al. (2021)).

To efficiently improve flood warning systems it is therefore necessary to evaluate the uncertainties of each modeling component adopted as input to forecast the hazard. Those analyses follow a forecast chain perspective since some output components are used as input in the next process of the chain (Viterbo et al. (2020); Hofmann and Schüttrumpf (2020) ). An issue currently addressed is how the observed rainfall (i.e., quantitative precipitation estimates - QPE) can influence the hydrological parameter calibration of operational models, and how the final choice of parameters can impact the runoff processes simulations in real-time (Wijayarathne et al. (2021)). The spatial and temporal representation of precipitation data has been shown to be one of the greatest sources of uncertainties in hydrological models, especially in smaller watersheds prone to flash flood events (Beven and Binley (1992); Beven (2010); Liao and Barros (2022)). Some flood warning system evaluations have also focused to show the limitations of short-term rainfall predictions when applied to operational hydrological models (Tao and Barros (2013)). The next generation of Quantitative precipitation forecast (QPF) relies on numerical weather prediction (NWP) models and short-term data assimilation techniques of weather radar data, which therefore brings to the forefront a wide range of uncertainty including the parameterization of cloud and precipitation processes in models and the measurement uncertainty of the observations themselves that are assimilated (Dowell et al. (2022)).

To diminish the propagation of short-term rainfall errors through hydrological models, the use of ensemble forecasts to quantify uncertainty in flood warning systems(FWS) and to communicate the uncertainty in probabilistic terms to the public (Wu et al. (2020)). Ensemble forecasts increase the overall FWS reliability since each probabilistic scenario helps bound the uncertainties driven by rainfall prediction (Cloke and Pappenberger (2009)). Recent studies have shown that rainfall ensembles generated via post-processing techniques can be as reliable as short-term QPF members originating from different initial conditions or atmospheric models (Crochemore et al. (2016)). Post-processing of deterministic rainfall fields using geostatistical approaches requires minimal computational resources and can be run using multiple outputs from atmospheric models to produce different rainfall members (Caseri et al. (2016), Cecinati et al. (2017) and Hartke et al. (2022)). Statistically based rainfall members are proven to increase the overall accuracy of streamflow prediction in distributed hydrological models (Falck et al. (2021) and Valdez et al. (2022)). Moreover, the cost-benefit of investing in new techniques for QPF ensemble could be beneficial to any application that relies on rainfall nowcasting, not only flood forecasting systems (Guzzetti et al. (2020)).

Although flood mapping has been analyzed in the context of the operational forecast chain uncertainty, few studies have addressed how flood mapping ensembles could improve current forecast systems (Zahmatkesh et al. (2021)). This article intends to 1) evaluate how the accuracy of forecast chain components influences flash flood prediction, and 2) investigate whether the probabilistic approach decreases the flood prediction uncertainty in complex terrain. We follow a forecasting framework similar to the National Water Model (NWM) and propose methodologies that could be beneficial for higher-resolution flash flood spatial analysis. In addition, we explored a new methodology for flood ensemble mapping based on geostatistical QPF, WRF-Hydro, and HAND.

The study region is the Greenbrier River Basin, in the Appalachian Mountains region of West Virginia which is highly prone to flash flood events. First, we assess the QPE and QPF accuracy tied to streamflow prediction for six extreme rainfall events. Probabilistic flash flood maps for the 2016 event that caused 23 fatalities (Martinaitis et al. (2020)) are evaluated in detail. Unlike the NWM, we approached calibrating the WRF-Hydro model relying on diffusive wave routing to capture flood propagation which is shown to be more suitable for extreme flash floods events due to abrupt changes in flood height (Moussa and Bocquillon (2009)). We further compared results using a new Lidar-based digital elevation model (DEM) at 1 m spatial resolution to forecast flash flood ensemble maps with results using the current NWM flood maps at 10 m spatial resolution.

## 2 Data and Methodology

In the following subsections, we will outline the data and experiments proposed as a workflow to evaluate the forecast chain depicted in Figure 1 (b). To build the forecast chain we include collection and preprocessing of meteorological (i.g. precipi-

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


tation) and surface description (i.g. digital elevation model) data. A physically based fully distributed hydrological-hydraulic model is the core component of the forecast chain. To configure the model we take the collected preprocessed data and calibrate the model parameters using historical flood events (Figure 1 (c)) to improve streamflow prediction.. Additionally, ensemble precipitation forecasts are used as inputs to the hydrological-hydraulic model for near-real-time flood extent uncertainties analysis.

## 2.1 Study area and flood events

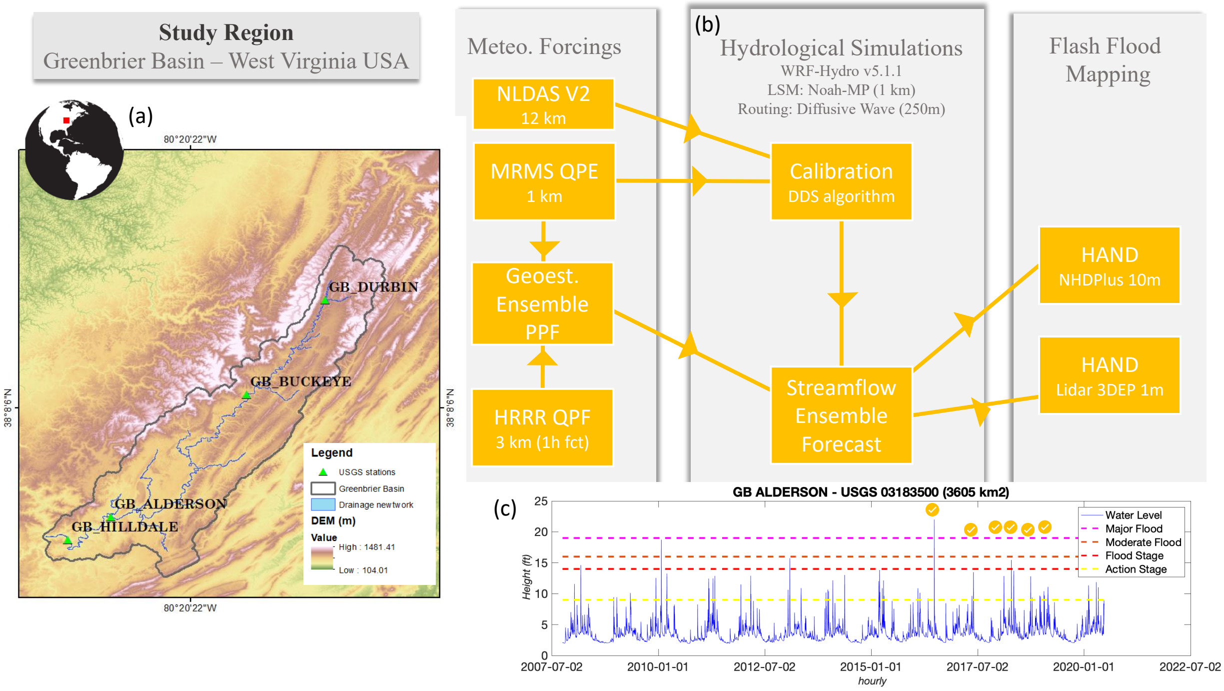



**Figure 1.** (a) Greenbrier River Basin (4290 km2) and Elevation map. (b) Data and Models for the Flood Mapping Forecast Chain. (c) 6 events selected based on activity stage threshold in Alderson gauge

As part of the Appalachian Mountains region in the Eastern United States, the Greenbrier River basin has approximately 4290 km$^2$ of drainage area (Figure 1 a). It is considered one of the longest undammed rivers in West Virginia. The river flows 261 km, from upstream areas in Randolph and Pocahontas counties to downstream in Greenbrier, Summers, and Monroe counties. The complex topography is characterized by steep valleys and thus Greenbrier tributaries are highly prone to flash flood events. One of the most recent events in 2016 led to 1 billion dollars in damages. It also caused 23 fatalities, most of them in the White Sulphur Springs neighborhood in West Virginia (Martinaitis et al. (2020)).

Constrained by QPE data availability, we selected six flash flood events between 2016 and 2018, in which the extreme rainfall scenarios caused the USGS station hydrological station in Alderson to exceed its action (high) water stage threshold of

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


9 ft (2.74 meters; Figure 1 (c)). Table 1 shows the water level peak time for each of the events. The approximate time to peak was calculated based on the hours between the initial time of the rainfall event (gauge measurement in mm/h) over the basin and the water level at the Alderson gauge.

**Table 1.** Description of extreme rainfall events used in this study. The peak time corresponds to the water level peak at Alderson station. The time to peak was calculated between the initial time of rainfall to the peak of water level. The precipitation values represent the maximum observed by the daily rain gauge and the maximum MRMS pixel over the basin domain.

| | Flow peak time | Time to peak (h) | Peak (ft) | Max. Accumulated Rainfall Gauge (mm) | Max. Accumulated Rainfall QPE (MRMS) (mm) |
|---|---|---|---|---|---|
| Event 1 | 2016-06-24 05:00 | 8 | 21.99 | 262.3 | 306.4 |
| Event 2 | 2018-04-17 01:00 | 16 | 15.46 | 116.3 | 78.18 |
| Event 3 | 2018-05-04 01:00 | 15 | 13.69 | 118.7 | 112.09 |
| Event 4 | 2018-09-28 14:00 | 10 | 12.81 | 93.3 | 74.91 |
| Event 5 | 2017-05-25 17:00 | 12 | 13.49 | 117.9 | 68.1 |
| Event 6 | 2018-02-20 07:00 | 23 | 12.86 | 101.8 | 59.35 |

### 2.2 Quantitative Precipitation Estimates (QPE)

Precipitation data plays a critical role in flood forecast chains, and its accuracy is subject to considerable uncertainty (Clark et al. (2014) and Valdez et al. (2022)). The spatial misrepresentation of rainfall patterns has been identified as a significant source of error that can substantially impact hydrological predictions, as discussed by Clark et al. (2008), Rakovec et al. (2014) and Clark et al. (2017).

In our study, we utilized the Multi-Radar/Multi-Sensor System (MRMS) data as the instantaneous Quantitative Precipitation Estimate (QPE). MRMS is an operational product that combines data from dual-polarization ground weather radar and satellite passive microwave measurements to derive surface rainfall estimates over continental areas (Zhang et al. (2016)). This dataset was selected due to its widespread use and capabilities in capturing precipitation information (Moazami and Najafi (2021)).

### 2.3 Quantitative Precipitation Forecasts (QPF)

To examine the influence of deterministic short-term rainfall predictions within the forecast chain, we incorporated the High-Resolution Rapid Refresh version 3 (HRRRv3) dataset (Dowell et al. (2022)). This dataset encompasses rainfall forecasts generated through the Weather Research and Forecasting-Advanced Research WRF (WRF-ARW) model version 3.8.1, employing a 3km spatial grid resolution. The HRRRv3 dataset leverages multiple data assimilation techniques and observation sources, including the assimilation of ground weather radar data. These methods enhance the reinitialization of atmospheric states such as air humidity and wind velocity on an hourly basis. Consequently, the accuracy of cloud formation and dissipation in short-term lead time forecasts is improved.

For our investigation, we specifically focused on the early range lead times of up to 3 hours. Starting from the model initialization at instant t=0, we extracted accumulated rainfall predictions (in mm/h) for the subsequent 1, 2, and 3 hours. These short-term predictions allowed us to evaluate the impact of deterministic rainfall forecasts on the overall performance of the flood forecast chain. Note that due to the short-response time of floods in headwater basins, actionable forecasts must be available 3-6 hours in advance.

### 2.4 Probabilistic Precipitation Forecast (PPF)

To optimize the utilization of Quantitative Precipitation Forecast (QPF) data and explore the uncertainties associated with capturing the spatial distribution of rainfall, we employed a Sequential Gaussian Distribution (SGD) methodology to generate QPF ensembles, referred to as Probabilistic Precipitation Forecast (PPF). This approach aligns with the findings discussed by Min et al. (2021), which highlight the challenges of accurately modeling extreme rainfall intensities (measured in mm/h) using atmospheric nowcasting models, particularly in comparison to the primary direction of thunderstorms. The complexity arises from the sub-grid representation of cloud and precipitation microphysics and challenges in the assimilation of radial motion velocity data obtained from ground-based weather radar systems (Sun et al. (2014)).

In our methodology, we assumed that it is still feasible to correct rainfall intensities at early forecast lead times, specifically within the range of up to 3 hours (t=t+1, t=t+2, and t=t+3), based on the Quantitative Precipitation Estimate (QPE) available at t=0. The underlying assumption is that precipitation processes that impact rainfall intensity have characteristic persistence of 1-3 hours even as the storm propagates. This is a critical assumption that will be discussed later in the manuscript. In other words, the PPF ensembles maintain the spatial representation of rainfall in the HRRR QPF. However, the intensities, corresponding to the high rainfall rates, are adjusted based on the most recent availability of MRMS data in real-time. This adjustment ensures that the PPF ensembles capture both the spatial distribution of rainfall (i.e., thunderstorm motion) from the HRRR QPF and the intensity corrections derived from the latest MRMS data.

To generate the multiplicative bias factor, a regular spatial grid with a resolution of 5 km was utilized over the study domain. Point values of the High-Resolution Rapid Refresh (HRRR) and Multi-Radar/Multi-Sensor System (MRMS) datasets were extracted to construct the bias factor, as illustrated in Figure 2. In this study, a total of 20 spatial error fields were derived using a semi-variogram model of multiplicative bias. These error fields served as the basis for 20 random fields of kriging interpolation. The multiplicative point values were required to satisfy the conditions for each of the 20 ensemble members, while the interpolated errors between the points, for each realization of the kriging field, followed a random Gaussian distribution, taking into account the statistical properties of the multiplicative bias variogram. The fundamental principle of the Sequential Gaussian Distribution (SGD) methodology is that as the number of random ensembles increases, the ensemble mean will converge to a simple kriging interpolation (Bai and Tahmasebi (2022)).

For each of the three groups of HRRR forecast lead times, up to 3 hours (t=t+1, t=t+2, and t=t+3), a different multiplicative error variogram was employed based on the Quantitative Precipitation Estimate (QPE) data at t=0. Figure 2 provides an example of the methodology for the HRRR rainfall forecast at 23:00 on June 23, 2016 (t=t+1), and the corresponding MRMS data at 22:00 on June 23, 2016 (t=0). The corrected HRRR rainfall for 23:00 on June 23, 2016 (t=t+1), utilizing the ensemble mean

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


from 20 ensembles, resulted in a reduction of extreme rainfall intensities (above 150 mm/h), while simultaneously increasing rainfall rates in less intense areas compared to the MRMS data.

Initially, we assessed the impact of correcting HRRR rainfall fields based on in-situ rain gauges, considering both daily and hourly accumulations. Subsequently, we analyzed the influence of the 20 rainfall ensemble members on the hydrological model, focusing on the Greenbrier River basin. This analysis demonstrated how different spatial patterns of precipitation can affect streamflow predictions along the length of the river.

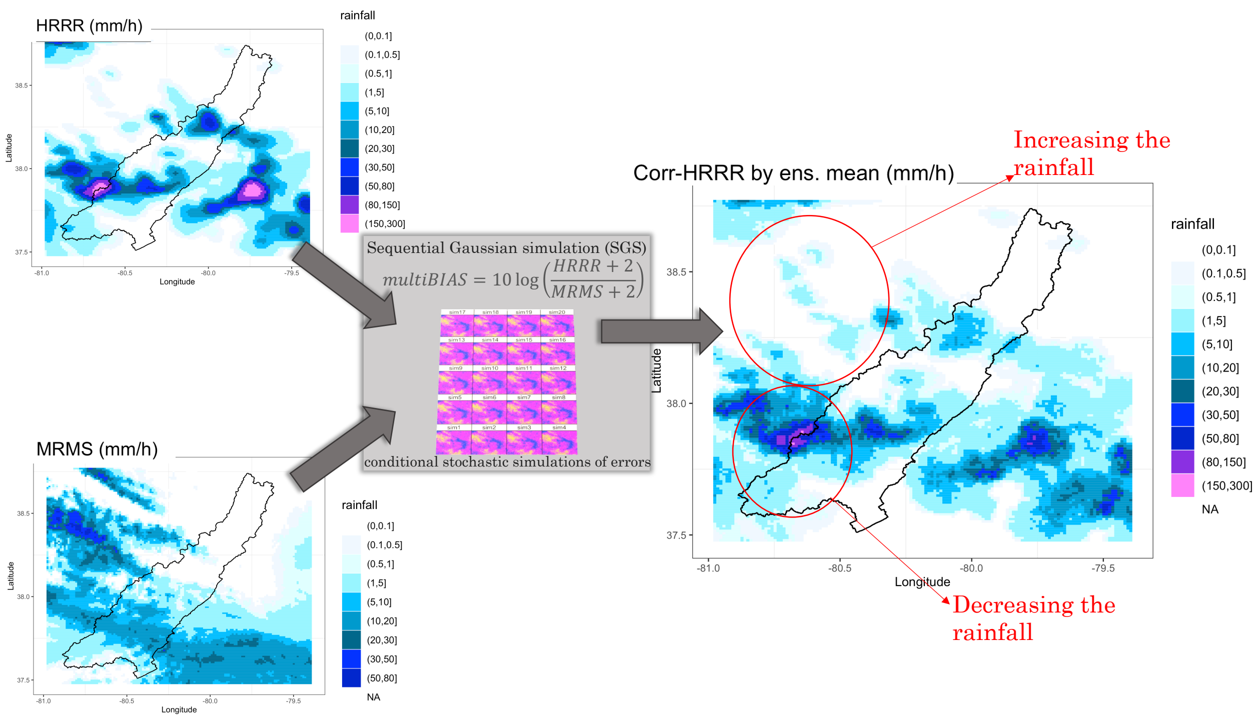



**Figure 2.** PFF methodology to generate 20 conditional stochastic simulations of error fields between HRRR and MRMS. The example shows the HRRR rainfall forecasted to 2016/06/23 23:00 (t=t+1) and the MRMS data at 2016/06/23 22:00 (t=0). Ensemble Mean of corrected HRRR for 2016/06/23 23:00 (t=t+1)

## 2.5 Calibrating parameters in the WRF-Hydro Model

To account for atmospheric interactions with previously neglected components of the Earth system in simulating the hydrological cycle, community models like the Weather Research and Forecast (WRF) model have been enhanced with extended parameterizations. The WRF-Hydro model (Gochis et al. (2015)) serves as an extension of the WRF framework specifically designed for hydrological simulations, enabling the simultaneous representation of atmospheric fluxes (such as rainfall) and river discharge within the same time-step.

In this integrated system, the Noah-MP Land Surface Model (LSM) (Niu et al. (2011)) is one-way coupled to river and terrain routing parametrizations, in a way that floodplain streamflow does not reinfiltrate back to the LSM grid cells. The

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


Noah-MP LSM simulates surface energy and water fluxes, providing runoff inputs to calculate the river discharge. When the atmospheric modeling component is deactivated, WRF-Hydro can function as a standalone gridded distributed hydrological model. Typically, it is common practice to implement the LSM at a coarser spatial grid resolution compared to the river network grid. For example, in our study, the surface and subsurface runoff are computed by the Noah-MP LSM at a resolution of 1 km, which then feeds into a river network grid with a resolution of 250 m, as depicted in Figure 3. Once the runoff reaches the channel banks, the streamflow is routed downstream using the diffusive wave routing approximation. This routing parametrization is well-suited for capturing abrupt changes in water height, particularly in regions with steep topography (**?**). Such hydraulic effects are shown to be observed during flash flood events (Ding et al. (2022)).

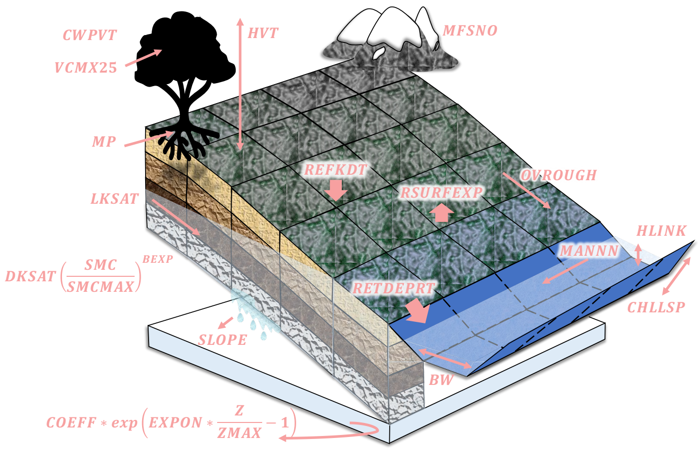
In order to ensure that the modeled mass and energy balance aligns with observed discharge measurements in the Greenbrier River, a calibration workflow was conducted for selected parameters within the Noah-MP Land Surface Model (LSM) and the diffusive wave routing scheme, as depicted in Figure 3. The parameters targeted for calibration were identified by the National Water Model developers as highly sensitive to changes in streamflow (Lahmers et al. (2021); Mascaro et al. (2023)), (Table in Figure 3). To streamline the calibration process, we leveraged the existing calibration conducted for operational purposes as a starting point. This allowed us to redefine the parameter ranges and facilitate the search for an optimal configuration. The Dynamic Dimensional Search (DDS) method (Tolson and Shoemaker (2007)) was employed for calibration, utilizing hourly streamflow observations as the calibration target. The DDS algorithm, implemented and evaluated within a Python interface workflow by the developers of WRF-Hydro, has demonstrated computational efficiency and suitability for hydrological applications at a continental scale (Silver et al. (2017)).

In this study, we chose the NSE as a metric for the optimum configuration in the Greenbrier River. In other words, the DDS searched for the best group of parameters which could give the highest NSE value between the simulated WRF-Hydro and observed USGS river discharges. The calibration workflow considered the Hilldale USGS station data as the optimum downstream point. Therefore, changes in the 20 parameters presented in Figure 3 would influence strongly the hydrology of the area upstream Hilldale' (drainage area of 4207 $km^2$). Firstly, we performed a simple test with 50 DDS iterations over 2016, separately for each parameter, to investigate changes in NSE values ($Fobj$). The saturated hydraulic conductivity multiplier (DKSAT), which controls the velocity of water transfer among soil grid layers in the Noah-MP, was considered the most individually sensitive parameter ($Fobj$ = 0.72; Figure 3) followed by the REFKDT parameter associated with scaling infiltration partitions to direct surface runoff and SMCMAX, that adjusts the maximum soil moisture capacity of all soil layers. River routing parameters describing the channel properties and geometry (roughness coefficient, width, and slope) were also highly relevant to changes in simulating streamflow. A more extensive Noah-MP parameter sensitivity analysis can be found at Cuntz et al. (2016).

We performed WRF-Hydro calibration considering 300 iterations over hourly simulations between 06/2015 and 12/2020. Therefore, all six extreme rainfall events were constrained to the calibration period. A more detailed description of WRF-Hydro parameters can be found in the supplementary material. The calibration relied on MRMS QPE at 1km for rainfall. Air temperature, pressure, humidity, incoming radiation, and wind were bilinearly interpolated to 1 km from the NLDAS v2 (Xia et al. (2012)) data at 12km. The complex terrain in Greenbrier Basin was originally described by the digital elevation model

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


from NHDPlus at 10 meters resolution (Moore et al. (2019)). We used the default WRF-Hydro Land Use and Land Cover (LULC) maps which rely on monthly MODIS images and look-up tables.




| | Δ NSE | Parameter | Component | Description |
|---|---|---|---|---|
| 1 | 0.72 | DKSAT | 1D LSM | Saturated soil hydraulic conductivity |
| 2 | 0.65 | REFKDT | 1D LSM | Surface infiltration parameter |
| 3 | 0.32 | SMCMAX | 1D LSM | Maximum soil moisture content |
| 4 | 0.30 | OVROUGH | T. ROUTING | Overland flow roughness coefficient |
| 5 | 0.30 | MANNN | C. ROUTING | Manning's roughness coefficient |
| 6 | 0.04 | BW | C. ROUTING | Channel bottom width [m] |
| 7 | 0.04 | CHSSLP | C. ROUTING | Channel side slope [m/m] |
| 8 | 0.03 | RSURFEXP | 1D LSM | Exponent in the resistance for soil evaporation |
| 9 | 0.02 | MP | 1D LSM | Slope of conductance to photosynthesis (Ball-Berry) |
| 10 | <0.01 | VCMX25 | 1D LSM | Maximum rate of carboxylation at 25 [umol co2/m/s] |
| 11 | <0.01 | LKSATFAC | 1D LSM | Multiplier on lateral saturated hydraulic conductivity |
| 12 | <0.01 | BEXP | 1D LSM | Beta parameter |
| 13 | <0.01 | SLOPE | 1D LSM | Linear reservoir coefficient from bottom drainage |
| 14 | <0.01 | REFDEPFAC | 1D LSM | Multiplier on retention depth for overland flow |
| 15 | <0.01 | CWVT | 1D LSM | Empirical canopy wind parameter |
| 16 | <0.01 | HVT | 1D LSM | Top of canopy [m] |
| 17 | <0.01 | COEFF | GW | Multiplier on bucket drainage as function of depth |
| 18 | <0.01 | EXPON | GW | Exponent of rate of bucket drainage |
| 19 | <0.01 | HLINK | C. ROUTING | Initial water depth in the channel [m] |
| 20 | <0.01 | MSFNO | 1D LSM | Snow melt factor for snow depletion curve |
| 21 | <0.01 | ZMAX | GW | Maximum groundwater bucket depth |

**Figure 3.** WRF-Hydro parameters considered for calibration using the DDS algorithm. The Table shows how the objective function NSE changes according to each individual parameter. 1D LSM represents the components from Noah-MP. Terrain Routing, Channel Routing and Groundwater are parametrizations from WRF-Hydro.

## 2.6 Probabilistic Streamflow Forecast (PSF)

Following the calibration of WRF-Hydro parameters using observed hourly streamflow data, we conducted near-real-time experiments to assess the model's streamflow prediction performance. For these experiments, we utilized both deterministic High-Resolution Rapid Refresh (HRRR) Quantitative Precipitation Forecast (QPF) with lead times up to three hours, as well as the corrected Probabilistic Precipitation Forecast (PPF) rainfall fields. During the evaluation, we focused on the same six extreme events listed at the table at Figure 3. For each event, we performed deterministic streamflow simulations using the HRRR QPF and generated a set of 20 streamflow predictions for each lead time. This ensemble of streamflow predictions formed our Probabilistic Streamflow Forecast (PSF). To assess the accuracy of the deterministic and probabilistic streamflow predictions, we compared them against available streamflow observations along the Greenbrier River. This evaluation aimed to determine the model's ability to capture the observed streamflow patterns and provide reliable forecasts during extreme events.

Through these near-real-time experiments, we aimed to evaluate and validate the performance of the WRF-Hydro model in predicting streamflow, both in deterministic and probabilistic modes, using a combination of HRRR QPF and corrected PPF rainfall fields as input.

## 2.7 Probabilistic Flood Mapping Forecast (PFF)

In our experiment using the Probabilistic Streamflow Forecast (PSF), we aimed to assess the potential of higher-resolution topographic data for generating probabilistic flood maps. Specifically, we utilized the National Elevation Dataset (NED) 3D

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


LIDAR product with a spatial resolution of 1 meter, as opposed to the operational version of the National Water Model (NWM) which relies on a 10-meter description of topography (Zheng et al. (2018)).

To evaluate the accuracy of the generated flood maps, we compared them against field-collected flood benchmarks provided by the US Geological Survey (USGS) at White Sulphur Springs. This neighborhood was significantly affected by the 2016 extreme flood event (Watson and Cauller (2017)). The observed flood extent was initially in vector format and was subsequently converted into raster format to match the spatial resolution of the High-Resolution National Hydrography Dataset-Digital Elevation Model (HAND-DEM) used in our analysis as reference. The Probabilistic Streamflow Forecast (PSF) was aimed at assessing the potential of using a higher spatial resolution topographic dataset for generating probabilistic flood maps in the study area.

### 2.8 Metrics

Ensemble metrics are necessary to demonstrate the accuracy, reliability, sharpness, and skill of any type of probabilistic forecast. To account for the forecast uncertainties of each component in our forecast chain, we applied statistical evaluation to the workflow proposed in Figure 1.

Equation 1 shows the Pearson Correlation coefficient used to verify the linear correlation between $O$ and $S$, the observed and simulated variables respectively. The Root Mean Squared Error (Equation 2) was used to measure the absolute accuracy differences between the simulated and observed variables. The multiplicative Bias (MBIAS; Equation 3) reveals how much (%) the simulations are being either overestimated or underestimated, when compared to the observations. Values of MBIAS lower than 1 indicate underestimation while values above 1 shows overestimation. The Nash-Sutcliffe Coefficient (NSE) was additionally applied between the time-series of the hydrological simulations (i.e. streamflow) and river gauges to verify the time-scale accuracy of the hydrologic model. The ensemble reliability was only verified though the ensemble mean for the precipitation ensembles and streamflow ensembles.

$$PearsonCor. = \frac{\sum_{t=1}^{n}(S_i - \bar{O})^2}{\sum_{t=1}^{n}(O_i - \bar{O})^2} \tag{1}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i - O_i)^2} \tag{2}$$

$$MBIAS = \frac{\sum_{i=1}^{N}(S_i - O_i)}{\sum_{i=1}^{N} O_i} \tag{3}$$

$$NSE = 1 - \frac{\sum_{t=1}^{n}(S_i - O_i)^2}{\sum_{t=1}^{n}(O_i - \bar{O})^2} \tag{4}$$

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


The package *rwfhydro* was used for some of the metrics at Session 3.3 to evaluate the streamflow calibration process of the hydrological model. The WRF-Hydro R-package has functions to calculate the accuracy statistics (i.e. RMSE, NSE ) against gauge observations at hourly, daily, and monthly time-scales.

The Equations 5, 6 and 7 were used to verify how the ensemble mean of simulated flood maps were comparable to the brenchmark collected by the USGS after the natural disaster. The Probability of Detection (POD), False Alarm Ratio (FAR) and Characteristic Stability Index (CSI) are calculated based on the number of pixels that successfully predicted ($n_{success}$), overpredicted ($n_{falsealarme}$) or underpredicted ($n_{fail}$) the flood extent.

$$POD = \frac{n_{success}}{n_{success} + n_{fail}} \tag{5}$$

$$FAR = \frac{n_{falsealarme}}{n_{success} + n_{falsealarm}} \tag{6}$$

$$CSI = \frac{n_{success}}{n_{success} + n_{fail} + n_{falsealarm}} \tag{7}$$

By following this proposed workflow, the forecast chain depicted in Figure 1 can be thoroughly evaluated, providing insights into the performance and limitations of the system for flash flood prediction. The workflow integrates data collection, model configuration, ensemble precipitation forecasting, flood simulation, evaluation, and uncertainty analysis to ensure a comprehensive assessment of the forecast chain's effectiveness over the Greenbrier River Basin.

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


## 3 Results

### 3.1 How accurate is the QPE in the complex terrain of the Greenbrier watershed?

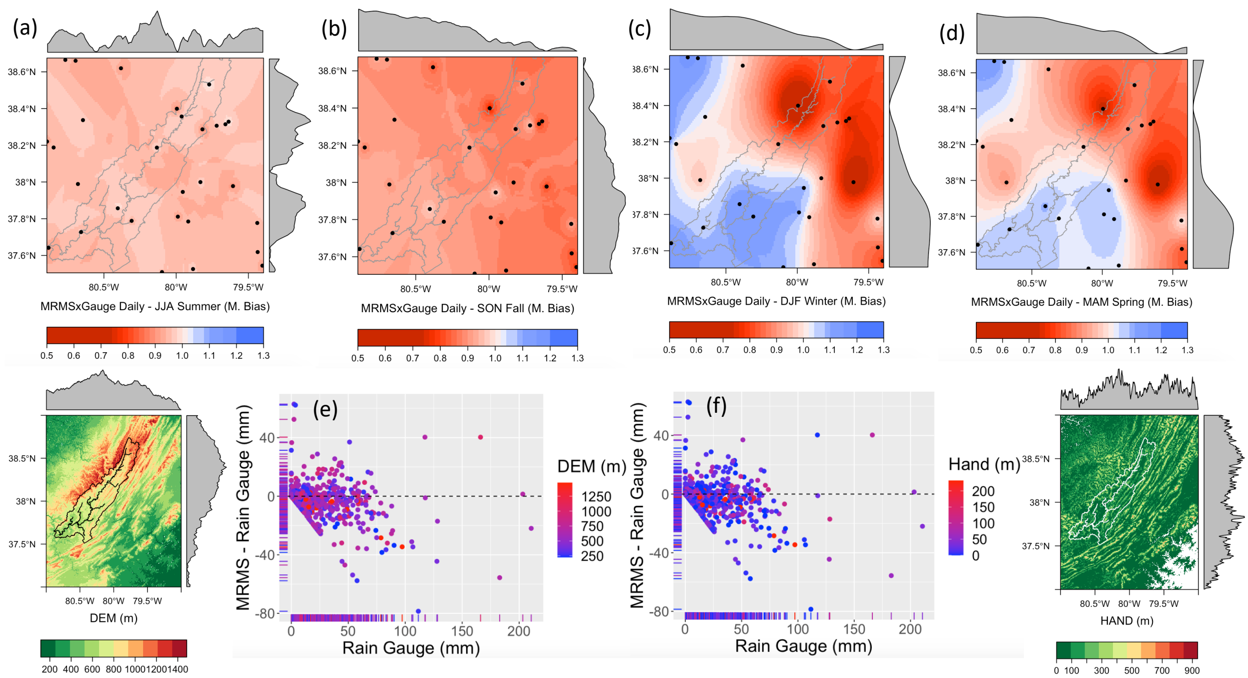



**Figure 4.** Kriging interpolation of MBIAS for (a) Summer (June, July, and August), (b) Fall (September, October, and November), (d) Winter (December, January, February), (d) Spring (March, April, May) between 05/2015 and 03/2020. The remaining figures illustrate the daily rainfall anomalies (mm) according to (e) Digital Elevation Model (m) and (f) Height above nearest drainage (m).

As mentioned in the methodology session, the hydrological model calibration, and its initial state for the forecast as well as the correction of QPF rely on the accuracy of QPE. Our first analysis focused on discussing the uncertainties of MRMS products to retrieve rainfall in complex terrain. As discussed by Cecinati et al. (2017) and Arulraj and Barros (2019), ground radar data can suffer inferences of the terrain according to the composition of the scan. This assumes that the rainfall would not be uniformly retrieved over the basin, and areas upstream would present a systematic error if the product was not real-time corrected with automatic hourly rain gauges, though this correction would be different depending on the storm and whether hourly rain gauges can capture the variability of rainfall. Over our study domain the USGS hourly rain gauges are scarce (Figure 2S) compared to daily rain gauges which can be found in higher density (Figure 4). Therefore, we decided to compare points of MRMS retrieval with observed daily rainfall accumulation to investigate possible systematic errors of estimation according to the steep terrain. This a limitation for flash flood forecasting and fast response in mountainous regions, where the diurnal cycle of rainfall is very strong (e.g., Liao and Barros (2022)). Figure 4 shows the kriging interpolation of MBIAS between observed daily rain gauges

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


and daily MRMS accumulation time series for (a) Summer, (b) Fall, (d) Winter, and (e) Spring between 05/2015 and 03/2020. It was detected an overall underestimation of MRMS rainfall (around 10 %) during the Summer and Fall. An overestimation of around 20% could be found in the lower Greenbrier River during the Winter and Spring. Without seasonal segregation, the upper basin remains in about 10% of underestimation against a reduction to 5% of overestimation in downstream areas at daily time-scales (Figure 1S). Furthermore, the supplementary material also demonstrates that the rainfall time-series correlation (i.e., R2) is lower with higher accuracy error (i.e., RMSE) for upstream areas.

We performed an additional MRMS analysis against the only available hourly rain gauge inside the basin, located at high elevations(Figure 2S). The seasonal analysis of the rainfall diurnal cycle revealed an overall MRMS underestimation, especially during winter months. However, it was observed that during the summer higher MRMS hourly rainfall rates between 12 and 21 local time, suggesting an overestimation of precipitation related to local diurnal convection in the warmer seasons. This is likely due to the overcorrection of radar products at high elevations due to overshooting (Arulraj and Barros (2021)). Additionally, when each one of the 6 extreme rainfall events was analyzed separately, we note the underestimation of QPE for 5 of those events (Figure 3S and Table 2).

### 3.2 How are uncertainties in deterministic QPF propagated to PPF Ensemble?

In this section, we analyzed the accuracy of HRRR deterministic QPF as well as the corrected HRRR probabilistic PPF during the 6 rainfall extreme events in the Greenbrier River Basin. We performed our statistical comparison considering the spatial mean over the 4 hourly rain gauges available in the domain shown in Figure 6 (b). The statistical metrics between the HRRR QPF at lead times up to 3 hours and the hourly rain gauges can be found in Table 3. The results show that the HRRR first hour forecast ($HRRR_{fct01}$), for instance, rainfall prediction for 01:00 with HRRR initialization at 00:00 compared to hourly rain gauges at 01:00, demonstrates higher time-series correlation (Pearson Cor) with observed rainfall when compared to lead time in 2 and 3 hour forecasts. Although, the second-hour forecast ($HRRR_{fct02}$) presented slightly lower RMSE and multiplicative Bias (MBIAS) closer to 1, suggesting that the second-hour forecasts, in general, could lead to more accurate rainfall magnitude, the first HRRR time-step ($HRRR_{fct01}$) is generally more accurate in to predict storms direction.

Figure 5 shows the spatial pattern of (a) MRMS QPE, (b) HRRR first, (c) second, and (d) third-hour forecasts of total accumulated rainfall during June 2016 flood event. The additional overlapping values of total rainfall accumulation from daily *in situ* gauges facilitate the interpretation of overestimation or underestimation across the domain. There is underestimation of QPE in upstream areas of high elevation , as mentioned previosuly, as well as the overestimation at the downstream areas at lower elevation. Comparing the MRMS QPE spatial patterns with the HRRR QPF, we can notice that the position of most intense rainfall accumulation was more accurately described by the second-hour forecast Figure 5(c). The QPE and QPF's ability to capture the peak time of intense rainfall rates is shown in Figure 6(c), where the time series of average rainfall is compared against rain gauge measurements. The MRMS showed a good agreement with USC00463669 rain gauge, as the peak of precipitation was placed at 06/23/2016 02:00 and 06/23/2016 22:00. All 3 lead-times QPFs predicted a certain level of precipitation for 06/23/2016 02:00 but concentrated the highest rainfall intensities in the early afternoon hours.

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


**Table 2.** Pearson Correlation, RMSE and multiplicative Bias (MBIAS) for HRRR deterministic QPF and MRMS QPE. Statistics considered 27 points of daily gauge measurements as observed rainfall. The last row represents the mean value over the six events.

| | | Event1 | Event2 | Event3 | Event 4 | Event 5 | Event 6 | All Events |
|---|---|---|---|---|---|---|---|---|
| Pearson Correlation | MRMS | 0.72 | 0.77 | 0.85 | 0.85 | 0.91 | 0.91 | 0.78 |
| | $HRRR_{fct01}$ | 0.27 | 0.62 | 0.25 | 0.42 | 0.23 | 0.49 | 0.34 |
| | $HRRR_{fct02}$ | 0.06 | 0.60 | 0.45 | 0.65 | 0.21 | 0.70 | 0.33 |
| | $HRRR_{fct03}$ | 0.11 | 0.42 | 0.41 | 0.32 | 0.42 | 0.55 | 0.31 |
| RMSE | MRMS | 2.57 | 1.84 | 0.93 | 0.79 | 0.54 | 0.34 | 1.32 |
| | $HRRR_{fct01}$ | 5.67 | 2.22 | 2.01 | 1.5 | 1.91 | 1.87 | 2.72 |
| | $HRRR_{fct02}$ | 5.14 | 2.34 | 1.67 | 1.36 | 1.83 | 1.95 | 2.55 |
| | $HRRR_{fct03}$ | 4.13 | 5.84 | 1.68 | 1.88 | 1.78 | 2.13 | 3.25 |
| MBIAS | MRMS | 0.86 | 1.05 | 1.12 | 1.03 | 0.91 | 1.06 | 0.99 |
| | $HRRR_{fct01}$ | 0.88 | 0.66 | 0.76 | 0.64 | 1.02 | 1.74 | 0.9 |
| | $HRRR_{fct02}$ | 0.89 | 0.72 | 0.84 | 0.83 | 0.97 | 2.25 | 1.01 |
| | $HRRR_{fct03}$ | 0.54 | 1.30 | 0.77 | 0.74 | 1.10 | 2.46 | 1.10 |

Figure 6 also shows the results of SGD methodology to adjust intensities of QPF based on the last QPE map, the PFF. Figure 6 (d), (e), and (f) represent the time series of accumulated rainfall considering the 20 PFF ensemble members generated for the first, second, and third hours of HRRR forecasts. The ensembles produced rainfall scenarios above the rain gauge values for the first and second forecast hours, which can explain why the ensemble mean reaches accumulation values closer to the observed. Considering the statistics presented in Table 3, the second-hour deterministic QPF showed lower RMSE and MBIAS, which lead to a lower overestimation of the ensemble members (Figure 6 (e)), compared to the first-hour forecast. However, the first-hour PPF ensemble mean (Figure 5 (e)) was able to adjust the rainfall intensities closer to the QPF (Figure 5 (b)) as well as the second-hour QPF (Figure 5 (c)) as previously demonstrated. To exemplify how is the effect of the correction over different rainfall intensities, Figure 6 (a) shows the histogram of first-hour PPF, by the percentage of pixels between certain rainfall rates over time. We noticed that most of the members started to differ from the ensemble mean for precipitation ranges between 30 and 100 mm/h (30 < P <100) and above 100 mm/h (P > 100). The PFF was not able to completely reduce areas where the QPF values were above 100 mm/h, not estimated by the MRMS. However, the ensemble members were able to reduce the percentage of values above 100 mm/h.

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023

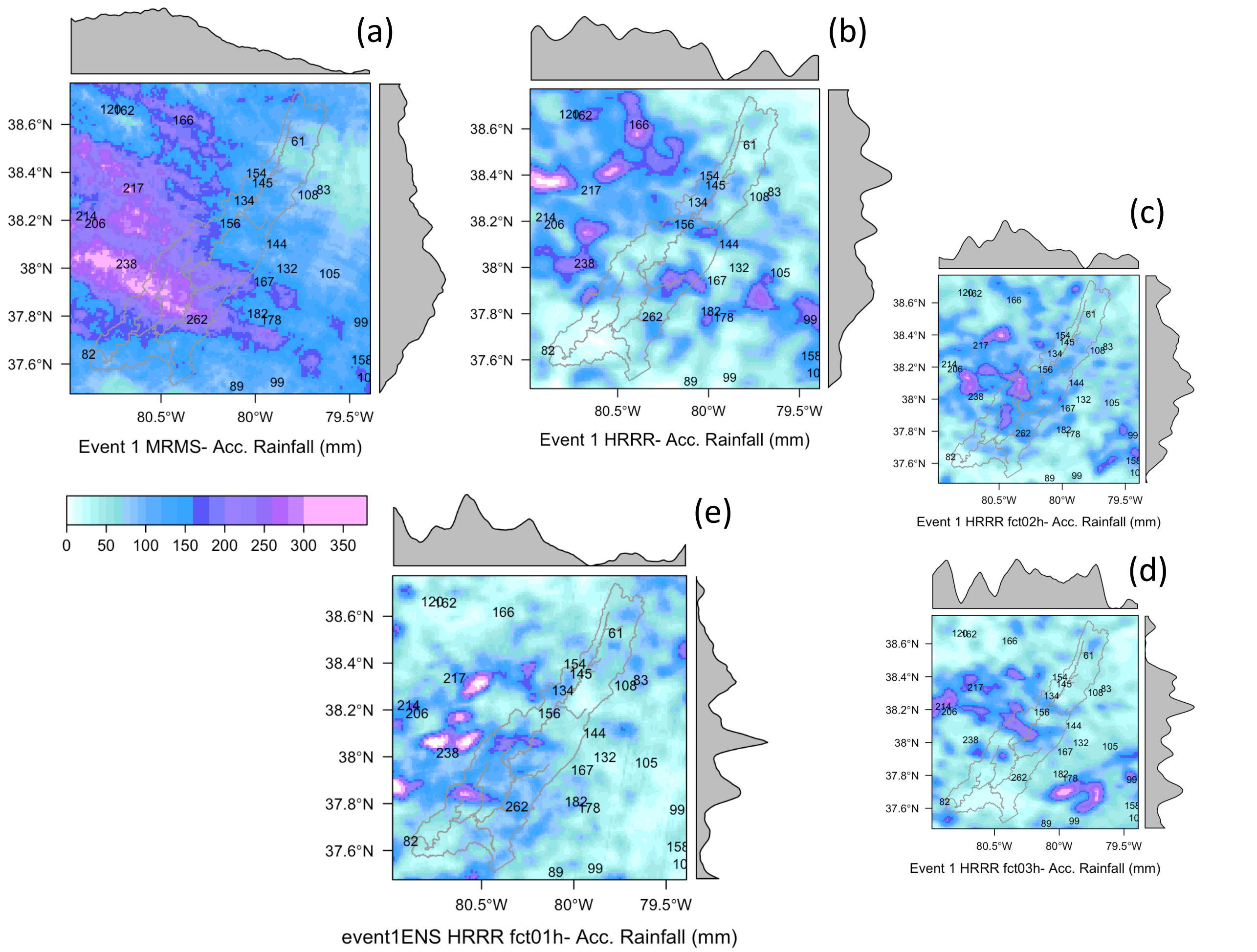




**Figure 5.** Total accumulated rainfall for the first extreme event (06/2016) using (a) MRMS QPE, (b) first hour, (c) second hour, (d) third hour QPF.(e) represents the total accumulated rainfall for first hour PPF.

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023

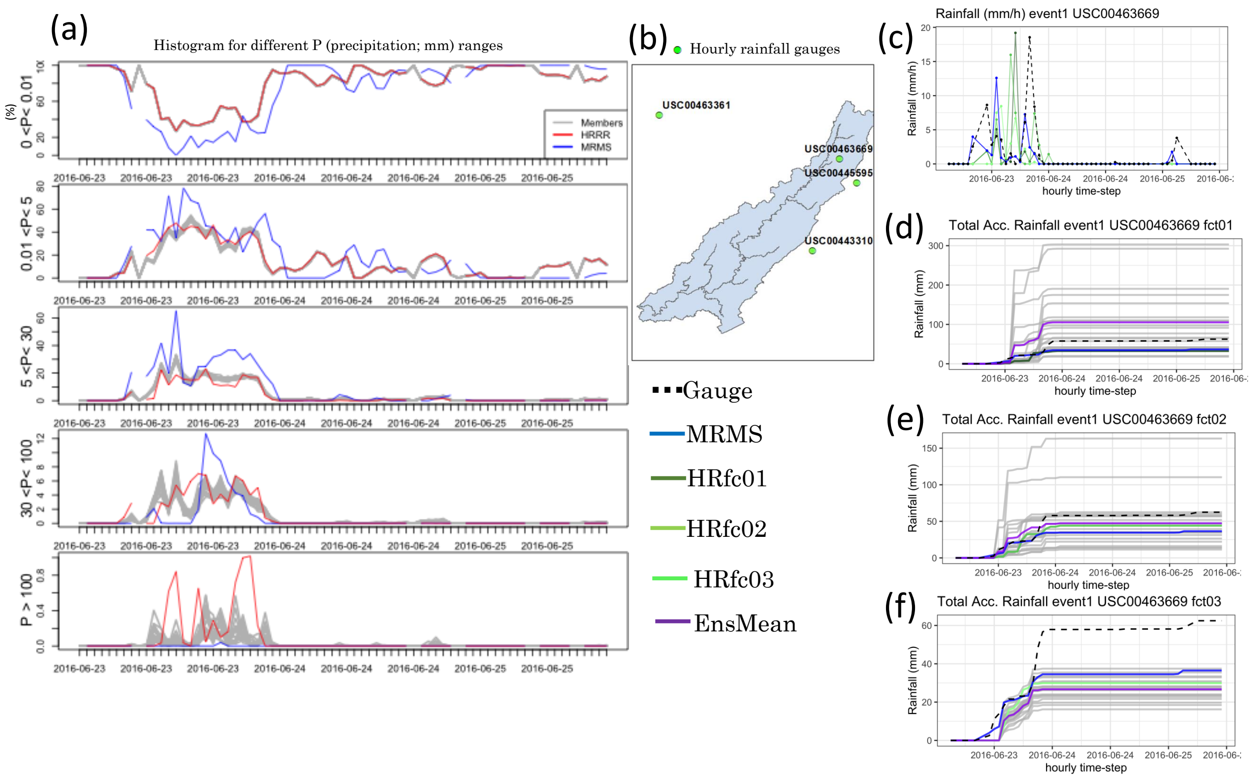



**Figure 6.** (a) The percentage of pixels for different rainfall thresholds over time considering the first-hour PPF. (b) Hourly rain gauge over the domain. (c) Rainfall rate time-series for USC00463669 rain gauge location. Total accumulated rainfall PPF time-series for the first hour (d), second-hour (e), third-hour forecast (f) rainfall for first hour PPF at USC00463669 rain gauge location

Table 3 represents the statistics that evaluate the accuracy of the ensemble mean. For extreme event 1 (06/2016), the Pearson correlation of the PFF ensemble mean slightly increases while the RMSE values decrease for the first-hour forecast. However, the second-hour PPF showed a deterioration in statistics compared to the deterministic QPF. The ensemble mean Pearson Correlation was significantly lower and the RMSE showed an increase. However, considering the 6 events, the overall bias was lower for the second-hour forecast.

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


**Table 3.** Pearson Correlation, RMSE and MBIAS for HRRR PPF. Statistics considered 4 points of hourly gauge measurements as observed rainfall over the domain. The last row represents the mean value over the six events.

| | Person Cor. | | | RMSE | | | MBIAS | | |
|---|---|---|---|---|---|---|---|---|---|
| | fct 01 | fct 02 | fct 03 | fct 01 | fct 02 | fct03 | fct 01 | fct 02 | fct 03 |
| Event 1 | 0.33 | 0.04 | 0.06 | 4.64 | 8.69 | 4.27 | 0.93 | 1.04 | 0.53 |
| Event 2 | 0.50 | 0.10 | 0.27 | 2.85 | 8.46 | 4.30 | 0.63 | 1.13 | 0.91 |
| Event 3 | 0.35 | 0.59 | 0.38 | 1.70 | 1.43 | 1.83 | 0.60 | 0.66 | 0.76 |
| Event 4 | 0.61 | 0.64 | 0.54 | 1.03 | 1.00 | 1.08 | 0.68 | 0.67 | 0.62 |
| Event 5 | 0.39 | 0.41 | 0.56 | 1.29 | 1.27 | 1.36 | 0.72 | 0.62 | 0.83 |
| Event 6 | 0.77 | 0.49 | 0.51 | 0.79 | 1.32 | 1.32 | 1.33 | 1.67 | 1.85 |
| All Events | 0.38 | 0.11 | 0.25 | 2.41 | 4.99 | 2.70 | 0.79 | 0.98 | 0.85 |

### 3.3 Does the hydrological calibration affect the streamflow simulations along the main river?

This section presents the results from the WRF-Hydro parameter calibration. As mentioned before, we performed the 16 Noah-MP parameter calibration considering the range provided by the operational version 2.0 of the National Water Model. Since we decided to use the gridded channel network with a diffusive wave routing scheme, we did not consider the NWM routing parameters which are related to a vector-based channel network applied to a kinematic wave approximation. The ranges of parameters can be found in Table 2S of the supplementary material. We calibrated the WRF-Hydro experiment considering the most downstream USGS streamflow gauge available in Greenbrier River (Hilldale - 4217 $km^2$ of the drainage area). The hydrograph resulting from this calibration can be found in the bottom row of Figure 7. The downstream region was fairly well calibrated providing an NSE of 0.75 and 0.79 on hourly and daily time scales, respectively.Nevertheless, the Pearson Correlation (0.94) and RMSE (31 m3/s) reached the best optimum configuration at monthly time scales. The overall multiplicative bias was approximately close to -28 m3/s, which reveals a systematic underestimation of streamflow for hourly, daily, and monthly time scales. The top row of Figure 7 shows the effect of downstream calibration in the the upstream region of Greenbrier River at Durbin (346 $km^2$). As we can observe, the NSE values had a considerable decrease to 0.43 and 0.44 for hourly and daily time scales. Comparing both flow duration curves in Figure 7 we show that the upstream region showed poorer agreement for moderate streamflow values. Although the overall bias is around -4.3 m3/s the underestimation is relative to the streamflow magnitude observed in the smaller watershed. Despite the underestimation happening in the upper and lower regions of Greenbrier River watershed, the additional comparison with the NWM version 2.0 components (Figure 6S) showed that this current configuration provided higher values of subsurface and channel runoff . The results could relate to a better ability to estimate the flood peaks, considering the system without real-time data assimilation. We also did not perform a separate streamflow validation against observed gauges for any other period than in calibration (2015-2020), instead we focused to evaluate the model outputs for different locations along the Greenbrier river. In the supplementary material (Figure 6Sb) we

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


show the simulated hydropgrahs using the NWM version 2.0. Our calibration mainly improved the streamflow simulations at Durbin station, decreasing the overestimation of discharge showed by the NWM version 2.0 parameters.

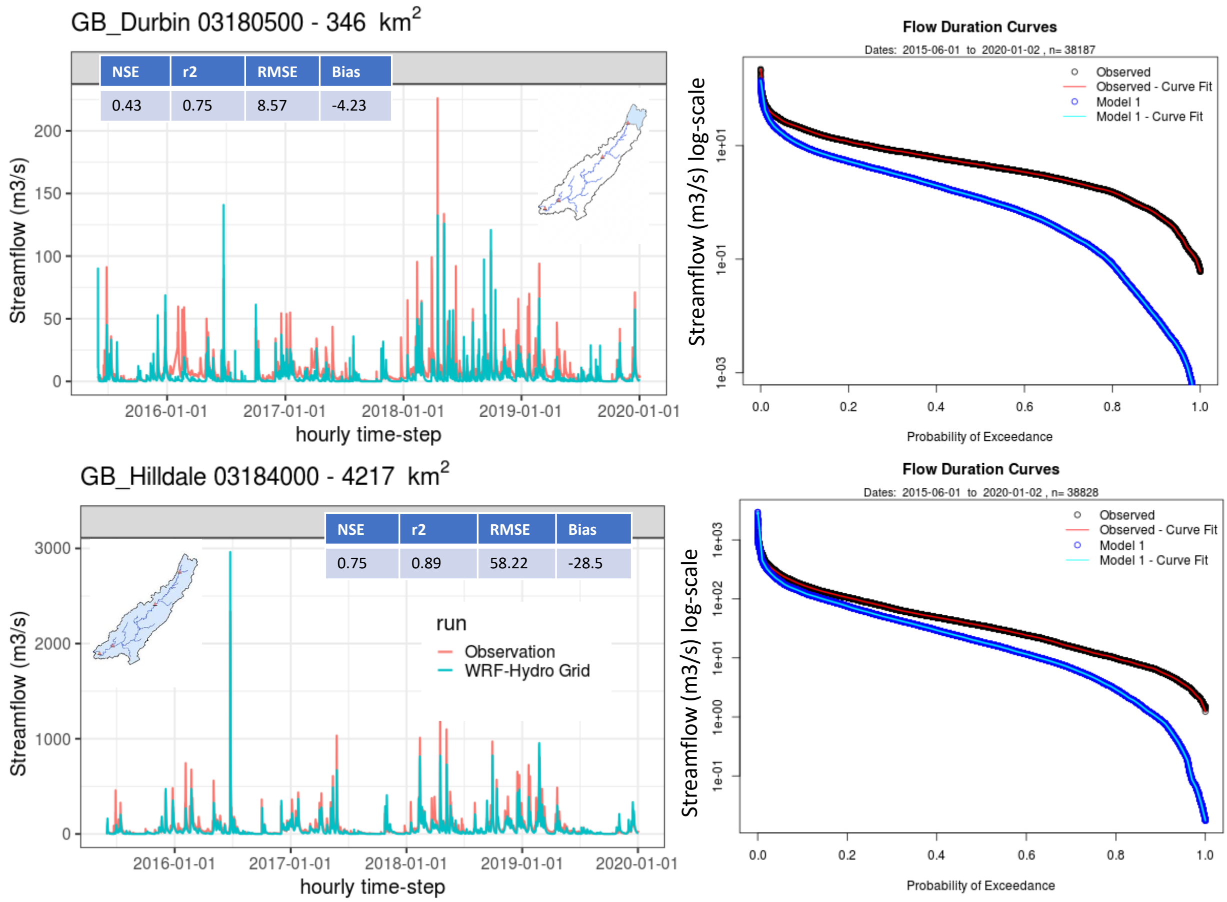



**Figure 7.** WRF-Hydro streamflow calibration results (lower row) for Hilldale (4217 $km^2$). Evaluation of calibration results for upstream Greenbrier (upper row) in Durbin (346 $km^2$). The statistics were calculated using the package *rwrfhydro*.

### 3.4 Can the PSF increase the accuracy of streamflow predictions?

After the WRF-Hydro calibration, the model's ability to simulate the streamflow in a real-time forecasting perspective using the rainfall ensembles from PPF at lead times up to 3 hours was evaluated. In particular, the streamflow ensemble PSF was statistically evaluated for the 2016 flood event. Figure 8 shows the hydrographs at Buckeye (1398 $km^2$) and Alderson (3531 $km^2$) USGS stations. These monitoring stations have the advantage to provide not only discharge and water level series but also water level early warning thresholds based on a constructed rating curve (i.e. relationship between discharge and local water level time series).

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


The red lines in Figure 8 are discharge rates associated with the flood warning stage, while the yellow line is associated with the action stage, which according to the USGS local stakeholders would call for evacuation precautions. The observed hydrographs showed that the flood wave was possibly under the flood threshold upstream of the Buckeye station. Most of the water overflowed to floodplain areas between Buckeye and Alderson stations. All modeled scenarios predicted the significant floodplain overflow at Alderson. However, the QPE simulation overestimated the flood peak, which was expected since during the calibration process this flood event was also overestimated. Both rainfall inputs from MRMS and the first hour of HRRR QPF anticipated the rising of the hydrograph for the downstream area in Alderson. However, the QPE simulation lead to a good agreement for the flood peak at the upstream area in Buckeye. This reveals a certain level of uncertainty to simulate extreme flood events even when they were part of the calibration process that could be attributed to hydrologic heterogeneity. As shown before in Figure 5, according to the QPE the most intense rainfall areas were placed between Buckeye and Alderson stations. In this case, we could possibly suggest that the hydrological model did not predict well the runoff response due to high-intensity rainfall in a short period of time. Table 5 also demonstrates that the QPE-driven hydrological simulations showed higher NSE and Pearson Correlation for Buckeye in comparison with Alderson. Both station locations presented an overall overestimation of hydrograph, mostly due to quick recession, which could be related to the routing parameter calibration or the simulation of interflow.

As discussed in the previous session and shown in Figure 6 (c) some of QPF fields presented a time delay to predict the most intense rainfall rates in comparison to QPE. However, the precipitation delay helped the hydrological model to adjust the rising limb time in hydrographs simulated with QPF. Table 5 and Figure 8 (f) showed that the second and third HRRR forecasts in fact outperformed the QPE for Alderson and Hilldale stations with higher Pearson correlation and NSE. At the same time, the first-hour QPF improved the streamflow prediction in Durbin, a location where the third-hour forecast showed a higher overestimation of discharge. Considering the mean statistic values for all stations presented in Table 5, the second hour QPF showed to be more accurate to predict the streamflow in the current WRF-Hydro configuration.

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


**Table 4.** Pearson Correlation, NSE and MBIAS between observed streamflow and WRF-Hydro predictions during the extreme flood event in 06-2016, considering rainfall inputs from MRMS QPE and deterministic HRRR QPF for the first, second-, and third-hour forecast.

| | | Durbin | Buckeye | Alderson | Hilldale | All Stations |
|---|---|---|---|---|---|---|
| Pearson Cor. | MRMS | 0.45 | 0.93 | 0.77 | 0.70 | 0.71 |
| | $HRRR_{fct01}$ | 0.76 | 0.81 | 0.63 | 0.53 | 0.68 |
| | $HRRR_{fct02}$ | 0.56 | 0.60 | 0.86 | 0.79 | 0.70 |
| | $HRRR_{fct03}$ | 0.63 | 0.85 | 0.95 | 0.94 | 0.84 |
| NSE | MRMS | 0.19 | 0.82 | 0.59 | 0.48 | 0.52 |
| | $HRRR_{fct01}$ | 0.48 | 0.57 | 0.38 | 0.26 | 0.42 |
| | $HRRR_{fct02}$ | 0.17 | 0.34 | 0.72 | 0.61 | 0.46 |
| | $HRRR_{fct03}$ | 0.06 | 0.50 | 0.90 | 0.88 | 0.58 |
| MBIAS | MRMS | 0.98 | 0.77 | 0.94 | 1.1 | 0.92 |
| | $HRRR_{fct01}$ | 0.79 | 1.18 | 0.97 | 1.08 | 1 |
| | $HRRR_{fct02}$ | 1.38 | 1 | 0.81 | 0.86 | 1.01 |
| | $HRRR_{fct03}$ | 3.69 | 1.56 | 0.89 | 0.92 | 1.77 |

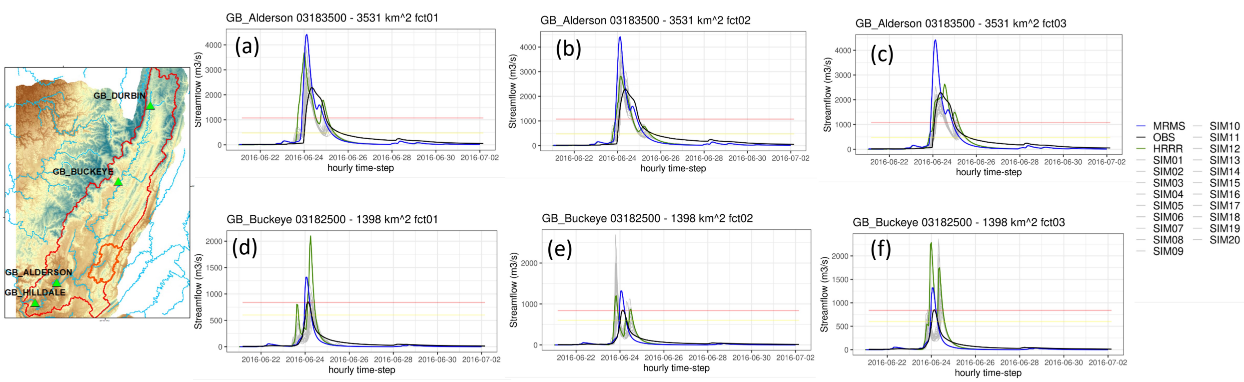



**Figure 8.** QPE, QPF and PPF driven hydrographs for the 2016 flood event at (a) upstream area in Buckeye and (b) downstream area in Alderson. Red lines represent the discharge stage threshold for flood (i.e. channel overflow) calculated by a rating curve. Yellow lines represent the action stage defined by local stakeholders.

Table 6 and 7 show the statistical results obtained with PSF considering different rainfall members from PPF for the first-, second-, and third-hour forecast. The ensemble mean of PSF was compared against observed streamflow data to compute the Pearson Correlation, NSE and MBIAS presented at Table 6. We noticed a considerable improvement in the first-hour

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


**Table 5.** Pearson Correlation, NSE and MBIAS between observed streamflow and PSF ensemble mean during the extreme flood event in 06-2016, considering rainfall inputs from PPF for the first-, second-, and third-hour forecast.

| | | Durbin | Buckeye | Alderson | Hilldale | All Stations |
|---|---|---|---|---|---|---|
| Pearson Cor. | $Ens_{fct01}$ | 0.77 | 0.84 | 0.82 | 0.74 | 0.79 |
| | $Ens_{fct02}$ | 0.65 | 0.58 | 0.90 | 0.84 | 0.74 |
| | $Ens_{fct03}$ | 0.54 | 0.68 | 0.95 | 0.92 | 0.77 |
| NSE | $Ens_{fct01}$ | 0.49 | 0.53 | 0.48 | 0.35 | 0.46 |
| | $Ens_{fct02}$ | 0.29 | 0.26 | 0.77 | 0.65 | 0.5 |
| | $Ens_{fct03}$ | 0.1 | 0.42 | 0.57 | 0.54 | 0.41 |
| MBIAS | $Ens_{fct01}$ | 0.84 | 0.66 | 0.64 | 0.71 | 0.71 |
| | $Ens_{fct02}$ | 1.07 | 0.84 | 0.71 | 0.76 | 0.84 |
| | $Ens_{fct03}$ | 2.13 | 0.84 | 0.60 | 0.65 | 1.05 |

and second-hour streamflow predictions, with increasing of Pearson correlation and NSE for the downstream areas next to Alderson and Hilldale stations. The predictions for Greenbrier upstream area in Buckeye suffered a slight decrease in NSE, but the overall mean accuracy still conserved the improvement along the river regarding the 4-gauge locations. There was a decrease in performance for the third-hour streamflow forecast. The third-hour PSF ensemble mean showed lower Pearson correlation and NSE values compared to the streamflow simulations driven by the third-hour deterministic QPF. However, the PSF helped to decrease the overestimation (i.e., lower percentual bias) at upstream areas in Durbin and Buckeye.

### 3.5 How is the predictability of flood extension related to terrain description?

The results presented in this section will highlight the influence of the topographic model to map flood areas from a probabilistic perspective. After the extreme flood event in June of 2016, USGS specialists benchmarked in the field the maximum flood extension around the Howard Creek floodplain. We used the observed USGS flood map scenario as the target for our flood map predictions. Based on the previous results evaluating the streamflow ensemble prediction, we decided to elaborate the probabilistic flood map forecasts (PFF) based on the first hour PSF members only. Figure 9 (a) shows the 20 PSF members at the output of Howard Creek (248 $km^2$ of drainage area – Figure 9 (b)). The streamflow spread among the PSF members was higher for the middle flood peak in comparison to the two other smaller discharge waves observed in the Howard Creek hydrograph. Unfortunately, there was no USGS streamflow gauge monitoring to evaluate the reliability of those predictions in the small watershed during the flood event. However, the PSF results in the previous session demonstrated that the discharge simulations downstream fo Alderson had a satisfactory level of accuracy (0.77 of Pearson Correlation and 0.59 of NSE) under the MRMS QPE.

The flood extension was estimated by a combination of the WRF-Hydro water level outputs and the HAND methodology, to describe the floodplain topography along Howard Creek. The first approach was to derive the HAND map using the NHDPlus DEM in 10 meters of spatial resolution (Figure 9(b)). The water level predictions at each WRF-Hydro 250 meters grid cell were overlapped by the Hand map considering the regions mapped as flooded by USGS. Figure 9 (c) shows that the flood delineation driven by the MRMS QPE (blue contours) overestimated the USGS benchmarks (black solid) flood extent. Moreover, the statistics in Figure 9 (d) showed a positive bias also when the forecast was driven by the first hour HRRR QPF and the PSF ensemble mean. The black ranges on top of the ensemble mean metrics indicate the upper and lower limits of the 20 PFF members, which enclose the score metrics found in the flood extension simulation driven by the MRMS QPE. The PFF ensemble means reduced the false alarm ratio (FAR) but also reduced the POD, which translated in a lower CSI in comparison with the MRMS. However, we observed that the CSI from the PFF ensemble mean was slightly higher than the CSI found by the HRRR first hour forecast QPF, which means that the overall ensemble mean also helped to slightly decrease the false alarm ratio in mapping the flood extension.

When we applied the WRF-Hydro water level outputs to the HAND map generated by the NED Lidar DEM in 1 meter of spatial resolution, there is a depreciation in the flood extension score metrics (Figure 9 (f)). For instance, the CSI score was considerably lower for the flood extension delineated in a 1-meter spatial resolution under the MRMS QPE, compared to the flood mapping obtained in 10 meters. The overall POD spread of the 20 members also increased with a higher spatial resolution of the floodplain, showing that the level of uncertainty increased when we mapped with the lidar DEM. However, the PFF members once again helped the POD and CSI scores to match the ones found in the simulation using MRMS QPE.

The overprediction was higher along the larger floodplain areas. Due to the lack of streamflow observations at Howard Creek, we could not validate the diffusive wave accuracy to predict the water levels used to estimate the flood extent. However, the streamflow evaluation showed that the MRMS QPE simulation was overpredicting the streamflow in Alderson, and the PFF helped decrease the overestimation of the peak flow across the basin. We conclude that this also helped to decrease the overestimation in Howard flood event.

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023

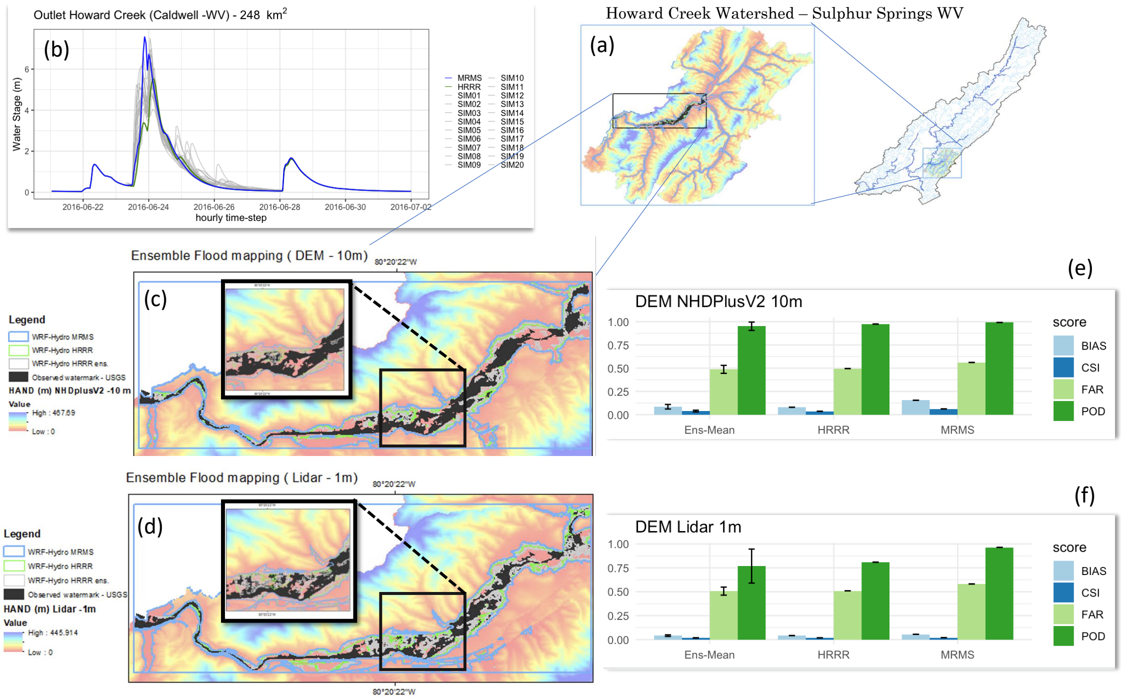




**Figure 9.** Flash flood probabilistic mapping for White Sulphur Springs (b) considering the stream flow members for 1 hour forecast at (a) Howard Creek. (c) shows the statistics for the forecasted maps considering the HAND-DEM at 10 m of spatial resolution (d). (e) present the maps and (f) statistics for the flood mapping considering the HAND-DEM at 1 m of spatial resolution.

## 4 Discussion

### 4.1 QPE

Identifying spatial uncertainties in real-time rainfall estimation products is relevant for any weather-related forecasting system. After a hydrometeorological disaster, the nature of storm severity is usually the number one fact discussed and pointed out as the main contributor to triggering the hazard. Martinaitis et al. (2020) discussed how the MRMS product was an important guideline to stakeholders in the operational environment to decide the time and level of flood warning during the 2016 flash flood event in West Virginia. Operational hydrological products available at the time contributed to early warnings issued between 0-24h lead-time and some relied on MRMS as a QPE, as initial conditions, or during the calibration process. However, our local analysis through the MRMS QPE product identified an overall underestimation of the rainfall during the fall and summer periods, which is apparent in the upstream region of the Greenbrier River basin area during the winter and spring months. Therefore, the systematic negative bias in upstream areas could have influenced capturing the real rainfall rate during the June 2016 flood event. A technical description of the MRMS retrieval algorithm shows that the final QPE could be gauge-corrected with hourly accumulations if in-situ observations are available (Zhang et al. (2016)). However, when we compared the MRMS rainfall diurnal means against the only available rain gauge inside the basin, we did not find evidence that it was in fact being used to correct the QPE during the 2016 June event. The difference in total accumulated rainfall between the USC00463669 rain gauge and the MRMS pixel was around 20 mm at the end of the flood event. As discussed by Gerard et al. (2021), the MRMS QPE tends to underestimate the rainfall intensity associated with deep convection storms during the warm season due to systematic overshooting artifacts in radar operations, which could be a case of the storm event in June of 2016. This systematic underestimation could be also related to the epistemic uncertainties of the weather radar to properly scan the storm and reflect the radio waves in mountainous regions, not only in the algorithm retrieval. Basins with scarce rain gauges could also benefit from QPE correction techniques derived from hydrological models, as shown by Liao and Barros (2022).

### 4.2 QPF

Martinaitis et al. (2020) also demonstrated that experimental deterministic QPFs (NAMRR (Carley et al. (2015)) and ESRL HRRRv2 (Benjamin et al. (2016))) were more accurate to capture the total accumulated rainfall than the operational HRRRv2 for the 2016 June event. In our evaluation focused on the Greenbrier River basin, the HRRRv2 first-hour QPF underestimated the rainfall around the White Sulphur Springs neighborhood, one of the regions most affected by flash flood events. The correction based on the MRMS QPE multiplicative bias was demonstrated through the ensemble means to relocate the maximum rainfall coverage more accurately. We noticed an improvement in both rainfall timing (by the Pearson correlation) and intensity (RMSE) mainly for the first-hour forecast. The ensemble PFF mean generated by the second and third HRRRv2 hour forecast maintained the percentual bias closer to 1, showing a slight overestimation. Experimental QPF ensembles were also evaluated (Martinaitis et al. (2020)). Previous work showed that for 5 experimental QPF ensemble products evaluated, the High-Resolution Ensemble Forecast System HREFX (Jirak et al. (2018)) was the most reliable to predict the spatial-temporal pattern of rainfall during the 2016 event in West Virginia. HREFX constructs the rainfall members based on a variation of multiple cycles of convective parametrizations, which could be considered more computationally expensive if applied operationally. The experimental HRRR-TLE (Alexander et al. (2010)) is a time-lagged ensemble, which means that the previous deterministic initialization of HRRRv2 was used to generate the members as a post-processing tool. According to Martinaitis et al. (2020), HRRR-TLE performed well since was able to predict a 20% probability chance to exceed 50.8 mm (2 in) of rainfall between 1800 UTC 23 Jun and 0000 24 Jun, which translated to a 10% change of flash food event. Our studies also demonstrated that a rainfall ensemble of previous initializations of HRRRv2 could still be valuable for hydrological prediction. We acknowledge that our QPF and PFF evaluation considered only 4 automatic rain gauges over our study domain, and object-based verification tools based on QPE, as MET's MODE (discussed by Viterbo et al. (2020)), may be more appropriate to track the evolution of isolated storms.

### 4.3 Hydrological model calibration and PSF

The WRF-Hydro hydrological model was calibrated using streamflow observations as well as considering the initial Noah-MP parameters closer to the operational version of the NWM. The set of parameters found for the catchment upstream of Hilldale station (4217 $km^2$) overestimated the streamflow for the 2016 flood event but underestimated the peak flows for the other

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


extreme event analyzed. We considered all the events we intended to forecast as part of the calibration period. Considering or not the most extreme events during the period of long-term calibration are an ongoing discussion in hydrological community. SIN (2012) developed a methodology to select usual events in a streamflow time series and compared the calibration process as well as transferability of parameters with the calibration being performed by the full time-series length. In their study, the authors concluded that the extreme event-based calibration was slightly worse than the full time-series length. However, if the events were randomly selected (without considering the most extreme events in the series) the calibration performed even worse. They discussed that methodology could be interesting to evaluate when a recalibration of operational hydrological models is necessary since not all new extreme events would influence how the state of hydrological processes is being modeled. We indeed noticed that in our case the overestimation of QPF beneficiated the streamflow forecasting. In the specific case of the 2016 flood event, the delay in rainfall timing from the third our forecast of HRRRv2 also helped to accurately forecast the rising limb o fteh flood hydrograph at Alderson (3531 $km^2$). Regarding the ability to simulate an accurate forecast for the June 2016 flood at different points along the Greenbrier River, the PFF ensemble mean increased the overall statistical performance when compared to the hydrological simulations by deterministic first-hour HRRRv2 QPF. The relevant effect was most rainfall members that diminished the original QPF rainfall intensity at headwaters, a region that was underestimated by the MRMS QPE, used as a reference to generate the members. Falck et al. (2021) showed that post-processing PFF techniques driven by NWP and rain gauges were mainly helpful to increase the streamflow prediction accuracy at watersheds smaller than 25 000 $km^2$ of drainage area, but the prediction evaluation range varied between 24h to 264h lead time. In the context of the short-term forecast, a radar-based PFF technique explored by Caseri et al. (2016) was able to increase the overall reliability of streamflow prediction in small basins around 250 $km^2$, up to 2 hours forecast lead-time compared to a persistence nowcasting method. In both cases, the streamflow prediction results were highly influenced by the rainfall reference used to generate the members, if it's either the space distribution of rain gauges or the weather radar-based QPE accuracy. It is important to note that in many regions of complex terrain there are no rain gauges or reliable radar products. AI-based methods to take advantage of the climatology of storm propagation (Kuligowski and Barros, 1998a, b; Kim and Barros, 2001) can extend the QPE lead time with success.

### 4.4 Flood ensemble mapping

The ability to translate the QPF uncertainties to real-time flood mapping also depends on the flood mapping methodology. We observed a higher flood extension ensemble spread (i.e. difference in flood scenarios) when we increased the resolution of the flood mapping areas by combining the LIDAR DEM with the HAND methodology. In other words, the resulting high false alarm ratio revealed that the overprediction of flood extent increases when the DEM resolution increases. Since our streamflow simulations were made using the WRF-Hydro gridded version, there was no difference in hydraulic properties (i.e. roughness parameter or river geometry) between the channel and floodplain inside the hydrological model. In this context, we preserved the original water level WRF-Hydro output calculated following the conservation of momentum and mass to intersect the HAND floodplain, unlike the methodology of synthetic rating curves in the NWM (Zheng et al. (2018)). Due to the epistemic limitations of Manning's equation to create those synthetic rating curves, Godbout et al. (2019) concluded that

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


HAND mapping tends to perform poorly at short river reaches with extreme slope values. In such cases, they suggested a more equally spaced streamflow input along the main river to increase the flood extent mapping performance. Therefore, applying the HAND methodology using the WRF-Hydro gridded version could be an alternative to finding the optimum reach length for mountainous regions prone to flash floods, such as the Greenbrier River. Due to the lack of streamflow observations at Howard Creek, we could not validate the accuracy of the diffusive wave approach to predict the water levels used to estimate the flood extent. However, the real-time PFF helped to decrease the peak flow compared to the offline simulation using the MRMS QPE based on the streamflow evaluation, which lead to higher POD and lower FAR.

## 5 Conclusions

This work intended to analyze uncertainties to forecast hydrometeorological components in a framework considered state-of-the-art in flash flood warning systems. We set up a physically based fully distributed hydrological-hydraulic model as the core component of the forecast chain. We conducted a comprehensive analysis of the uncertainties associated with different components of the forecast chain. We quantified the contributions of meteorological uncertainty, model structure, parameter uncertainty, and data input uncertainty to the overall forecast uncertainty.

By conducting our own experiment, we chose some extreme rainfall events in the Greenbrier River basin to create PSF and particularly took a closer look at the flash flood event in June of 2016 to generate PFF. The starting point of our analysis was the evaluation of the operational MRMS QPE. We noticed a systematic underestimation of the rainfall on a daily scale mostly during the summer and fall months. The high elevation with steep valleys could have some influence on weather radar scanning of severe storms especially in the upstream area of the basin. At hourly scales, the scarcity of rain gauges made the results not very conclusive. We indeed noticed that for the 2016 June flood event, the hourly rain gauges were probably not entirely presented in the retrieval algorithm to bias correct the MRMS QPE product in real-time. Under the epistemic uncertainties in weather radar scanning, the accuracy of hourly QPE could be majorly improved if more rain gauges were installed over the Greenbrier River basin. Because this is not likely to occur and it is not generalized, alternative techniques such as AI-based storm propagation can be helpful in improving QPE at a basin scale.

We demonstrated that a simple geostatistical methodology to generate rainfall ensembles using the HRRRv2 and MRMS QPE could be effective to increase the overall reliability of the first-hour forecast. More sophisticated methodologies which incorporate storm tracking had been shown to increase reliability by up to 3 hours of lead time and longer. In such a case, we did not include patterns of previous initializations of HRRRv2, which could have been important to diminish the temporal bias of each forecast to each initialization of SGD methodology.

The WRF-Hydro parameter calibration was satisfactory to simulate the hourly streamflow during The WRF-Hydro parameter calibration was satisfactory to simulate the hourly streamflow during extreme flood events. When we introduced the rainfall ensemble members, it helped to adjust the streamflow prediction at upstream areas of the basin. The hydrological model also benefited from the second and third hours PPF forecast delays in predicting the most intense rainfall. The results show that the forecast skill along the Greenbrier River network varies significantly from the skill at the outlet that was used as the forecast

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


point for calibration. This highlights the hydrologic and hydraulic heterogeneity across the basin and the complexity of the coupled hydrologic-hydraulic networks in complex terrain. Therefore, to go beyond the point of flood forecast operations towards effective flood mapping, the calibration methodology for process-based models requires accounting for the scale-dependent behavior of physical parameterizations and watershed physiographic heterogeneity, even at very high spatial resolution. We did not test data assimilation techniques, which could significantly improve the initial conditions prior to the streamflow prediction. In ensemble forecast scenarios, a combination of real-time streamflow data assimilation would possibly decrease the PPF bias, especially in the timing of flood waves.

The forecasted flash flood maps were considerably overpredicted at regions with longer floodplain banks. However, the regularly spaced distributed streamflow prediction members created more dispersed flood scenarios when we increased the DEM resolution (i.e. the floodplain detail). Following the previous literature in flood mapping using the HAND methodology, hydraulic properties (i.e. roughness coefficient, channel slope) used in the diffusive wave formulation were more sensitive to the changes in height description by the DEM. For that reason, simply updating the terrain description in the final step of a flood forecast chain cannot guarantee a higher accuracy to simulate flood extension. Identifying those sources of uncertainty that have the most significant impact on flood predictions is important to continuously keep strategies to mitigate or reduce near-real-time flood mapping uncertainties.

*Author contributions.*

LB - Conceptualized the research, including data curation, methodology, modeling application and development, formal analysis, figures, and writing the original draft. AN - Participated in formal analysis of calibration methodology and editing of the original draft. NC - edited the original draft. AB - Supervised the research, suggested formal analysis and edited the original draft.

*Competing interests.*

The authors declare no competing interests at present.

*Acknowledgements.* This study was partially funded by a doctorate scholarship from CAPES Brazil - project code 8 8 8 8 1 . 1 7 4 6 5 5 / 2 0 1 8 - 0 1 D O C - P L E N O

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


## References

Calibration of hydrological models on hydrologically unusual events, Advances in Water Resources, 38, 81–91, https://doi.org/https://doi.org/10.1016/j.advwatres.2011.12.006, 2012.

Alexander, C. S., Weygandt, S., Benjamin, S. G., G., T., and Smirnova, J. M. e. a.: Recent and future enhancements ,time-lagged ensembling, and 2010 forecast evaluation, 24th Conf. on Weather and Forecasting/20th Conf. on Numerical Weather Prediction, Seattle, WA, Amer, 2010.

Arulraj, M. and Barros, A. P.: Improving quantitative precipitation estimates in mountainous regions by modelling low-level seeder-feeder interactions constrained by Global Precipitation Measurement Dual-frequency Precipitation Radar measurements, Remote Sensing of Environment, 231, 111 213, https://doi.org/https://doi.org/10.1016/j.rse.2019.111213, 2019.

Arulraj, M. and Barros, A. P.: Automatic detection and classification of low-level orographic precipitation processes from space-borne radars using machine learning, Remote Sensing of Environment, 257, 112 355, https://doi.org/https://doi.org/10.1016/j.rse.2021.112355, 2021.

Bai, T. and Tahmasebi, P.: Sequential Gaussian simulation for geosystems modeling: A machine learning approach, Geoscience Frontiers, 13, 101 258, https://doi.org/https://doi.org/10.1016/j.gsf.2021.101258, 2022.

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., Olson, J. B., James, E. P., Dowell, D. C., Grell, G. A., Lin, H., Peckham, S. E., Smith, T. L., Moninger, W. R., Kenyon, J. S., and Manikin, G. S.: A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh, Monthly Weather Review, 144, 1669 – 1694, https://doi.org/https://doi.org/10.1175/MWR-D-15-0242.1, 2016.

Beven, K.: Distributed Models and Uncertainty in Flood Risk Management, chap. 14, pp. 289–312, John Wiley Sons, Ltd, https://doi.org/https://doi.org/10.1002/9781444324846.ch14, 2010.

Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrological Processes, 6, 279–298, https://doi.org/https://doi.org/10.1002/hyp.3360060305, 1992.

Bocanegra, R. A. and Francés, F.: Assessing the risk of vehicle instability due to flooding, Journal of Flood Risk Management, 14, e12 738, https://doi.org/https://doi.org/10.1111/jfr3.12738, 2021.

Braud, I., Vincendon, B., Anquetin, S., Ducrocq, V., and Creutin, J.-D.: 3 - The Challenges of Flash Flood Forecasting, in: Mobility in the Face of Extreme Hydrometeorological Events 1, edited by Lutoff, C. and Durand, S., pp. 63–88, Elsevier, https://doi.org/https://doi.org/10.1016/B978-1-78548-289-2.50003-3, 2018.

Carley, J., E. Rogers, B. S., Ferrier, E. Aligo, W. S., Wu, S., and et al.: Ongoing Development of the Hourly-Updated Version of the NAM Forecast System, 27th Conference On Weather Analysis And Forecasting/23rd Conference On Numerical Weather Prediction, https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273567.html, 2015.

Caseri, A., Javelle, P., Ramos, M., and Leblois, E.: Generating precipitation ensembles for flood alert and risk management, Journal of Flood Risk Management, 9, 402–415, https://doi.org/https://doi.org/10.1111/jfr3.12203, 2016.

Cecinati, F., Rico-Ramirez, M. A., Heuvelink, G. B., and Han, D.: Representing radar rainfall uncertainty with ensembles based on a time-variant geostatistical error modelling approach, Journal of Hydrology, 548, 391–405, https://doi.org/https://doi.org/10.1016/j.jhydrol.2017.02.053, 2017.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resources Research, 44, https://doi.org/https://doi.org/10.1029/2007WR006735, 2008.

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., Pauwels, V. R. N., Cai, X., Wood, A. W., and Peters-Lidard, C. D.: The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism, Hydrology and Earth System Sciences, 21, 3427–3440, https://doi.org/10.5194/hess-21-3427-2017, 2017.

Clark, R. A., Gourley, J. J., Flamig, Z. L., Hong, Y., and Clark, E.: CONUS-Wide Evaluation of National Weather Service Flash Flood Guidance Products, Weather and Forecasting, 29, 377 – 392, https://doi.org/https://doi.org/10.1175/WAF-D-12-00124.1, 2014.

Cloke, H. and Pappenberger, F.: Ensemble flood forecasting: A review, Journal of Hydrology, 375, 613–626, https://doi.org/https://doi.org/10.1016/j.jhydrol.2009.06.005, 2009.

Crochemore, L., Ramos, M.-H., and Pappenberger, F.: Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts, Hydrology and Earth System Sciences, 20, 3601–3618, https://doi.org/10.5194/hess-20-3601-2016, 2016.

Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., Attinger, S., and Thober, S.: The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model, Journal of Geophysical Research: Atmospheres, 121, 10,676–10,700, https://doi.org/https://doi.org/10.1002/2016JD025097, 2016.

Di Baldassarre, G., Nardi, F., Annis, A., Odongo, V., Rusca, M., and Grimaldi, S.: Brief communication: Comparing hydrological and hydrogeomorphic paradigms for global flood hazard mapping, Natural Hazards and Earth System Sciences, 20, 1415–1419, https://doi.org/10.5194/nhess-20-1415-2020, 2020.

Ding, Y., Liu, X., and Chen, R.: Numerical Simulation of Alpine Flash Flood Flow and Sedimentation in Gullies With Large Gradient Variations, Frontiers in Environmental Science, 10, https://doi.org/10.3389/fenvs.2022.858692, 2022.

Dowell, D. C., Alexander, C. R., James, E. P., Weygandt, S. S., Benjamin, S. G., Manikin, G. S., Blake, B. T., Brown, J. M., Olson, J. B., Hu, M., Smirnova, T. G., Ladwig, T., Kenyon, J. S., Ahmadov, R., Turner, D. D., Duda, J. D., and Alcott, T. I.: The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part I: Motivation and System Description, Weather and Forecasting, 37, 1371 – 1395, https://doi.org/https://doi.org/10.1175/WAF-D-21-0151.1, 2022.

Evers, M., Jonoski, A., Maksimovič, v., Lange, L., Ochoa Rodriguez, S., Teklesadik, A., Cortes Arevalo, J., Almoradie, A., Eduardo Simões, N., Wang, L., and Makropoulos, C.: Collaborative modelling for active involvement of stakeholders in urban flood risk management, Natural Hazards and Earth System Sciences, 12, 2821–2842, https://doi.org/10.5194/nhess-12-2821-2012, 2012.

Falck, A. S., Tomasella, J., Diniz, F. L., and Maggioni, V.: Applying a precipitation error model to numerical weather predictions for probabilistic flood forecasts, Journal of Hydrology, 598, 126 374, https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.126374, 2021.

Flack, D. L. A., Skinner, C. J., Hawkness-Smith, L., O'Donnell, G., Thompson, R. J., Waller, J. A., Chen, A. S., Moloney, J., Largeron, C., Xia, X., Blenkinsop, S., Champion, A. J., Perks, M. T., Quinn, N., and Speight, L. J.: Recommendations for Improving Integration in National End-to-End Flood Forecasting Systems: An Overview of the FFIR (Flooding From Intense Rainfall) Programme, Water, 11, https://doi.org/10.3390/w11040725, 2019.

Gerard, A., Martinaitis, S. M., Gourley, J. J., Howard, K. W., and Zhang, J.: An Overview of the Performance and Operational Applications of the MRMS and FLASH Systems in Recent Significant Urban Flash Flood Events, Bulletin of the American Meteorological Society, 102, E2165 – E2176, https://doi.org/https://doi.org/10.1175/BAMS-D-19-0273.1, 2021.

Gochis, D., Yu, W., Yates, D., and et al: The WRF-Hydro modeltechnical description and user's guide, version 3.0., NCAR Tech., Doc., 120 pp.,, 2015.

Godbout, L., Zheng, J. Y., Dey, S., Eyelade, D., Maidment, D., and Passalacqua, P.: Error Assessment for Height Above the Nearest Drainage Inundation Mapping, JAWRA Journal of the American Water Resources Association, 55, 952–963, https://doi.org/https://doi.org/10.1111/1752-1688.12783, 2019.

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


Gourley, J. J., Flamig, Z. L., Vergara, H., Kirstetter, P.-E., Clark, R. A., Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J. M., Hong, Y., and Howard, K. W.: The FLASH Project: Improving the Tools for Flash Flood Monitoring and Prediction across the United States, Bulletin of the American Meteorological Society, 98, 361 – 372, https://doi.org/https://doi.org/10.1175/BAMS-D-15-00247.1, 2017.

Guzzetti, F., Gariano, S. L., Peruccacci, S., Brunetti, M. T., Marchesini, I., Rossi, M., and Melillo, M.: Geographical landslide early warning systems, Earth-Science Reviews, 200, 102 973, https://doi.org/https://doi.org/10.1016/j.earscirev.2019.102973, 2020.

Hartke, S. H., Wright, D. B., Li, Z., Maggioni, V., Kirschbaum, D. B., and Khan, S.: Ensemble Representation of Satellite Precipitation Uncertainty Using a Nonstationary, Anisotropic Autocorrelation Model, Water Resources Research, 58, e2021WR031 650, https://doi.org/https://doi.org/10.1029/2021WR031650, e2021WR031650 2021WR031650, 2022.

Hocini, N., Payrastre, O., Bourgin, F., Gaume, E., Davy, P., Lague, D., Poinsignon, L., and Pons, F.: Performance of automated methods for flash flood inundation mapping: a comparison of a digital terrain model (DTM) filling and two hydrodynamic methods, Hydrology and Earth System Sciences, 25, 2979–2995, https://doi.org/10.5194/hess-25-2979-2021, 2021.

Hofmann, J. and Schüttrumpf, H.: Risk-Based and Hydrodynamic Pluvial Flood Forecasts in Real Time, Water, 12, https://doi.org/10.3390/w12071895, 2020.

Hu, A. and Demir, I.: Real-Time Flood Mapping on Client-Side Web Systems Using HAND Model, Hydrology, 8, https://doi.org/10.3390/hydrology8020065, 2021.

Jirak, I. L., Clark, A. J., B. Roberts, B. T. G., and Weiss, S. J.: Exploring the optimal configuration of the High Resolution Ensemble Forecast system, 25th Conf. on Numerical Weather Prediction Denver, CO, Amer. Meteor. Soc., 14B.6,, 2018.

Kim, G. and Barros, A. P.: Quantitative flood forecasting using multisensor data and neural networks, Journal of Hydrology, 246, 45–62, https://doi.org/https://doi.org/10.1016/S0022-1694(01)00353-5, 2001.

Knighton, J., Buchanan, B., Guzman, C., Elliott, R., White, E., and Rahm, B.: Predicting flood insurance claims with hydrologic and socioeconomic demographics via machine learning: Exploring the roles of topography, minority populations, and political dissimilarity, Journal of Environmental Management, 272, 111 051, https://doi.org/https://doi.org/10.1016/j.jenvman.2020.111051, 2020.

Kuligowski, R. J. and Barros, A. P.: USING ARTIFICIAL NEURAL NETWORKS TO ESTIMATE MISSING RAINFALL DATA1, JAWRA Journal of the American Water Resources Association, 34, 1437–1447, https://doi.org/https://doi.org/10.1111/j.1752-1688.1998.tb05443.x, 1998a.

Kuligowski, R. J. and Barros, A. P.: Localized Precipitation Forecasts from a Numerical Weather Prediction Model Using Artificial Neural Networks, Weather and Forecasting, 13, 1194 – 1204, https://doi.org/https://doi.org/10.1175/1520-0434(1998)013<1194:LPFFAN>2.0.CO;2, 1998b.

Kuller, M., Schoenholzer, K., and Lienert, J.: Creating effective flood warnings: A framework from a critical review, Journal of Hydrology, 602, 126 708, https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.126708, 2021.

Lahmers, T. M., Hazenberg, P., Gupta, H., Castro, C., Gochis, D., Dugger, A., Yates, D., Read, L., Karsten, L., and Wang, Y.-H.: Evaluation of NOAA National Water Model Parameter Calibration in Semiarid Environments Prone to Channel Infiltration, Journal of Hydrometeorology, 22, 2939 – 2969, https://doi.org/https://doi.org/10.1175/JHM-D-20-0198.1, 2021.

Li, Z., Gao, S., Chen, M., and et al: The conterminous United States are projected to become more prone to flash floods in a high-end emissions scenario., Commun Earth Environ 3, 86, https://doi.org/https://doi.org/10.1038/s43247-022-00409-6, 2022.

Liao, M. and Barros, A. P.: Toward optimal rainfall – Hydrologic QPE correction in headwater basins, Remote Sensing of Environment, 279, 113 107, https://doi.org/https://doi.org/10.1016/j.rse.2022.113107, 2022.

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, https://doi.org/https://doi.org/10.1175/WAF-D-20-0016.1, 2020.

Mascaro, G., Hussein, A., Dugger, A., and Gochis, D. J.: Process-based calibration of WRF-Hydro in a mountainous basin in southwestern U.S., JAWRA Journal of the American Water Resources Association, 59, 49–70, https://doi.org/https://doi.org/10.1111/1752-1688.13076, 2023.

Min, L., Fitzjarrald, D. R., Du, Y., Rose, B. E. J., Hong, J., and Min, Q.: Exploring Sources of Surface Bias in HRRR Using New York State Mesonet, Journal of Geophysical Research: Atmospheres, 126, e2021JD034 989, https://doi.org/https://doi.org/10.1029/2021JD034989, e2021JD034989 2021JD034989, 2021.

Moazami, S. and Najafi, M.: A comprehensive evaluation of GPM-IMERG V06 and MRMS with hourly ground-based precipitation observations across Canada, Journal of Hydrology, 594, 125 929, https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125929, 2021.

Moges, E., Demissie, Y., Larsen, L., and Yassin, F.: Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis, Water, 13, https://doi.org/10.3390/w13010028, 2021.

Mondal, M. S. H., Murayama, T., and Nishikizawa, S.: Examining the determinants of flood risk mitigation measures at the household level in Bangladesh, International Journal of Disaster Risk Reduction, 64, 102 492, https://doi.org/https://doi.org/10.1016/j.ijdrr.2021.102492, 2021.

Moore, R., McKay, L., Rea, A., Bondelid, T., Price, C., Dewald, T., and Johnston, C.: User's Guide for the National Hydrography Dataset Plus (NHDPlus) High Resolution, Open-File Report 2019–1096, https://pubs.usgs.gov/of/2019/1096/ofr20191096.pdf, 2019.

Moussa, R. and Bocquillon, C.: On the use of the diffusive wave for modelling extreme flood events with overbank flow in the floodplain, Journal of Hydrology, 374, 116–135, https://doi.org/https://doi.org/10.1016/j.jhydrol.2009.06.006, 2009.

Mudashiru, R. B., Sabtu, N., Abustan, I., and Balogun, W.: Flood hazard mapping methods: A review, Journal of Hydrology, 603, 126 846, https://doi.org/https://doi.org/10.1016/j.jhydrol.2021.126846, 2021.

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

Oddo, P. C. and Bolten, J. D.: The Value of Near Real-Time Earth Observations for Improved Flood Disaster Response, Frontiers in Environmental Science, 7, https://doi.org/10.3389/fenvs.2019.00127, 2019.

Rakovec, O., Hill, M. C., Clark, M. P., Weerts, A. H., Teuling, A. J., and Uijlenhoet, R.: Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models, Water Resources Research, 50, 409–426, https://doi.org/https://doi.org/10.1002/2013WR014063, 2014.

Ran, Q., Wang, J., Chen, X., Liu, L., Li, J., and Ye, S.: The relative importance of antecedent soil moisture and precipitation in flood generation in the middle and lower Yangtze River basin, Hydrology and Earth System Sciences, 26, 4919–4931, https://doi.org/10.5194/hess-26-4919-2022, 2022.

Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., and Waterloo, M. J.: HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia, Remote Sensing of Environment, 112, 3469–3481, https://doi.org/https://doi.org/10.1016/j.rse.2008.03.018, 2008.

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


Scriven, B. W. G., McGrath, H., and Stefanakis, E.: GIS derived synthetic rating curves and HAND model to support on-the-fly flood mapping, Natural Hazards, 109, 1629–1653, https://doi.org/10.1007/s11069-021-04892-6, 2021.

Silver, M., Karnieli, A., Ginat, H., Meiri, E., and Fredj, E.: An innovative method for determining hydrological calibration parameters for the WRF-Hydro model in arid regions, Environmental Modelling Software, 91, 47–69, https://doi.org/https://doi.org/10.1016/j.envsoft.2017.01.010, 2017.

Sun, J., Xue, M., Wilson, J. W., Zawadzki, I., Ballard, S. P., Onvlee-Hooimeyer, J., Joe, P., Barker, D. M., Li, P.-W., Golding, B., Xu, M., and Pinto, J.: Use of NWP for Nowcasting Convective Precipitation: Recent Progress and Challenges, Bulletin of the American Meteorological Society, 95, 409 – 426, https://doi.org/https://doi.org/10.1175/BAMS-D-11-00263.1, 2014.

Tao, J. and Barros, A. P.: Prospects for flash flood forecasting in mountainous regions – An investigation of Tropical Storm Fay in the Southern Appalachians, Journal of Hydrology, 506, 69–89, https://doi.org/https://doi.org/10.1016/j.jhydrol.2013.02.052, typhoon Hydrometeorology, 2013.

Teng, J., Jakeman, A., Vaze, J., Croke, B., Dutta, D., and Kim, S.: Flood inundation modelling: A review of methods, recent advances and uncertainty analysis, Environmental Modelling Software, 90, 201–216, https://doi.org/https://doi.org/10.1016/j.envsoft.2017.01.006, 2017.

Tolson, B. A. and Shoemaker, C. A.: Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, Water Resources Research, 43, https://doi.org/https://doi.org/10.1029/2005WR004723, 2007.

Valdez, E. S., Anctil, F., and Ramos, M.-H.: Choosing between post-processing precipitation forecasts or chaining several uncertainty quantification tools in hydrological forecasting systems, Hydrology and Earth System Sciences, 26, 197–220, https://doi.org/10.5194/hess-26-197-2022, 2022.

Viterbo, F., Mahoney, K., Read, L., Salas, F., Bates, B., Elliott, J., Cosgrove, B., Dugger, A., Gochis, D., and Cifelli, R.: A Multiscale, Hydrometeorological Forecast Evaluation of National Water Model Forecasts of the May 2018 Ellicott City, Maryland, Flood, Journal of Hydrometeorology, 21, 475 – 499, https://doi.org/https://doi.org/10.1175/JHM-D-19-0125.1, 2020.

Ward, P. J., Blauhut, V., Bloemendaal, N., Daniell, J. E., de Ruiter, M. C., Duncan, M. J., Emberson, R., Jenkins, S. F., Kirschbaum, D., Kunz, M., Mohr, S., Muis, S., Riddell, G. A., Schäfer, A., Stanley, T., Veldkamp, T. I. E., and Winsemius, H. C.: Review article: Natural hazard risk assessments at the global scale, Natural Hazards and Earth System Sciences, 20, 1069–1096, https://doi.org/10.5194/nhess-20-1069-2020, 2020.

Watson, K. and Cauller, S.: Flood Inundation, Flood Depth, and High-Water Marks for Selected Areas in West Virginia from the June 2016 Flood: U.S. Geological Survey data release, USGS, https://doi.org/https://doi.org/10.5066/F76T0K4K, 2017.

Wijayarathne, D., Coulibaly, P., Boodoo, S., and Sills, D.: Use of Radar Quantitative Precipitation Estimates (QPEs) for Improved Hydrological Model Calibration and Flood Forecasting, Journal of Hydrometeorology, 22, 2033 – 2053, https://doi.org/https://doi.org/10.1175/JHM-D-20-0267.1, 2021.

Wu, W., Emerton, R., Duan, Q., Wood, A. W., Wetterhall, F., and Robertson, D. E.: Ensemble flood forecasting: Current status and future opportunities, WIREs Water, 7, e1432, https://doi.org/https://doi.org/10.1002/wat2.1432, 2020.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., and Mocko, D.: Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/https://doi.org/10.1029/2011JD016048, 2012.

https://doi.org/10.5194/egusphere-2023-2088
Preprint. Discussion started: 9 October 2023


Yu, I., Park, K., and Lee, E. H.: Flood Risk Analysis by Building Use in Urban Planning for Disaster Risk Reduction and Climate Change Adaptation, Sustainability, 13, https://doi.org/10.3390/su132313006, 2021.

Zahmatkesh, Z., Han, S., and Coulibaly, P.: Understanding Uncertainty in Probabilistic Floodplain Mapping in the Time of Climate Change, Water, 13, https://doi.org/10.3390/w13091248, 2021.

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

Zheng, X., Tarboton, D. G., Maidment, D. R., Liu, Y. Y., and Passalacqua, P.: River Channel Geometry and Rating Curve Estimation Using Height above the Nearest Drainage, JAWRA Journal of the American Water Resources Association, 54, 785–806, https://doi.org/https://doi.org/10.1111/1752-1688.12661, 2018.