

Brief Communication: Rise of the Guadalupe River- A Multifaceted Post Event Analysis of the July 4, 2025, Flood in Central Texas

Anupal Baruah¹, Dinuke Munasinghe¹, Sagy Cohen¹, Mohamed Abdelkader², Dipsikha Devi¹, Yixian Chen¹, Humberto Vergara², [Riley McDermott](#)

¹Department of Geography and Environment, University of Alabama, Tuscaloosa, USA

²IIHR—Hydroscience & Engineering, University of Iowa, Iowa, USA

[3 NOAA National Weather Service, Office of Water Prediction –Federal](#)

Correspondence to: Anupal Baruah (abaruah@ua.edu), Sagy Cohen (sagy.cohen@ua.edu)

Abstract

The flash flooding across Central Texas on July 4th, 2025, caused more than 130 fatalities and property losses exceeding 20 billion dollars. The objective of this study is to evaluate the performance of the NOAA operational flood forecasting pipeline, including Quantitative Precipitation Forecasts, National Water Model short-range streamflow forecasts, and Office of Water Prediction flood inundation mapping during this catastrophic event, and to characterize how forecast skill and impact-based predictions varied with forecast lead time at gauged and ungauged locations. ~~The objective of this study is to diagnose the drivers of this catastrophic event and to analyze the temporal variability in forecasting flood inundation dynamics in the hours leading up to and during the event.~~ Using the Operational National Water Model short-range streamflow forecast product, we generated 306 forecasted flood inundation maps between July 3rd and July 4th, 2025. For evaluation, we constructed an inundation extent benchmark derived from USGS high water marks. Both impact-based and ~~pixel~~ skill-based assessments are presented.

1. Introduction

On July 4th at ~~dawn~~, dawn, a catastrophic flash flood hit Kerr and Kendall Counties in Texas, claiming at least 135 lives, making it one of the deadliest inland flooding events in United States history (<https://psl.noaa.gov/news/2025/texasfloods.html>). The event was triggered by intense rainfall up to 3-4 inches (76-100 mm) per hour in some of the impacted areas, causing flash flooding with exceptionally high peaks throughout the Guadalupe River basin. The Texas Hill Country, including Kerr and Kendall County, is prone to flash flooding, known as “Flash Flood Alley” (Saharia et al., 2017), partly due to the geomorphology of the area, having steep slopes and clay-rich soils, which result in high runoff rates (Sharif et al., 2010). Based on United States

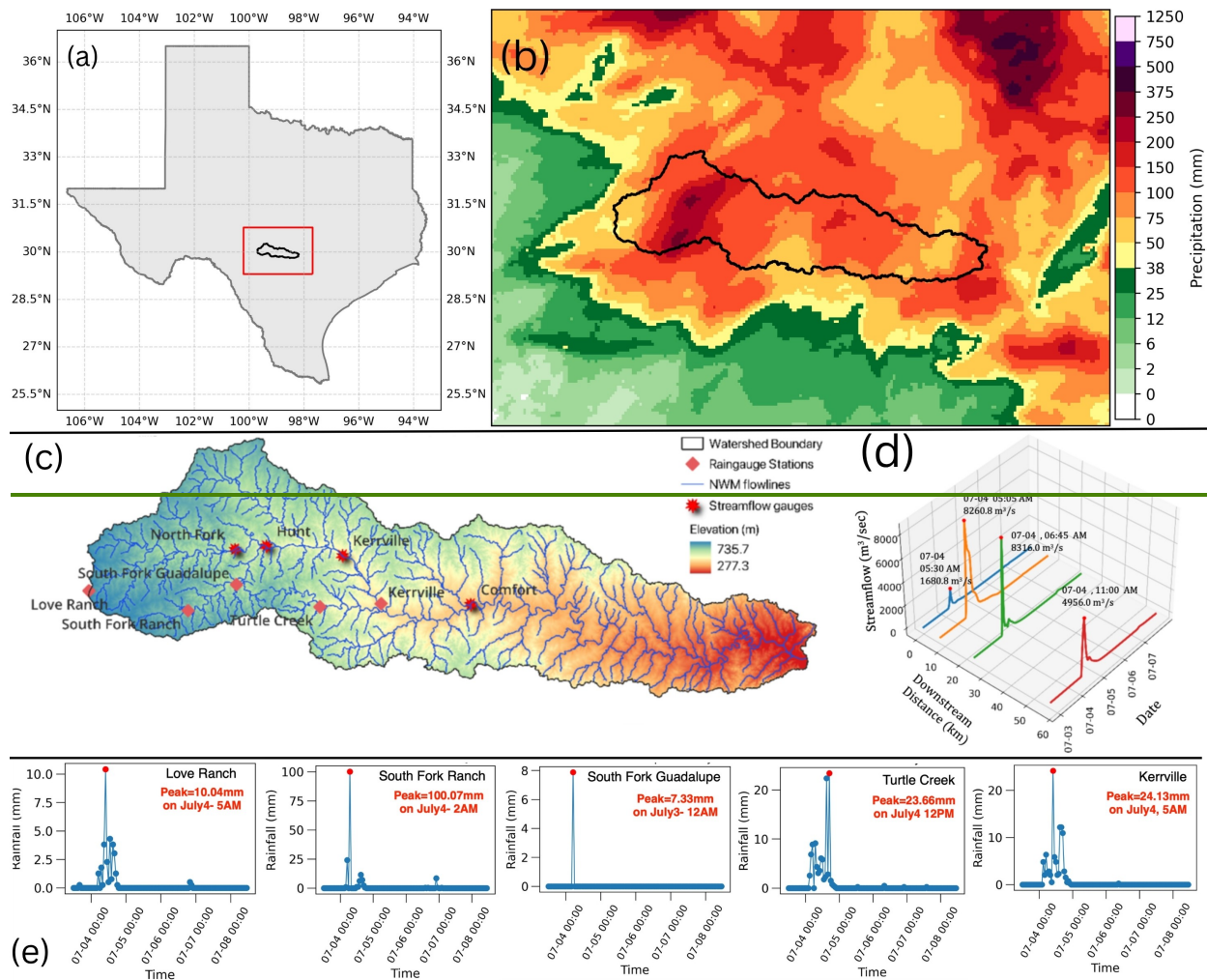
Geological Survey (USGS) streamflow observations, extreme flows were recorded downstream of the confluence of the South Fork Guadalupe River and the North Fork River (Figure 1d). At Hunt (Gage ID: 08615500), the streamflow reached 8,260.8 m³/s at 5:05 AM, far exceeding the 500-year return period flow of 4,160.8 m³/s. A significant portion of this peak was contributed by the ungauged South Fork Guadalupe River, which caused numerous fatalities near Camp Mystic (Nevitt 2025). Further downstream, at Kerrville, the streamflow peaked at 8,316.0 m³/s at 6:45 AM, crossing the 500-year return period flow of 7,937.6 m³/s. Apart from the peak values, the hydrographs at both upstream gauge stations exhibited exceptionally short times to peak, reaching the 500-year return period level within just 1–1.5 hours. This rapid rise of flow is indicative of extremely high runoff response, amplifying the severity of the event.

The main motivation of this study was to utilize the operational forecast pipeline and conduct an extensive impact-based and categorical assessment of the event across different forecast hours against a high-water-mark-derived benchmark FIM. We employed the National Weather Service (NWS) National Water Model (NWM) short-range forecast (Cosgrove et al., 2024) and the FIMserv tool (Baruah et al., 2025) to generate 306 FIMs at the Hydrologic Unit Code (HUC)-8 scale. Additionally, we developed a methodology to use post-event USGS observed high-water marks to create a spatially continuous FIM benchmark for evaluating the forecast FIMs. We investigate the trade-off between NWM forecast range and the reliability of flood impact predictions and translate the forecast error into meaningful information by quantifying building-level impacts in addition to standard pixel-level skill metrics. Specifically, we examined how FIM forecast skill, including building-level impacts, varied with respect to different forecast hours at five different locations (four streamflow gauges and near Camp Mystic) ~~The main motivation of this study was to utilize multiple operational forecast Flood Inundation Maps (FIM) and conduct an extensive impact-based and categorical assessment of the event across different forecast hours against high water mark derived benchmark FIM. We employed the National Weather Service (NWS) National Water Model (NWM) short-range forecast and the FIMserv tool (Baruah et al., 2025) to generate 306 FIMs at the Hydrologic Unit Code (HUC)-8 scale. Additionally, we developed a methodology to use post-event USGS observed high-water marks to create a spatially continuous FIM benchmark for evaluating the forecast FIMs. Our objective is to investigate the trade-off between NWM forecast range and the reliability of flood impact predictions. Specifically,~~

~~we examined how FIM forecast skill, including building-level impacts, varied with respect to different forecast hours at five different locations (four streamflow gauges and near Camp Mystic).~~

2. Hydrometeorological Assessment

Between July 3rd and 6th, 2025, the Texas Hill County region experienced an extraordinary rainfall amounts that produced catastrophic flash flooding along the Guadalupe River (Figure 1a-1c). The event was initiated by remnant moisture from Tropical Storm Barry, which had made landfall in Mexico days earlier, and was further intensified by a mesoscale convective vortex embedded within a persistent mid-level trough. These factors, combined with a strengthening low-level jet to sustain widespread convection, favor back-building successive thunderstorms across the Hill County (WPC, 2025). Hourly precipitation amounts from the Multi-Radar Multi-Sensor (MRMS) gauge-corrected ~~q~~Quantitative ~~p~~Precipitation ~~e~~Estimates ~~ion~~ (QPE) ~~dataset~~product (Zhang et al., 2016), ~~supplemented~~along with rain gauge observations, reveal accumulations exceeding 500 mm (20 inch) in parts of the upper Guadalupe basin between July 3 to July 6 (Figure 1b). Rainfall intensities surpassed 100-year return period thresholds for durations ranging from 3 to 24 hours at the rainfall gauges when compared against NOAA Atlas 14 precipitation frequency curves (Perica et al., 2018; NSSL, 2025). One of the most severe bursts occurred between 2 AM and 5 AM on July 4, when localized rainfall rates reached 100 mm/h (Figure 1e, Figure S2). The persistence of convective regeneration amplified accumulations and produced rainfall totals approaching 460 mm in some localized areas on July 4. Gauge records showed peaks at stations in the western part of the basin that aligned with MRMS QPE observations, where total precipitation indicated higher accumulations (Figure 1b, e, Figure S2).



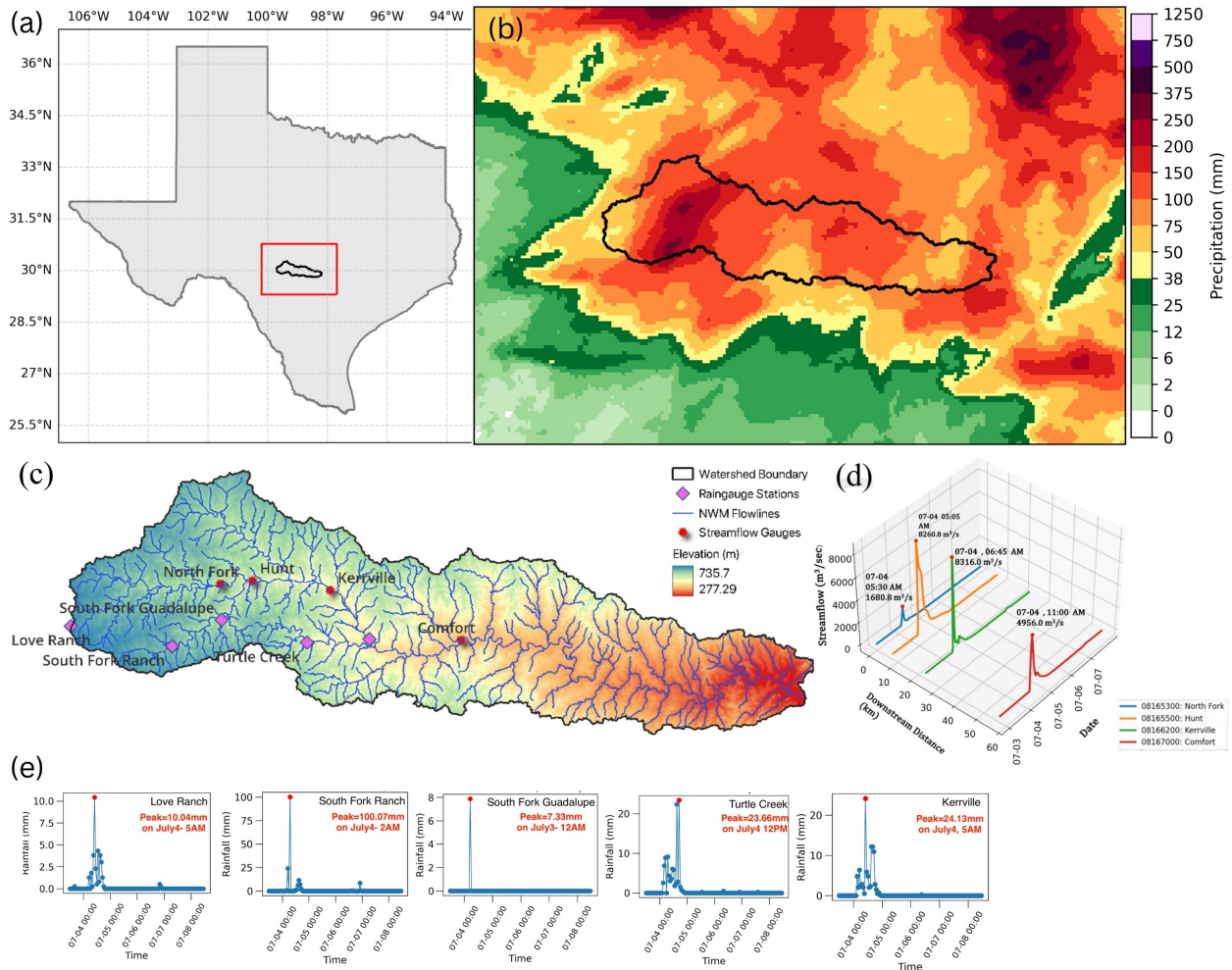
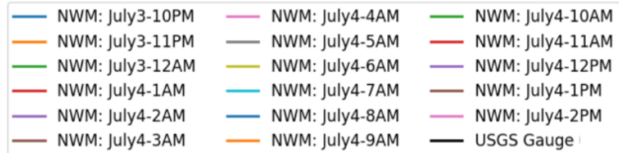
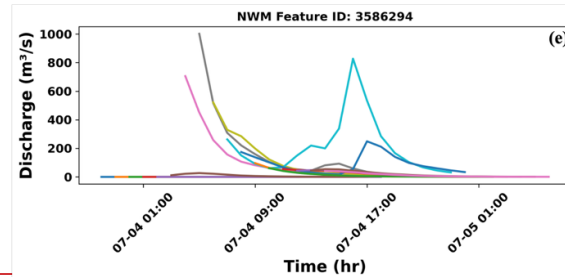
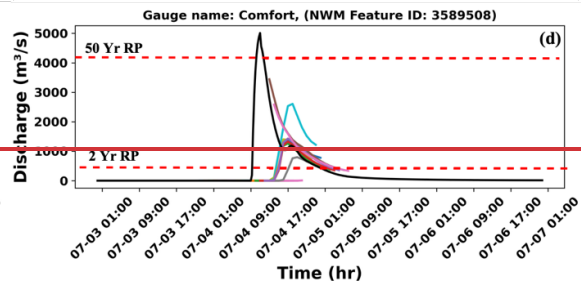
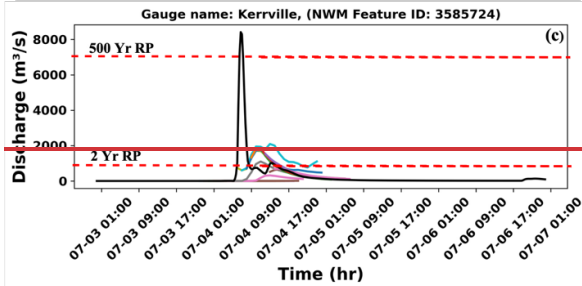
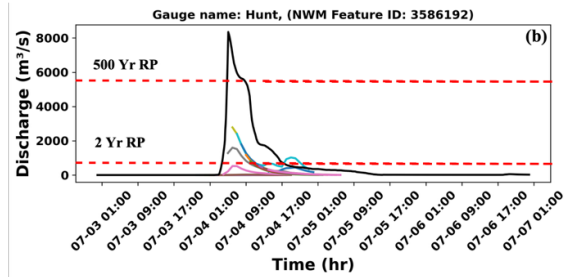
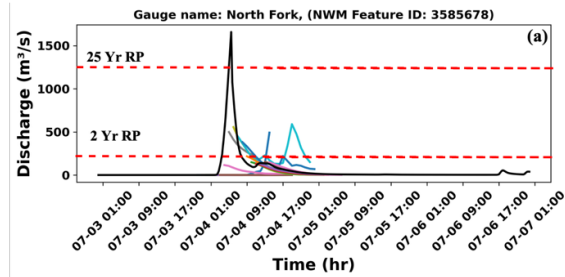


Figure 1. (a) Geographical location of the study area, (b) MRMS-derived total precipitation between 4–6 July 2025, (c) locations of USGS streamflow and rainfall gauges, (d) Gauge streamflow records along the reach, and (e) hourly rainfall time series.

The NWM short-range forecasts (SRF) provide hourly streamflow predictions extending to 18 hours forecast range (<https://water.noaa.gov/about/nwm>). We identified the NWM flowlines (feature_id) that intersect with the USGS gagesgauges and downloaded the SRF generated (Figure 2) from 10 PM, July 3rd till 2 PM, July 4th using FIMserv (Figure 2). At North Fork (peak at 5.30 AM), earlier SRFs show no flooding conditions, but there is a sudden jump on the forecast generated on July 4, between 4 AM to 7AM (120 m³/sec to 600 m³/sec). Near Camp Mystic, for which we do not have a USGS gauge, the forecast for 4 AM in South Fork Guadalupe River predicted a peak of 750 m³/sec. A Zoom-in view of the generated forecast closer to the event is shown in Figure-S1 of the supplementary material. At Hunt (peak at 5:05 AM), the forecast

generated for 4 AM indicates a possibility that flow will cross 500 m³/sec (above 2 Year RP flow of 266.08 m³/sec) in the next two hours, while the forecast generated for 5 AM, reaches the 10-year RP (1276.52 m³/sec). ~~At this Station (USGS 08165500), the peak discharge was recorded approximately at 05:00 AM CDT on July 4, 2025, while the upstream gauge on the North Fork Guadalupe River (USGS 08165300) peaked roughly one hour later, at 06:00 CDT. As shown in Figure 1a, the Hunt gauge is located immediately downstream of the confluence of the North Fork and South Fork branches of the Guadalupe River. The observed flow at the North Fork gauge, did not exceed the 25-year return-period, whereas the peak at Hunt exceeded the 500-year return-period level. This huge difference cannot be explained by inflow from the North Fork River alone and is consistent only with a substantial contribution from the South Fork Guadalupe River, which drains the Camp Mystic catchment and joins the North Fork just upstream of the Hunt gauge. Since no operational USGS stream gauge existsexisted at that time on the South Fork reach, the streamflow contribution from the South Fork River cannot be directly verified from in-situ observations. However, the timing of the peak and the spatial distribution of rainfall during the event strongly indicate that the South Fork River was the dominant contributor to the very high flashy streamflow peak at Hunt.~~ Further downstream, at Kerrville (peak at 6:45 AM), there was no indication of a high flow event; however, the forecast generated for 6 AM shows that flow will exceed the 2-year return period flow (284.22 m³/sec). At Comfort (peak at 11:00 AM), the initial forecast shows low flows, while the forecast generated after 10 AM indicates that flow will exceed the 10-year RP flow (1842.73 m³/sec), and the subsequent forecast follows the same trend. ~~Near Camp Mystic, for which we do not have a USGS gage, the forecast for 4 AM in South Fork Guadalupe River predicted a peak of 750 m³/sec. A Zoom-in view of the generated forecast closer to the event is shown in Figure S1 of the supplementary material.~~



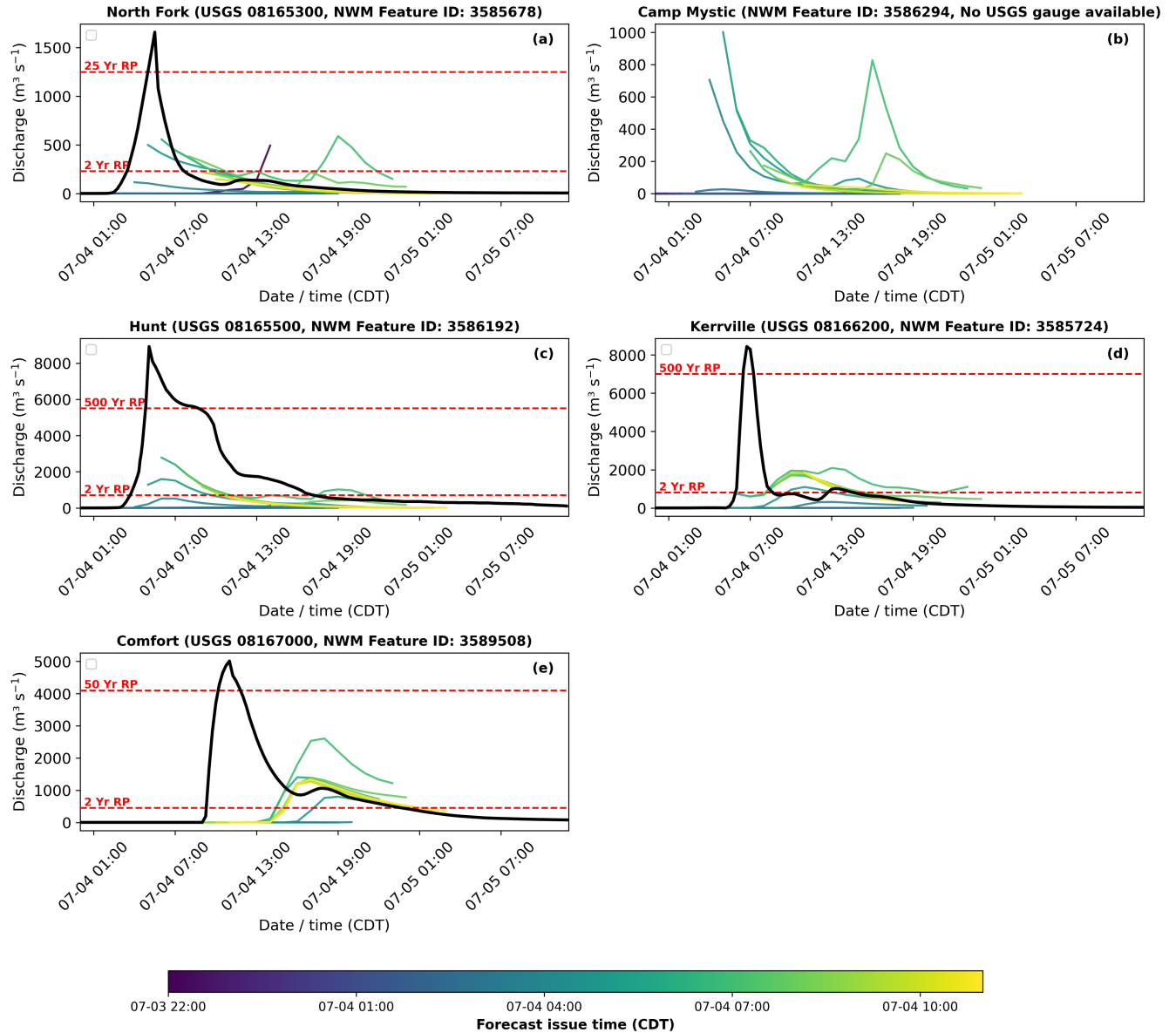


Figure 2. NWM short-range forecast at different lead times at (a) North Fork (b) ~~Hunt~~Camp Mystic (c) ~~Kerrville~~Hunt (d) ~~Comfort~~Kerrville and (e) ~~Camp-Mystie~~Comfort. Near Camp Mystic, there is no USGS gauge data available. The red dash line indicates the return period flow exceeded by the gauge readings.

Overall, the NWM SRF deviated substantially from the USGS observations in this case (Figure 2). These underpredictions in NWM SRF during the flood event were strongly affected by two primary sources of uncertainty: 1) Inconsistency in the location and magnitude of precipitation rates in the precipitation forcing from the High-Resolution Rapid Refresh (HRRR) model leading up to the event, and ii) the failure of USGS stream gauges used for data assimilation. The NWM

short-range configuration relies on HRRR Quantitative Precipitation Forecast (QPF) at 3-km resolution ([Dowel et al., 2022](#)), to drive ~~runoff~~streamflow forecast generation ([Dowel et al., 2022](#)).

A rainfall-to-rainfall comparison between the HRRR-QPF and MRMS ~~Quantitative Precipitation Estimate~~ (QPE) indicated a notable underestimation in the HRRR rainfall forecasts (Figure S3 in the supplementary material). ~~The MRMS Quantitative Precipitation Estimate (QPE) is a near-real-time precipitation analysis product derived primarily from weather radar observations and rain gauge data.~~ –When QPF underestimates or spatially misplaces convective rainfall cells, the resulting simulated discharge and flood peaks are either lower in magnitude or displaced in time and location relative to ~~a reference data series observations~~ ([Krajewski et al., 2025](#); [Vergara et al., 2023](#)). This bias has a pronounced effect on streamflow forecasts, especially for ~~small~~ basins (~~typically~~ smaller than 1,000 km²–) that are prone to flash flooding (~~Krajewski et al., 2025~~). Although a full atmospheric diagnostic of HRRR is beyond the scope of this Brief Communication, the systematic limitations of operational HRRR forecasts of warm-season convective rainfall are well documented in the literature and provide important context for the underprediction reported here. [James et al. \(2022\)](#) report regional warm-season precipitation biases over the southern and central United States, including a tendency for HRRR to dissipate nocturnal mesoscale convective systems over the Mississippi and southern Plains region too rapidly. [Min et al. \(2021\)](#) document a seasonally dependent warm and dry near-surface bias in HRRR during the warm season and an associated 2 to 4 h lag in the diurnal evolution of forecast convective available potential energy (CAPE) relative to Atmospheric Emitted Radiance Interferometer (AERI) observations. ~~This, a~~ lag time ~~that~~ has been attributed in part to the representation of sub-grid clouds and the planetary boundary layer and that is directly relevant to the timing of convective initiation. [Bytheway et al. \(2017\)](#) and [Bytheway and Kummerow \(2015\)](#) document object-based precipitation displacement errors in HRRR warm-season QPF on the order of 100 to 150 km, with no systematic geographical preference but with the largest errors over the southeastern United States. [Lawson et al. \(2022\)](#) similarly find centroid displacement errors and a slow propagation bias for some convective systems in the HRRR ensembles. [Henderson et al. \(2021\)](#) show that HRRR convective initiation is frequently mistimed relative to GOES-16 brightness-temperature evolution. [Chen et al. \(2024\)](#) further document that HRRR systematically struggles with the spatial organization of warm-season precipitation modes. In a nutshell, the errors QPE impacting the FIM forecast could be attributed to (i) localized convective rainfall rates can be substantially underestimated when nocturnal MCSs

dissipate prematurely in the model, (ii) the location of intense precipitation cells can be displaced between successive forecast initializations, and (iii) the timing of convective initiation and CAPE evolution may lag by a few hours. Each of these behaviors is consistent with what we observe for the July 4, 2025, Guadalupe event. For instance, successive HRRR initializations placed the convective core in different locations, under-resolved the localized rainfall rates, and produced a streamflow that consistently underrepresented both the magnitude and the timing of the rising limb.

The combination of the spatially misplaced QPF, smaller basin sizes, and flash flood prone geomorphology caused the NWM to signal intense flooding potential in the basins surrounding the Guadalupe River in the hours leading up to the event. However, the changing signal location in each timestep resulted in underprediction for the Guadalupe basin. By "changing signal location" we mean that there is a spatial mismatch in successive HRRR forecast cycles in the hours leading up to the event over different geographic locations in Upper Guadalupe Basin from one cycle to the next. Consequently, the NWM short-range forecasts driven by these HRRR cycles produced spatially inconsistent runoff predictions. The underestimation in the NWM during the event was exacerbated by the failure of the USGS gauge (08165500) near Hunt on the Guadalupe River during the fast-rising streamflow, which is confirmed on the USGS National Water Information System page for the site, where the peak of $315,000 \text{ ft}^3 \text{ s}^{-1}$ ($\approx 8,910 \text{ m}^3 \text{ s}^{-1}$) is documented as estimated value rather than an observed one (USGS, 2025; <https://waterdata.usgs.gov/monitoring-location/USGS-08165500>) as seen from the outage of data on the USGS water data gauge page from July 4th to the 5th (<https://wik.ly/3Wjb6hk>). The streamflow values shown in Figure 2(c) between approximately 04:50 on July 4 and 15:20 on July 5 are not direct gauge measurements, those are post-event reconstructions developed by USGS from surveyed high-water marks using indirect measurement techniques (Benson & Dalrymple, 1967). This outage disrupted the NWM's data assimilation process (Seo et al., 2021) from performing streamflow "nudging" to correct model states toward observed discharge. These errors were extrapolated with downstream flow propagation, resulting in systematic underprediction of streamflow. In the National Water Model (NWM), forecast errors are partially constrained through streamflow nudging during the Analysis and Assimilation (AnA) cycle, in which simulated discharge at USGS-gauged reaches is adjusted toward observations by adding a time-weighted correction to the channel-routing equation. The corrected discharge state then propagates

downstream through the Muskingum – Cunge routing scheme, thereby reducing forecast errors at downstream locations (Cosgrove et al., 2024; McCreight et al., 2024). In this event the failure of the Hunt gauge on the rising limb of the flood hydrograph prevents the assimilation of streamflow observations at this upstream location. Consequently, the incoming flood wave entered the routing network without observational constraint, allowing rainfall-driven errors in upstream inflows to propagate and accumulate downstream rather than being corrected. This resulted in progressively degraded discharge forecasts at downstream gauges, which in turn systematically reduced the accuracy of the corresponding flood inundation map (FIM) forecasts. However, even with the continuous gauge observations without any failure, the current NWM streamflow nudging scheme provides limited correction at longer forecast lead times because the observational discharge adjustment decays with forecast time (Seo et al. 2021).

3. Forecast FIM Generation and Evaluation

We used FIMserv to generate 306 flood inundation maps using the above described NWM forecasted flows. FIMserv assigns each forecast as an inflow to the OWP HAND-FIM model, which uses reach-averaged synthetic rating curves to generate binary FIMs, based on the Height Above Nearest Drainage (HAND) approach (Aristizabal et al., 2023; Chen et al., 2025; Baruah et al., 2025).

3.1 Derivation of Maximum Forecasted Inundation

The maximum extents from the forecast FIMs at a given reference time are estimated by combining multiple short-range forecast FIMs (Figure S4 in the supplementary material). For example, to assess flooding on July 4th, 5:00 am (reference time), we used the maximum inundation extent from a total of 7 forecasts from earlier timestamps (10 pm July 3 through 4:00 am July 4) that predict conditions at 5:00 am. Each raster (i.e., each of the 7 forecasts) was stacked, and for each pixel, if it was marked as flooded in any raster, it was considered flooded in the final composite. This approach produces a temporally integrated, conservative estimate of potential flooding (i.e., the largest possible forecasted extent for that hour), capturing uncertainty and variability in subsequent short-range forecasts.

3.2 Flood benchmark map creation using High Water Marks (HWMs)

We used the USGS High-water marks (HWMs) as the primary reference data for the creation of the benchmark FIM. HWM observations were obtained from the USGS Short-Term Network (STN) database (<https://stn.wim.usgs.gov/STNDataPortal/#>) and subjected to multi-stage preprocessing. First HWMs were filtered based on data quality (“Good” or “Excellent” quality were retained). Surveyed elevations (elev_ft) were converted to metric units (elev_m) relative to mean sea level. Next, global outlier removal was performed using the interquartile range (IQR) method (Tukey, 1977) to eliminate extreme data points. Subsequently, spatial outliers were identified using Local Moran’s I, which evaluates each HWM in the context of its neighboring points rather than the dataset as a whole. The spatial relationships are defined using a weights matrix, with the neighborhood distance empirically determined through semivariogram analysis. This approach ensures that the detection of local outliers reflects the natural decay of spatial autocorrelation across the floodplain, improving the reliability of the interpolated water surface elevation. The filtered HWM dataset was then interpolated into a continuous Water Surface Elevation (WSE) raster using the Topo to Raster algorithm in ArcGIS Pro 3.4.0, a hydrologically conditioned interpolation approach that enforces drainage continuity (Feaster and Koeing.,2017; Koeing et al., 2016). The interpolated WSE surface was subtracted from a 10 m Digital Elevation Model (DEM), yielding flood depth estimates. Cells with positive residuals ($DEM \leq WSE$) were classified as inundated, while negative residuals were assigned as non-flooded. This procedure yielded a benchmark flood inundation map derived solely from empirical HWM data that was used in FIM forecast evaluations in subsequent sections. This benchmark is particularly valuable because it enables spatial evaluation of forecast FIMs at locations where in-situ streamflow observations are unavailable either because no USGS gauge exists, as is the case near Camp Mystic, or because the gauge failed during the event, as occurred at Hunt.

3.3 Evaluation of the forecasted FIM against the benchmark FIM

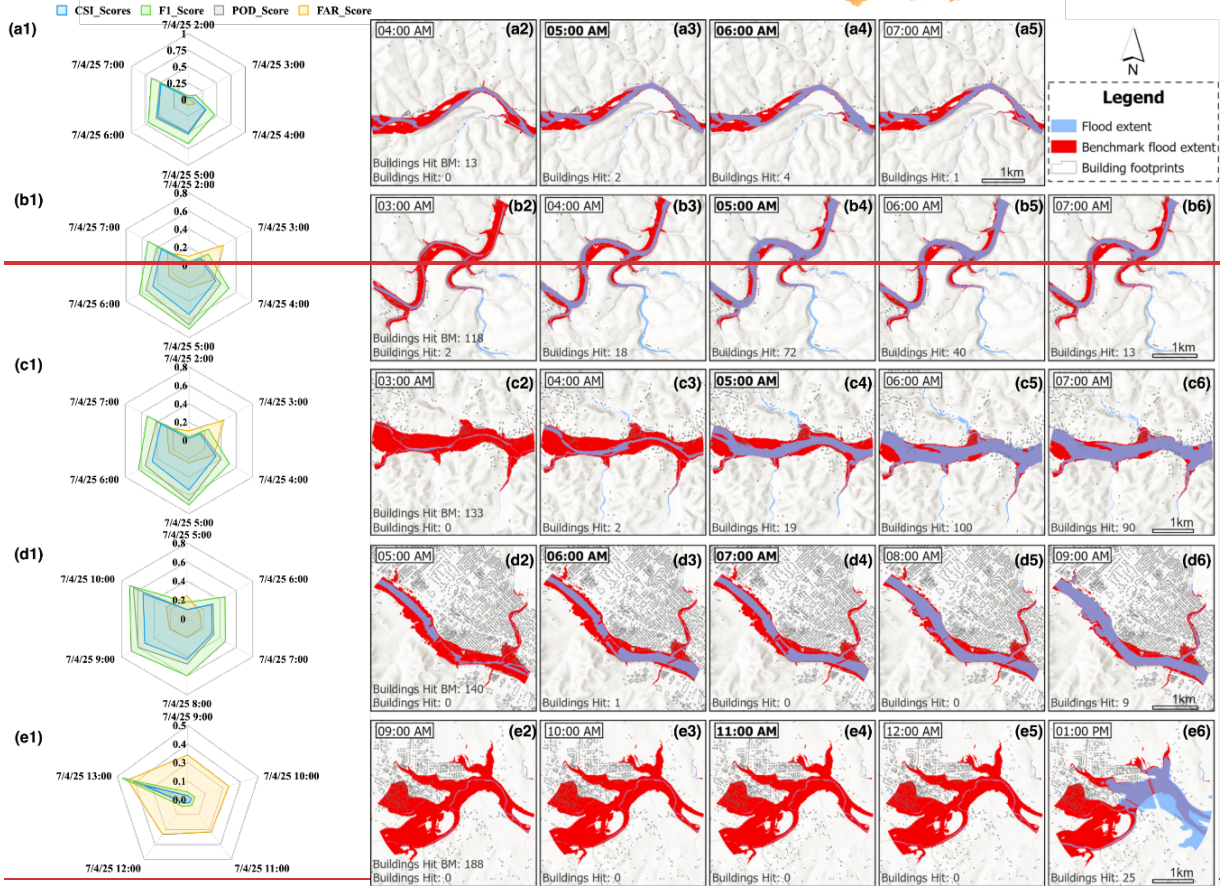
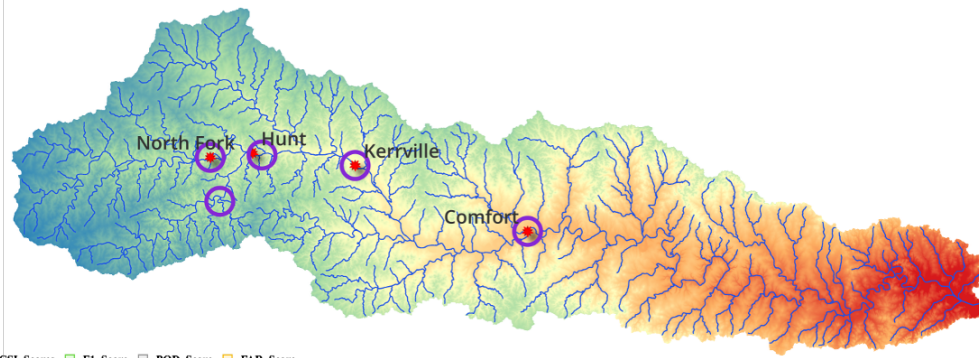
We used FIMeval (Devi et al., 2025) to evaluate the forecasted FIMs at five locations (Figure 3) across different forecast times against the benchmark FIM. ~~Since the HWM-derived benchmark is static (maximum inundation) and the exact time at which the HWMs were attained during the flood event is unknown, the gage-recorded peak time was used as the reference, and forecasted FIMs were evaluated within a ± 2 -hour window around this reference hour. This introduces an additional~~

~~source of temporal uncertainty into the overall evaluation process. Since, the HWMs do not carry information about when each mark was attained during the event, and there is no guarantee that all HWMs across the basin correspond to the same time, different locations have reached their maximum at different hours. The HWM-derived benchmark FIM is therefore inherently a maximum-extent product rather than a snapshot at a specific time. We used the gauge-recorded peak time as a physically meaningful temporal anchor (the closest available proxy for when peak conditions occurred at each location), and the ± 2 -hour window allows the evaluation to accommodate the fact that the HWM recorded maximum may have been attained slightly before or after the gauged peak. Aligning to the forecasted peak would not remove this ambiguity, because the source of the temporal uncertainty lies in the benchmark itself, not in the choice of reference time.~~

For pixel-based assessment, we used Critical Success Index (CSI), F1-Score, Probability of Detection (POD), and False Alarm Rate (FAR) as the evaluation matrix (Figure 3), and for impact assessment, we compared the number of predicted and observed flooded buildings at different forecast times.

~~At North Fork, the flood peak occurred at 5:30 AM. Compared to earlier forecasts, the FIM forecasts for 4 AM to 6 AM show gradual improvements in CSI, F1, and POD scores, although the 4 AM forecast also exhibits a higher FAR (Figure 3a1).~~ At North Fork, the flood peak occurred at 5:30 AM. Compared to earlier forecasts, the FIM forecasts for 4 AM to 6 AM show gradual improvements in CSI, F1, and POD scores, while FAR decreases over the same period, indicating an overall improvement in performance (Figure 3b1). The impact assessment (Figures 3b2–b6) indicates that the benchmark FIM estimates 133 buildings impacted, compared to only 2 for the 4 AM forecast and 19 for the 5 AM forecast. ~~The impact assessment (Figures 3a2–a5) indicates that 4 AM forecast FIM predicts no building flooded, 5 AM forecast predicts two flooded buildings, whereas the benchmark FIM shows 13 flooded buildings.~~ Near Camp Mystic, the evaluation scores for the 4 and 5 AM forecasts improved substantially compared to earlier forecasts (Figure 3**cb**1). In terms of impacts, the benchmark FIM shows 118 buildings affected, while the 4 AM and 5 AM forecasts predict only 18 and 72 flooded buildings, respectively (Figures 3**cb**2–**cb**6). At Hunt, the flood peak was recorded at 5:05 AM. Here, the evaluation scores also improved, with CSI values increasing from 0.34 for the 4 AM forecast to 0.52 for the 5 AM forecast, while

earlier forecasts yielded much lower scores (<0.25) (Figure 3~~de~~1). The impact assessment (Figures 3~~de~~2–~~de~~6) indicates that the benchmark FIM estimates 133 buildings impacted, compared to only 19 for the 5 AM forecast and only 2 for the 4 AM forecast. At Kerr, the peak was recorded at 6:45 AM. Here, we found higher evaluation scores in the post-event forecast. For instance, the CSI scores are below 0.30 for 6 AM, but the scores are increasing at 10 AM, at 0.53. A similar trend was observed in POD and F1 score. From the impact assessment, we found that the benchmark FIM shows 140 flooded buildings, while the forecast FIMs have only one building flooded at 6 AM. We also observed notable underprediction in forecast FIM extents downstream of hydraulic structures (Figure 3~~ed~~2–~~ed~~64). Such discrepancies likely stem from artifacts in the HAND raster and insufficient representation of dam release flows in the NWM streamflow forecasts (Aristizabal et al., 2023; Kim et al., 2020). At Comfort, the flood peak occurred at 11:00 AM. Early forecasts showed very low scores (<0.1), but the scores improved at 1 PM (Figure 3~~fe~~1). Eventually, we found no buildings hit from the predicted FIMs until 1 PM; At 1 PM for Comfort, there are 25 buildings predicted to be flooded, while the benchmark indicates 188 flooded buildings (Figure 3 ~~fe~~2–~~fe~~6).



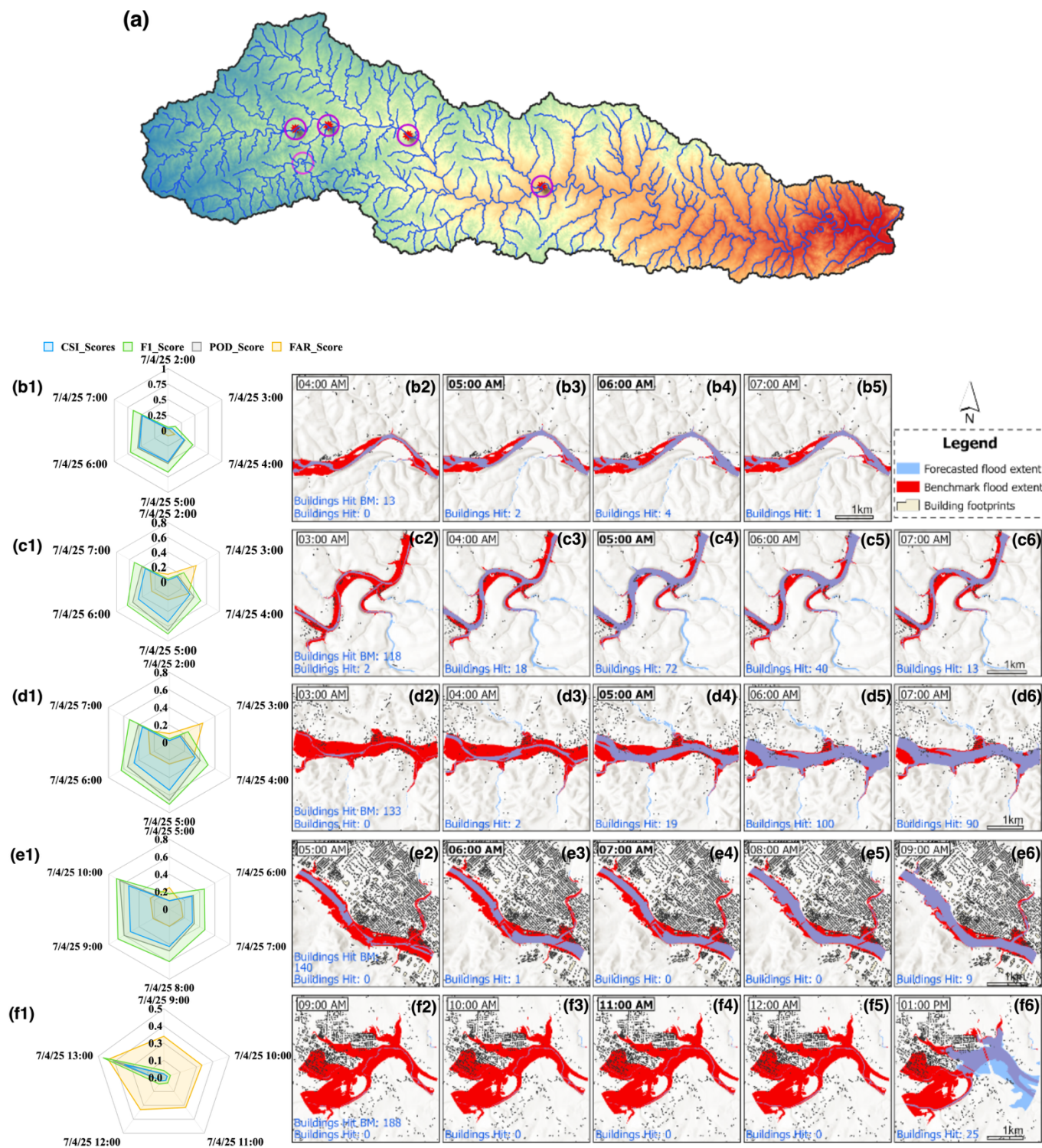


Figure 3. Top panel shows the FIM evaluation extent. (a) North Fork (b) Camp Mystic (c) Hunt (d) Kerrville and (e) Comfort. The bottom left panel (ba1-fe1) shows the accuracy metrics of FIM evaluation and the right panel (ba2-fe6) shows the forecast flood extent, benchmark flood extent, and number of building hits at different forecast hours. Each subpanel lists the forecast hour in the upper-left box; times that coincide with, or cover, the USGS-observed peak discharge are highlighted in bold.

4. Closing Remarks

~~In this brief communication, we highlight the hydroclimatic drivers and impacts of the Texas flash flood event on July 4, 2025. Using USGS gauge observations, we analyzed the NWM short-range streamflow forecasts. Across all four gauge locations, the NWM short-range forecasts considerably underpredicted the peak flow, with a pronounced lag between observed and forecasted flows at downstream sites. We found that underprediction and lag in short-range forecasts are mainly due to the errors in rainfall estimation and the failure of data assimilation ("nudging") caused by malfunctioning USGS gauges during the peak flow. In this brief communication, we have presented an end-to-end evaluation of the NOAA operational forecasting chain for the Texas flash flood event on July 4, 2025, using precipitation forecasts, streamflow predictions, and flood inundation mapping. Using USGS gauge observations and a benchmark flood inundation extent generated from post-event ground-sampled high-water marks, we analyzed the NWM short-range streamflow forecasts and the resulting OWP HAND-derived flood inundation maps. Across all four gauged locations (North Fork, Hunt, Kerrville, and Comfort), the NWM short-range forecasts considerably underpredicted the peak flow, with a pronounced lag between observed and forecasted flows at downstream sites. No USGS gauge data is available on the South Fork Guadalupe River near Camp Mystic for that time period, so a direct streamflow evaluation was not possible at that location. We found that underprediction and lag in short-range forecasts are mainly due to errors in rainfall estimation and the failure of data assimilation ("nudging") caused by malfunctioning USGS gauges during the peakpeak flow. The streamflow forecast underpredictions seem to have propagated, resulting in substantially underestimated flood extents in forecasted FIMs at different lead times. Given the risk of in-situ gauge failure at peak~~

flows, complementary observations beyond the USGS telemetered network should be considered to support continuous data assimilations in the NWM. Realistically, no single alternative observation type can match the sub-hourly temporal resolution required to capture a flash flood that, in this case, exceeded the 500-year return period at Hunt in less than 1.5 hours. At the in-situ level, dense networks of low-cost water-level sensors (Nemnem et al., 2026), bridge-mounted flood sensors (radar or acoustic), and image-based discharge estimation from fixed cameras (Young et al., 2015) using Large Scale Particle Image Velocimetry (LSPIV) can provide sub-hourly observations that are far less likely to be damaged during extreme flow. At the spaceborne level, satellite altimetry such as SWOT (Patidar et al., 2025), Sentinel-3 (Kittel et al., 2020), and Synthetic Aperture Radar (SAR) inundation retrievals provide spatially distributed water-surface and extent information, but their fixed revisit cycles (typically multiple days) and orbit timing rarely coincide with the few hour duration of a flash flood scale. We see particular potential, however, in commercial SAR satellite constellations with tasking capabilities for all weather day/night imaging. When pre-configured for tasking over flash flood-prone basins based on NWS watches or QPF-based heavy rain outlooks, these constellations can provide near-real-time inundation snapshots at multiple times during a single event. ~~These streamflow underpredictions seem to have propagated, resulting in substantially underestimated flood extents in forecasted FIMs at different lead times. Using the gauge-recorded peak time as a reference, we evaluated forecast FIMs at different hours, against a benchmark flood inundation extent generated using ground-sampled high-water marks from USGS. At upstream gagesgauges, forecast skill improved as the forecast time approached the reference peak, whereas at downstream locations (Kerr and Comfort), evaluation scores were very poor at the reference time but improved after the observed peak flow. It is also clear from the impact assessment that forecast FIM was unable to capture the majority of flooded buildings compared to the benchmark. This case study points to the need for communicating the uncertainty and error bounds in operational forecast FIMs for early response and decision-making. One promising approach is probabilistic FIM which can effectively represent and leverage uncertainty in streamflow forecasts, providing more informative and actionable guidance for emergency responders. The streamflow forecast underpredictions seem to have propagated, resulting in substantially underestimated flood extents in forecasted FIMs at different lead times.~~

Using the gauge-recorded peak time as a reference, we evaluated forecast FIMs at different hours against the HWM-derived benchmark. At upstream gauges, forecast skill improved as the forecast time approached the reference peak, whereas at downstream locations (Kerr and Comfort), evaluation scores were very poor at the reference time but improved after the observed peak flow. The HWM-derived benchmark also enabled forecast FIM evaluation at the ungauged Camp Mystic location. It is also clear from the impact assessment that forecast FIM was unable to capture the majority of the building impacts. This study also points to the need for communicating the uncertainty and error bounds in operational forecast FIMs for early response and decision-making. One promising emerging approach is probabilistic FIM which can effectively represent and leverage uncertainty in streamflow forecasts and terrain parameters, providing more informative and actionable guidance for emergency responders. –Probabilistic FIM can be most useful in operational settings when paired with predefined action triggers. By contrast, a binary deterministic forecast offers only "flood" or "no flood," and repeated false alarms or false sense of security, which can erode public trust. A probabilistic map, on the other hand, allows responders to decide if, e.g., a 10% probability triggers preparation, while a 90% probability may trigger action or evacuation. Specifically, ensemble exceedance probabilities can be tied to (i) impact thresholds, e.g., the probability that critical infrastructure, schools, or evacuation routes are inundated; and (ii) jointly developed emergency-management protocols in which a given probability level automatically activates a defined response action. For instance, action triggers can be defined as a joint function of the probability of inundation and the severity of its impact on people and infrastructure. A low-to-moderate probability (e.g., $P(\text{flooding}) \geq 0.2$) over high-consequence assets such as low-water crossings, school routes, or campsites can be sufficient to initiate precautionary road and bridge closures and to pre-position emergency resources, whereas a higher probability (e.g., for $P \geq 0.5-0.7$) over residential zones can trigger formal evacuation preparation (mobilization of buses, notification of vulnerable residents, opening of shelters). Embedding both impact and probability may help emergency managers to prioritize the alerts, situating probabilistic forecasting as a compelling focus for future studies aimed at improving the accuracy of prediction in flash flood events. While this study focuses on the technical components of the operational forecasting pipeline, we acknowledge that forecast quality is only one of the components, while warning dissemination, warning interpretations by at-risk populations, evacuation timing, and the

operational decisions of emergency responders all contribute substantially to the alert dissemination process

Code and data availability

Code for downloading the NWM short-range forecast and for producing the flood inundation maps is available at <https://github.com/sdmlua/FIMserv>. The benchmark high water flood inundation map for the Texas Flood event is available in <https://github.com/sdmlua/fimbench>

Author contributions

Conceptualization: Anupal Baruah, Dinuke Munasinghe, Sagy Cohen. Formal analysis and methodology: Anupal Baruah, Dinuke Munasinghe, Mohamed Abdelkader, Dipsikha Devi, Yixian Chen, ~~Derek~~ Supervision: Sagy Cohen. Writing (original draft preparation): Anupal Baruah, Mohamed Abdelkader, Dinuke Munasinghe, Dipsikha Devi, Riley McDermott, ~~Derek Giardino~~, Humberto Vergara. All authors reviewed the final paper.

Acknowledgments

Funding for this project was provided by the National Oceanic & Atmospheric Administration (NOAA), awarded to the Cooperative Institute for Research to Operations in Hydrology (CIROH) through the Cooperative Agreement with The University of Alabama (NA22NWS4320003).

Conflict of Interest

None of the authors has any conflict of interests

Disclaimer

The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the authors and do not necessarily reflect the views of NOAA.

References

1. Aristizabal, F., Salas, F., Petrochenkov, G., Grout, T., Avant, B., Bates, B., ... & Judge, J. (2023). Extending height above nearest drainage to model multiple fluvial sources in flood

inundation mapping applications for the US National Water Model. Water resources research, 59(5), e2022WR032039.

~~1-2.~~

3. Baruah, A., Dhital, S., Cohen, S., Tran, T. N. D., Elhaddad, H., Watts, C. L., Devi, D., Chen, Y., & Pruitt, C. (2025). FIMserv v. 1.0: A tool for streamlining Flood Inundation Mapping (FIM) using the United States operational hydrological forecasting framework. Environmental Modelling & Software, 106581.

4. Benson, M. A., & Dalrymple, T. (1967). General field and office procedures for indirect discharge measurements (U.S. Geological Survey Techniques of Water-Resources Investigations, Book 3, Chapter A1). U.S. Government Printing Office. <https://doi.org/10.3133/twri03A1>

5. Bytheway, J. L., & Kummerow, C. D. (2015). Toward an object-based assessment of high-resolution forecasts of long-lived convective precipitation in the central US. Journal of Advances in Modeling Earth Systems, 7(3), 1248-1264.

6. Bytheway, J. L., Kummerow, C. D., & Alexander, C. (2017). A features-based assessment of the evolution of warm season precipitation forecasts from the HRRR model over three years of development. Weather and Forecasting, 32(5), 1841-1856.

~~2-7.~~ Chen, I. H., Berner, J., Keil, C., Kuo, Y. H., & Craig, G. (2024). Classification of warm-season precipitation in High-Resolution Rapid Refresh (HRRR) model forecasts over the contiguous United States. Monthly Weather Review, 152(1), 187-201.

8. Chen, Y., Cohen, S., Baruah, A. et al. Merging Remote Sensing Derived River Slope Datasets with High-Resolution Hydrofabrics for the United States. Sci Data 12, 1657 (2025). <https://doi.org/10.1038/s41597-025-05941-6>

~~3-9.~~ Cosgrove, B., Gochis, D., Flowers, T., Dugger, A., Ogden, F., Graziano, T., et al. (2024). NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling. JAWRA Journal of the American Water Resources Association, 60(2), 247-272.

10. Devi, D., Dhital, S., Munasinghe, D., Baruah, A., Chen, Y., Tian, D., & Pruitt, C. (2026). A framework for the evaluation of flood inundation predictions over extensive benchmark databases. Environmental Modelling & Software, 196, 106786. <https://doi.org/10.1016/j.envsoft.2025.106786>

~~4.—~~

~~5-11.~~ Dowell, D. C., Alexander, C. R., James, E. P., Weygandt, S. S., Benjamin, S. G., Manikin, G. S., & Alcott, T. I. (2022). The High-Resolution Rapid Refresh (HRRR): An hourly updating convection-allowing forecast model. Part I: Motivation and system description. Weather and Forecasting, 37(8), 1371-1395.

~~6-12.~~ Feaster, T. D., & Koenig, T. A. (2017). Field manual for identifying and preserving high-water mark data (No. 2017-1105). US Geological Survey.

13. July 2025 Central Texas floods. (n.d.). In Wikipedia. Retrieved January 12, 2026, https://en.wikipedia.org/wiki/July_2025_Central_Texas_floods

- 7.14. James, E. P., Alexander, C. R., Dowell, D. C., Weygandt, S. S., Benjamin, S. G., Manikin, G. S., ... & Turner, D. D. (2022). The High-Resolution Rapid Refresh (HRRR): an hourly updating convection-allowing forecast model. Part II: Forecast performance. Weather and Forecasting, 37(8), 1397-1417.
15. Kim, J., Read, L., Johnson, L. E., Gochis, D., Cifelli, R., & Han, H. (2020). An experiment on reservoir representation schemes to improve hydrologic prediction: Coupling the national water model with the HEC-ResSim. *Hydrological Sciences Journal*, 65(10), 1652-1666
16. Kiel, B. M., Gallus Jr, W. A., Franz, K. J., & Erickson, N. (2022). A preliminary examination of warm season precipitation displacement errors in the upper Midwest in the HRRRE and HREF ensembles. Journal of Hydrometeorology, 23(6), 1007-1024.
17. Kittel, C. M., Jiang, L., Tøttrup, C., & Bauer-Gottwein, P. (2020). Sentinel-3 radar altimetry for river monitoring—a catchment-scale evaluation of satellite water surface elevation from Sentinel-3A and Sentinel-3B. Hydrology and Earth System Sciences Discussions, 2020, 1-35.
18. Henderson, D. S., Otkin, J. A., & Mecikalski, J. R. (2021). Evaluating convective initiation in high-resolution numerical weather prediction models using GOES-16 infrared brightness temperatures. Monthly Weather Review, 149(4), 1153-1172.
- 8.19. _____
- 9.20. Krajewski, W. F. (2025). Is this rainfall forecast good or bad? For flood forecasting, the answer is scale-dependent. Bulletin of the American Meteorological Society, 106(9). <https://doi.org/10.1175/BAMS-D-24-0166.1>
21. Koenig, T. A., Bruce, J. L., O'Connor, J., McGee, B. D., Holmes Jr, R. R., Hollins, R., ... & Pepler, M. C. (2016). Identifying and preserving high-water mark data (No. 3-A24). US Geological Survey.
- 10.22. Min, L., Fitzjarrald, D. R., Du, Y., Rose, B. E., Hong, J., & Min, Q. (2021). Exploring sources of surface bias in HRRR using New York State Mesonet. Journal of Geophysical Research: Atmospheres, 126(20), e2021JD034989.
23. Musser, J. W., Watson, K. M., Painter, J. A., & Gotvald, A. J. (2016). Flood-inundation maps of selected areas affected by the flood of October 2015 in central and coastal South Carolina (No. 2016-1019). US Geological Survey.
24. Nemnem, A. M., Khan, M. A., Downey, A. R. J., & Imran, J. (2026). Early Warning Potential of Low-Cost Sensors Evaluated Using the Camp Mystic Flash-Flooding Event. In Proceedings of the World Environmental and Water Resources Congress 2026. American Society of Civil Engineers (ASCE). <https://doi.org/10.1061/9780784486931.024>
- 11.25. _____
26. National Research Council (NRC). (2009). Mapping the Zone: Improving Flood Map Accuracy. National Academies Press.

27. Patidar, G., Paris, A., Indu, J., & Karmakar, S. (2025). How can SWOT-derived water surface elevations help calibrating a distributed hydrological model? *Journal of Hydrology*, 658, 133240. <https://doi.org/10.1016/j.jhydrol.2025.133240>
28. Saharia, M., Kirstetter, P.-E., Vergara, H., Gourley, J. J., Hong, Y., & Giroud, M. (2017). Mapping flash flood severity in the United States. *Journal of Hydrometeorology*, 18(2), 397–411
30. Seo, B. C., Krajewski, W. F., & Quintero, F. (2021). Multi-scale hydrologic evaluation of the national water model streamflow data assimilation. *JAWRA Journal of the American Water Resources Association*, 57(6), 875-884.
31. Sharif, H. O., Hunter, N. M., & Wang, Q. J. (2010). Hydrologic modeling of an extreme flood event using distributed physically-based models. *Journal of the American Water Resources Association*, 46(2), 292–307. <https://doi.org/10.1111/j.1752-1688.2010.00459.x>
32. Vergara, H., Gourley, J. J., & Erickson, M. (2023). An efficient ensemble technique for hydrologic forecasting driven by quantitative precipitation forecasts. *Journal of Hydrometeorology*, 24(3), 479–495. <https://doi.org/10.1175/JHM-D-22-0109.1>
33. Watson, K. M., Storm, J. B., Breaker, B. K., & Rose, C. E. (2017). Characterization of peak streamflows and flood inundation of selected areas in Louisiana from the August 2016 flood (No. 2017-5005). US Geological Survey.
34. Weather Prediction Center (WPC), Mesoscale Precipitation Discussion: #0582, https://www.wpc.ncep.noaa.gov/metwatch/metwatch_mpd_multi.php?md=582&yr=2025
35. Young, D. S., Hart, J. K., & Martinez, K. (2015). Image analysis techniques to estimate river discharge using time-lapse cameras in remote locations. *Computers & Geosciences*, 76, 1–10.
36. Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., ... & Kitzmiller, D. (2016). Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, 97(4), 621-638.
38. National Severe Storms Laboratory, Multi-Radar Multi-Sensor Operational Product Viewer, https://mrms.nssl.noaa.gov/qvs/product_viewer/.