

Mesoscale Model Forecast Skill for Hourly Rainfall During Postfire Flood and Debris-Flow Events in California

Nina Oakley

nina.oakley@gmail.com

California Geological Survey <https://orcid.org/0000-0001-5680-9296>

Luke A. McGuire

The University of Arizona

Deanna Nash

University of California San Diego Scripps Institution of Oceanography

Research Article

Keywords: wildfire, postfire, debris flow, flash flood, forecast, rainfall

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11296532/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Mesoscale Model Forecast Skill for Hourly Rainfall During Postfire Flood and Debris-Flow
Events in California

Nina S. Oakley^{1*}, Luke A. McGuire², Deanna Nash³

¹California Geological Survey, Burned Watershed Geohazards Program, Santa Rosa, CA, USA

²Department of Geosciences, University of Arizona, Tucson, AZ, USA

³Center for Western Weather and Water Extremes, Scripps Institution of Oceanography,
University of California, San Diego, La Jolla, CA, USA

*Corresponding author, nina.oakley@conservation.ca.gov

Abstract

Fire can increase flood and debris-flow susceptibility. Steep, recently burned watersheds can produce damaging postfire floods and debris flows (PFDFs) in response to short-duration (≤ 60 minutes), high-intensity rainfall with a 1-to-2-year recurrence interval. Rainfall forecasts could identify PFDF-triggering rainfall in recently burned areas, providing lead time for risk-reduction decisions. However, we lack quantitative evaluations of how well mesoscale weather models forecast the rainfall that triggers PFDFs. We evaluated two deterministic mesoscale weather forecast models, WestWRF and HRRR, for forecasting hourly rainfall associated with observed PFDF events in California, USA. We examined 10 PFDF events triggered by cool-season rainfall and compared forecasts with observations from at least three rain gauges within and surrounding each burn area using multiple spatial neighborhoods and temporal windows. Using a 9-km neighborhood and 5-hour window, both models reliably predicted that hourly rainfall would exceed 10 mm h^{-1} at lead times ≤ 24 hours, as indicated by a high probability of detection. Still, both models struggled to reproduce the magnitude of observed rainfall, with forecast intensities typically less than observed across all lead times. Forecast error was smallest within 24 hours of the event, with a median percent error of 6–16% among events, but increased in magnitude at longer lead times. Error was typically greatest for the most intense rainfall observations ($> 20 \text{ mm h}^{-1}$) and often persisted throughout the forecast horizon. These results support incorporating mesoscale weather forecasts into postfire debris-flow early warning systems while highlighting forecast limitations that could be considered in operational decision-making and hazard communication.

Keywords: wildfire, postfire, debris flow, flash flood, forecast, rainfall

1. Introduction

Postfire floods and debris flows (PFDFs) pose a threat to life, property, and infrastructure within and downstream of recently burned areas around the world, including western North America (e.g., Cannon and DeGraff 2009; McGuire et al. 2024) and specifically, California (CA) (Oakley et al. 2025). These events are triggered in steep watersheds by short-duration (≤ 1 hour), high-intensity rainfall. Rainfall forecasts are a key component of PFDF warning systems (NOAA-USGS Debris Flow Task Force 2005; Guzzetti et al. 2020).

To incorporate rainfall forecasts into a PFDF warning system, it is beneficial to understand which forecast models are most appropriate and their performance for the rainfall characteristics known to trigger events. This requires consideration of the scales of atmospheric motion. Atmospheric phenomena span spatial and temporal scales, from meters and seconds at the microscale to $> 10,000 \text{ km}$ and a month at the synoptic scale (Orlanski 1975; Lin 2007). The timeframes for skillful prediction generally increase with spatial scale; larger-scale features tend to be more predictable at longer lead times (Fig. 1). For example, atmospheric rivers (ARs), synoptic-scale features that typically exceed 2000 km in length with a length-to-width ratio greater than 2, and their associated extratropical cyclones can be skillfully forecast at lead times

of up to ~7 days (Wick et al. 2013; Guan and Waliser 2024). Similarly, operational forecasts of hurricane track and intensity, also synoptic scale phenomenon (hundreds to ~1000 km in diameter) are skillful out to ~5 days (Landsea and Cangialosi 2018; Kieu and Moon 2016).

In contrast, smaller-scale atmospheric features exhibit shorter skillful forecast lead times. Convective features embedded within the larger extratropical cyclone/atmospheric river systems that can trigger debris flows, such as squall lines (e.g., narrow cold-frontal rain bands), typically span only a few to tens of kilometers in horizontal extent (e.g., Houze Jr., 2014; Oakley et al. 2017; Collins et al. 2020; Schwartz et al. 2021; Zou et al. 2023). These finer-scale features fall within the mesoscale and generally have skillful forecast lead times on the order of hours to approximately one day (Markowski & Richardson, 2010); they have shorter skillful lead times than the larger synoptic context in which they occur. Accurately forecasting these mesoscale features requires high-resolution atmospheric modeling, with spatial resolutions ≤ 3 km and temporal resolutions ≤ 1 hour commonly needed to capture their evolution (e.g., Cannon et al. 2020; Oakley et al. 2023; Zou et al. 2023).

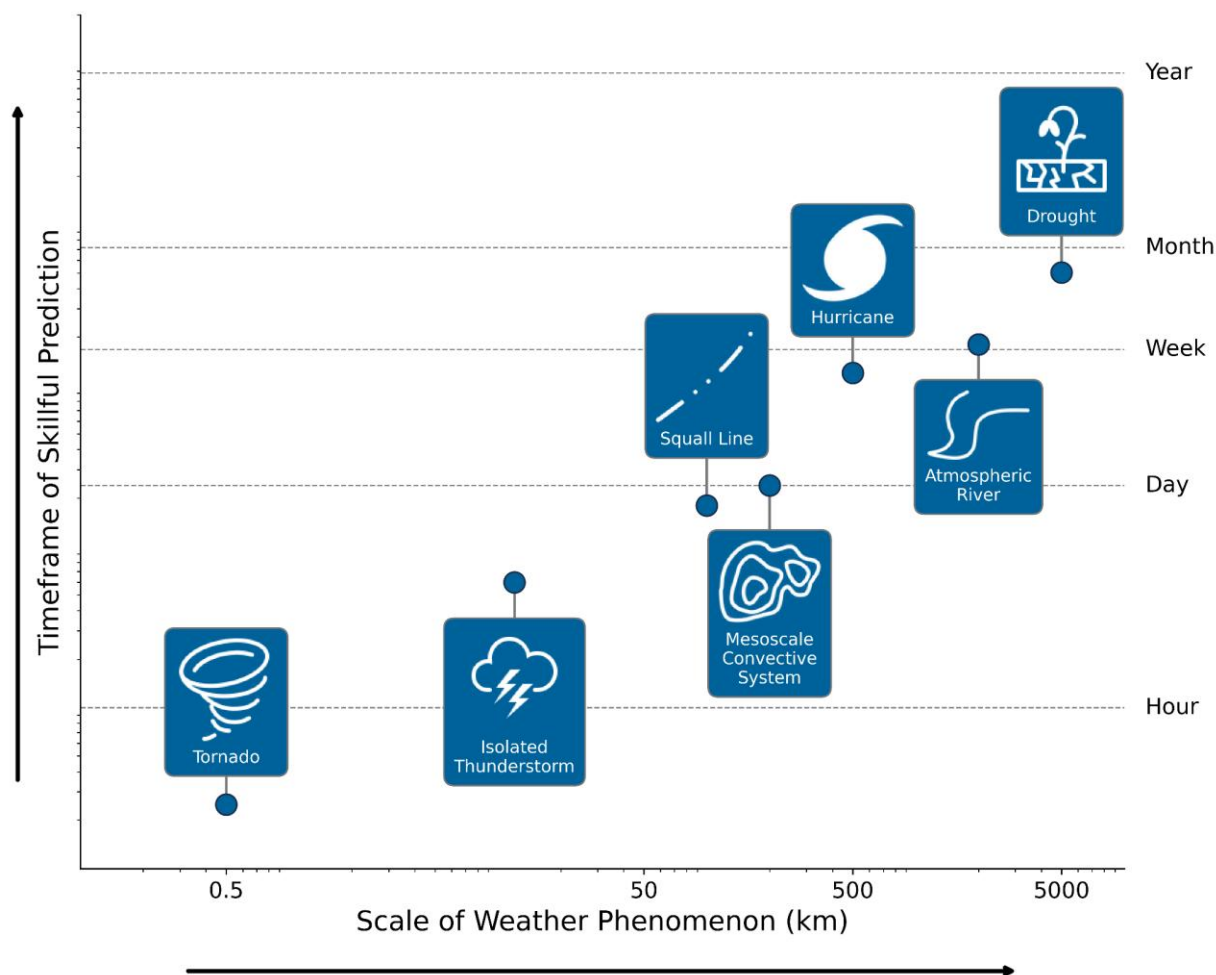


Fig. 1 Scales of predictability for several weather phenomena. Schematic illustrating the relationship between the characteristic spatial scale of selected weather phenomena (x-axis, km; log scale) and the approximate timeframe over which skillful prediction is possible (y-axis, hours; log scale). Points are representative estimates for tornadoes (200 m-2 km), isolated thunderstorms (2-20 km), squall lines, mesoscale convective systems (20-200 km), hurricanes (200-2000 km), atmospheric rivers (length > 2000 km), and drought (Markowski and Richardson, 2010; Su et al., 2023; Guan and Waliser, 2024). Dashed horizontal lines indicate reference predictability time scales from hours to years. Icons are included to aid visual identification of each phenomenon and are not to scale. Note that finer scale features, such as squall lines and tornadoes, typically have shorter forecast lead times than the larger synoptic features in which they occur.

While the general timeframes of skillful prediction of atmospheric phenomena across scales are known, it is valuable to quantify forecast performance for the particular events and region of interest. Most research assessing mesoscale model performance for hourly to sub-hourly precipitation in the US focuses on areas east of the Rockies with limited terrain and good radar coverage (e.g., James et al. 2022). This is typically due to the lack of a reliable gridded quantitative precipitation estimate (QPE) products at relevant spatial and temporal resolutions in the mountainous West that would facilitate evaluation of common forecast performance metrics (e.g., Bytheway et al. 2020, 2022; White and Nelson 2024; Sinclair et al. 2026). The existing precipitation forecast verification research for California primarily evaluates cool season multi-hour rainfall accumulation during atmospheric river storms (e.g., English et al. 2021; Delle Monache et al. 2025; Mascioli et al. 2026) or evaluates hourly warm season precipitation for a large region that includes California (Bytheway et al. 2025).

Our goal is to assess whether mesoscale models tend to over- or under-forecast the magnitude of hourly precipitation associated with PFDF events, by how much, and how the error varies with forecast lead time. To do this, we evaluated forecasts from two mesoscale weather forecast models against 1-hour accumulations at rain gauges for 10 PFDF events in central and southern California between 2021 and 2025. While rainfall accumulation over a 15-minute duration typically has the best relationship with postfire runoff response (Kean et al. 2011), hourly model output and gauge observations are more widely available and the 1-hour duration is most commonly used by weather forecasters in communications with partners and the public (e.g., Oakley et al. 2023; IEM 2026).

Emergency management actions, such as evacuations or road closures, are best implemented “before the rain is falling” and during daylight. This implies that often at least 24 hours of lead time is needed to act. For example, the evacuation order was issued roughly 30 hours ahead of the 2018 Montecito, CA, debris flow event (Hayden 2018; Goto et al 2021). Documenting forecast performance for PFDF-triggering rainfall at forecast lead times within and beyond this 24-hour window can improve situational awareness and understanding of the benefits and challenges of high-resolution precipitation forecasts for emergency managers and others

concerned with PFDF impacts. A better understanding of forecast uncertainty can support decision-making around evacuations, road closures, debris basin cleanouts, barrier installation, and other mitigation measures that may be taken ahead of a storm of concern. PFDF likelihood, volume, runout, and inundation models are very sensitive to rainfall inputs (Kean and Staley 2021; Oakley et al. 2023, Prescott et al. 2024; Barnhart et al. 2026). Thus, to effectively integrate forecast models with PFDF models, it is valuable to have context for the uncertainty in forecasts of short-duration, high-intensity rainfall.

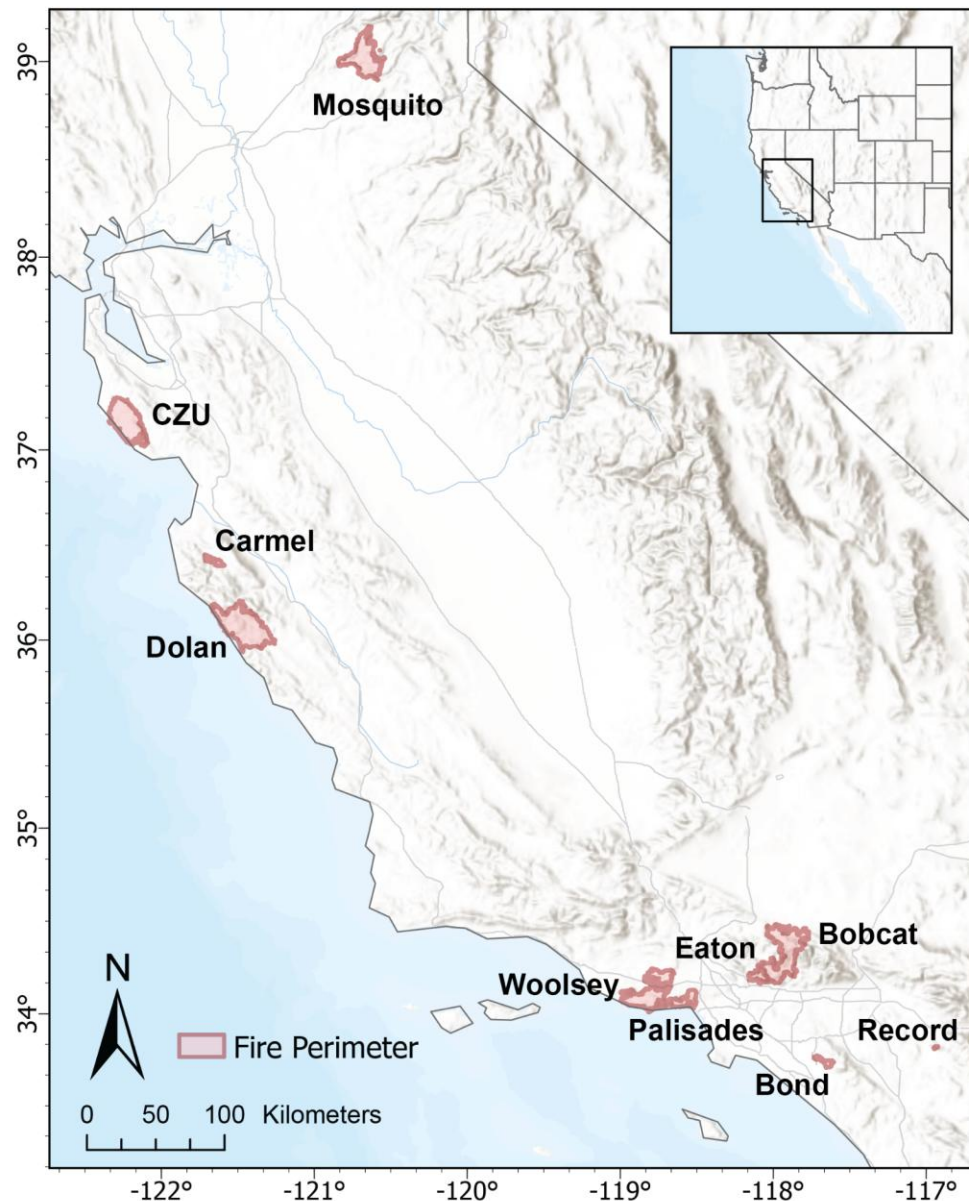


Fig. 2 Map of fire perimeters for events listed in Table 1. Fire perimeters sourced from CAL FIRE (2026).

2. Methods

2.1 Postfire flood and debris flow event selection

We identified PFDF events, or dates on which one or more PFDF responses occurred within a three-hour window on an individual burn area, for the forecast evaluation. We did not require PFDFs to exceed a certain volume or level of impact. All events occurred in Central or Southern California. This area experiences frequent PFDFs that often have well-constrained timing.

The events included in this analysis occurred during the cool season (October-March) between January 2021 and December 2025. PFDFs initiated within areas that had burned within five years preceding each event. PFDF timing could be constrained within three hours for all events and there were at least three rain gauges available in the vicinity of the burn area that recorded peak hourly rainfall exceeding 10 mm h^{-1} during the 3 h window. Screening by these criteria resulted in 10 events occurring on 10 unique wildfire burn areas during six unique storms (Table 1).

Constraining events to 2021 and later allows us to examine two mesoscale weather forecast models in consistent versions (Section 2.3). We consider only events in the cool season as one of the models of interest is only available during this period. PFDFs typically occur within the first five years following wildfire (e.g., McGuire et al. 2024), thus we focus on events in this window. Table 1 provides details on each event's timing and response characteristics.

Event Date (UTC)	Window Center Time (UTC)	Fire Name (Year)	Response Type	Impacts	Storm Description
27 Jan 2021	0900	CZU (2020)	debris flow ¹	One remote small-volume debris flow, no impacts.	Moderate AR (AR2), squall line along cold front ¹
27 Jan 2021	1200	Carmel (2020)	flood, debris flood, debris flow ¹	One debris flow, multiple other minor responses. Debris impacted rural roads, sediment impacted fisheries.	Moderate AR (AR2), squall line along cold front ¹
28 Jan 2021	0500	Dolan (2020)	debris flood, debris flow ¹	Multiple debris floods and flows. Several large debris flows, major impacts to Highway 1.	Moderate AR (AR2), squall line along cold front ¹
14 Dec 2021	1900	Bond (2020)	debris flood, debris flow ¹	One high-impact debris flow blocked roads and knocked a home off its foundation.	Moderate AR (AR2), squall line along cold front ¹
14 Dec 2021	1500	Bobcat (2020)	flood ¹	High-energy flood flows observed, debris flow occurrence could not be confirmed.	Moderate AR (AR2), squall line along cold front ¹

30 Dec 2021	1100	Woolsey (2018)	debris flood, shallow landslide ^{1,3}	Debris floods and shallow landslides that mobilized into debris flows. Mud and rocks on roads, road closures.	Cutoff low, no AR, broad area of convective showers ⁴
9 Jan 2023	1700	Mosquito (2022)	debris flood, debris flow ¹	One confirmed debris flow discharged into Oxbow Reservoir; two other responses that could not be confirmed as debris flood or debris flow.	Moderate AR (AR3), squall line along cold front ¹
13-14 Feb 2025	0100	Palisades (2025)	flood, debris flood, debris flow ²	One debris flow overtopped Topanga Cyn. Blvd. Another overtopped Pacific Coast Hwy., pushing an emergency response vehicle off the road. Floods and debris floods also impacted roads.	Weak AR (AR1), squall line along cold front ⁵
13-14 Feb 2025	0100	Eaton (2025)	flood, debris flood and debris flow ²	Several debris structures filled and overtopped. Several debris basins partially filled and overtopped, sending clearwater flow down spillways. Mud and debris removal from properties and roadways.	Weak AR (AR1), squall line along cold front ⁵
13 Mar 2025	1400	Record (2024)	debris flow ²	Two debris flows overtopped the road, trapping vehicles in mud and debris.	No AR, squall line along cold front ⁶

Table 1 Information on each PFDF event evaluated in this study. The first column provides the date of the event in UTC. The second indicates the center time in UTC of the 3 h or 5 h temporal window applied in the forecast assessment. The third column provides the fire name and year of ignition. The fourth column classifies the response type as flood, debris flood, debris flow or shallow landslide following descriptions provided in Oakley et al. (2025). A superscript of 1 in this column indicates documentation of the event in Oakley et al. (2025) and references therein. A superscript of 2 indicates that documentation of this event is available in California Geological Survey (2026), and a superscript of 3 indicates documentation in Sawicki (2021). The fifth column provides a description of the impacts of the event. The final column describes the storm conditions associated with the event. “AR” indicates atmospheric river, and value in parentheses (e.g., AR2) indicates strength of AR on the AR scale of 1-5, where 5 is strongest (Ralph et al. 2019). A superscript of 1 in this column indicates storm information sourced from Oakley et al. (2025) and references therein. Sources for superscripts 4, 5, and 6 are CW3E (2022), CW3E (2025a), and CW3E (2025b), respectively. All events were also evaluated using archive composite radar reflectivity imagery (CNRFC 2026), archive surface analysis (NOAA WPC 2026), and archived National Weather Service Area Forecast Discussions (IEM 2026).

2.2 Rain gauge selection

We used rain gauge data primarily from federal, state, county, and university research networks as well as a few gauges from the Citizen Weather Observer Program accessed through sources

documented in Supplementary Information A. We considered stations within the burn area and/or within a 4-km radius of the fire perimeter. The 2025 Record Fire burned a very small area; thus, we extended our criteria to 10 km to incorporate the nearest three rain gauges. As available gauges are often not collocated with precise locations where a PFDF response was observed, the observations here do not necessarily define the triggering intensity of the response. There can be high spatial variability in rainfall intensity across the burn area. However, understanding model performance at points with gauge observations can give insight into forecast performance at other ungauged points in the burn area.

We required each event to have at least three rain gauges where the hourly intensity exceeded 10 mm h⁻¹ during the event window, which we defined as a three hour time period during which a PFDF response was observed, such that we had multiple data points to define performance for each individual event. The 10 mm h⁻¹ threshold serves as a general minimum threshold for rainfall intensities conducive to postfire response. Staley et al. (2020) found that the average triggering rainfall at a 1-hour duration was equivalent to a 1-year recurrence interval storm, which can be as low as ~10 mm h⁻¹ in Southern California.

These stations primarily use unshielded tipping bucket rain gauges, though equipment cannot be confirmed in many cases. Tipping bucket rain gauges are prone to underestimation of precipitation depending on equipment, siting, and environmental conditions (Dunn et al. 2025). We manually quality controlled data for each event, considering whether hourly values were reasonable given atmospheric conditions, nearby gauge observations, radar observations, and climatology of the area.

2.3 Mesoscale forecast model selection

We assessed two mesoscale models for their representation of hourly rainfall during the 10 PFDF events, namely the High-Resolution Rapid Refresh (HRRR, version 4) and WestWRF.

The HRRR is a National Oceanic and Atmospheric Administration real-time 3-km horizontal resolution convection-allowing atmospheric model. The HRRR produces hourly precipitation forecasts out to 18 hours every hour and out to 48 hours four times daily at 00, 06, 12, and 18 UTC. Herein, we assessed these 48-hour forecasts at 6h intervals. HRRR version 4, the final version, was implemented in December 2020 (Benjamin et al. 2016; Dowell et al. 2022). The HRRR is one of the convection-allowing models commonly used by the National Weather Service to forecast high-intensity rainfall. While the HRRR was superseded by Rapid Refresh Forecast System (RRFS) in 2026 (NOAA 2026; Banos et al. 2022), it is still valuable to assess HRRR for these events given its past operational value and available archive. We downloaded the HRRR hourly accumulated precipitation from the NOAA Amazon Web Services archive (<https://registry.opendata.aws/noaa-hrrr-pds/>) using the Herbie Python package (Blaylock 2026).

The WestWRF is the Center for Western Weather and Water Extremes (CW3E) near real-time, convection-allowing, 3-km horizontal resolution, regional weather prediction model based on the WRF-ARW version 4.4.1 (Skamarock et al. 2019) and tailored to western US hydrologic extremes (Martin et al. 2018). WestWRF is supported in part by the California Department of Water Resources to inform on flood hazards and water resource management in the state. The model is run October-March and provides hourly forecasts out to 120 hours twice daily at 00 and 12 UTC. Here, we used WestWRF initialized with the NOAA Global Forecast System. The WestWRF archives extend back to 2021. See data availability statement.

We selected PFDF events from 2021 onward as archive forecasts for these events are available for both HRRRv4 and WestWRF. As we only evaluated two models, our goal was not to determine which of all available models is best for forecasting PFDF-triggering rainfall, but rather to develop a general understanding of how mesoscale forecast models represent PFDF-triggering rainfall.

We used data from each model initialization running forward through the representative event time (Table 1). Lead time varies across events since event times vary relative to the fixed model initialization times and timesteps. When we aggregated data across all events, we binned forecast lead times into groups representing the first, second, third, and so on, closest lead time to the event. The shortest lead times range from 3-9 h for the 6-hourly updated HRRR and from 3-15 h for the 12-hourly updated WestWRF. We needed a minimum of 3 hours lead time in both cases to compare forecast rainfall within a temporal window around the event time, as described in the following section. The maximum duration of the temporal windows applied is 7 h, requiring at least 3 h of forecast lead time prior to the event time for the analysis.

2.4 Spatial and temporal windowing

To avoid over-penalizing a high spatial and temporal resolution weather forecast model, a temporal windowing and spatial “neighborhood” approach can be applied (e.g., Ebert 2008; English et al. 2021; Gaudet et al. 2024). While it may be difficult for a high-resolution forecast model to exactly match an observed value at a point, comparing the observation with modeled values close in time and/or in nearby grid cells allows for a flexible evaluation that permits the user to determine at what scales the model is skillful.

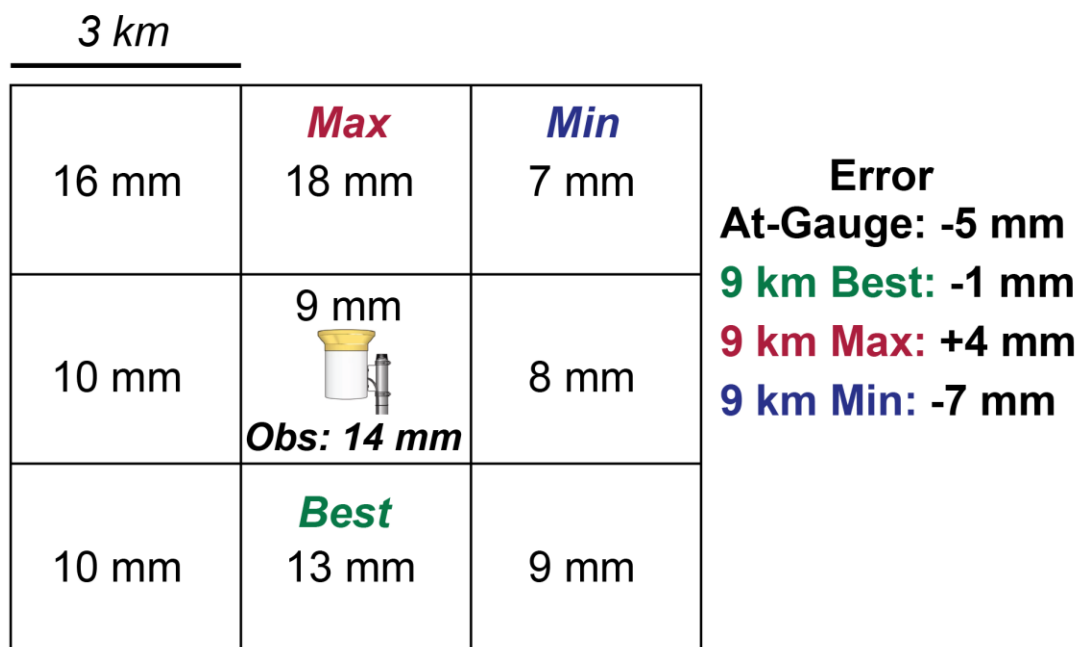


Fig. 3 Example of a 9 km neighborhood assessment for a given forecast lead time. The rain gauge observation lies in the center grid cell, indicated as “Obs”. The grid cell where the gauge is located as well as neighboring grid cells all have forecast 1 h precipitation accumulation values. The error is the difference between the forecast value and the gauge observation.

Our analysis compared the modeled and observed peak 1-hour rainfall intensity. We computed 1-hour rainfall intensities from gauge observations on fixed-hour intervals for comparison with rainfall intensities derived from fixed-hour model timesteps. When comparing modeled and observed intensities, we considered windows of 3 h, 5 h, and 7 h. The windows are centered on a representative time ascribed to each event (Table 1). In space, we consider the grid cell in which the rain gauge is located, a 9 km neighborhood (3x3 grid cell matrix, Fig. 3), and a 15 km neighborhood (5x5 grid cell matrix). Going beyond the 15 km neighborhood begins to exceed useful skill for considering impacts on a particular burn area and using a window greater than 7 hours limits our ability to quantify variations in model performance at short lead times with models that are updated every 6 to 12 h. We assessed the space and time neighborhoods for the value closest to what was observed at the gauge (“best” match).

Additionally, we compared the gauge observation with the maximum, minimum, and mean modeled rainfall within each temporal window and spatial neighborhood. The maximum forecast intensity within the window provides information on the most intense rainfall that the model produces in the area. Maximum modeled values within a window could be viewed from a forecast perspective as a worst-case scenario. It would therefore be beneficial to know if this modeled metric is typically less than or greater than the observation.

2.5 Forecast evaluation metrics

We evaluated forecasts using three quantitative metrics: error, percent error, and probability of detection.

At each gauge, and for each temporal window and neighborhood size at each forecast lead time for the event, we determined the forecast 1-hour intensity of interest from the spatial and temporal window, i_{60}^{model} . Metrics of interest included the value within the window that is closest to the observed gauge value, i_{60}^{obs} , as well as the maximum, minimum, and mean within the temporal window and spatial neighborhood.

Error is then calculated as:

$$Error = i_{60}^{model} - i_{60}^{obs}.$$

Percent error is calculated as:

$$Percent\ Error = (1 - (i_{60}^{model} \div i_{60}^{obs})) * 100.$$

Unless noted otherwise, the model metric i_{60}^{model} is the value within the window that is closest in absolute value to the observed gauge value.

As we are only evaluating forecasts pertinent to an observed event, we cannot calculate many of the typical metrics used to evaluate weather forecast models (e.g., Roebber 2009; Gaudet et al. 2024). In a 2x2 contingency table, we can only calculate a true positive (i.e., an event was forecast and observed) and a false negative (i.e., an event was observed but not forecast). We cannot calculate a false positive (i.e., an event was forecast but not observed) or a true negative (i.e., an event was not forecast and not observed). The forecast metric we can calculate is Probability of Detection:

$$POD = \frac{true\ positives}{true\ positives + false\ negatives} = \frac{true\ positives}{observed\ events}$$

We consider a true positive to be when both $i_{60}^{obs} \geq 10\ mm\ h^{-1}$ and $i_{60}^{model} \geq 10\ mm\ h^{-1}$ and a false negative when $i_{60}^{obs} \geq 10\ mm\ h^{-1}$ and $i_{60}^{model} < 10\ mm\ h^{-1}$.

We produced maps for each PFDF event and for each model to visually represent the spatial variations in forecasts of high-intensity rainfall. The maps show the peak 60-minute accumulation in the 3 h event window for each forecast lead time as well as the gauge locations and perimeter of the wildfire where the PFDFs initiated.

We developed an event-level error index to facilitate comparisons among the 10 PFDF-triggering storm events to determine how storm characteristics, precipitation intensity, and other factors may relate to model performance. Specifically, we defined the event-level error index as the mean of the forecast error among all gauges for an event at the two shortest lead times for

WestWRF and four shortest lead times for the HRRR. This index facilitates comparisons of model performance across events for the ≤ 24 -hour lead time, when we expect forecast skill to be greatest.

3. Results

3.1 Sensitivity to spatial and temporal window size

We evaluated the sensitivity of results to different window sizes and selected two combinations for which to display results (Table 2). While increasing the spatial window from 9 km to 15 km generally reduces error, we determined that a 15 km neighborhood, or 225 km² area, is too large to be relevant to evaluating rainfall associated with PFDF hazards at the gauge location. Since PFDFs are triggered by intense rainfall over a burned area, the spatial scale of the neighborhood should be relevant to the area typically affected by a fire. Most fires evaluated for PFDF hazards in California are < 225 km² in area (US Geological Survey 2026), suggesting a 15 km neighborhood could often include substantial area outside of a fire perimeter. There is a substantial reduction in error from the 3 h to 5 h window when using a 9 km neighborhood, but only modest improvement when expanding from a 5 h to 7 h window. Thus, we present results for the 9 km neighborhood and 5 h window. In addition, we present results using the at-gauge neighborhood (i.e., a single model grid cell) and 3 h window since this is of interest as the most restrictive evaluation metric. Figs. S1 and S2 compare four of these windows across all forecast lead times for WestWRF and HRRR, respectively.

Space, Time Window	WestWRF Median error (%) 3-15 h lead	WestWRF Median error (%) 39-51 h lead	HRRR median error (%) 3-9 h lead	HRRR median error (%) 39-45 h lead
At gauge, 3 h	-26	-53	-35	-44
At gauge, 5 h	-21	-27	-23	-33
At gauge, 7 h	-21	-27	-23	-20
9 km, 3 h	-10	-42	-15	-21
9 km, 5 h	-9	-17	-13	-11
9 km, 7 h	-9	-17	-11	-6
15 km, 3 h	-4	-34	-9	-15
15 km, 5 h	-3	-8	-2	-8
15 km, 7 h	-3	-8	-2	-2

Table 2 Comparison of median percent error for WestWRF and HRRR for best fit to gauge observation at two different lead times across nine spatial and temporal windows. The rows of time windows/neighborhood sizes selected for analysis are in bold.

3.2 WestWRF evaluation

WestWRF forecast lead times range from 3-15 h at the shortest (lead time 1) out to 99-111 h at the longest (lead time 9) at 12-hour intervals (Fig. 4). The forecast evolution (change in error magnitude with lead time) varies in each event as well as individual gauges within an event. Results described in the main text reflect the 9 km neighborhood, 5 h window analysis.

WestWRF 1 h precipitation forecasts were generally less intense than observations across all lead times. The median event-level error was -1.4 mm at lead time 1 (3-15 h), -1.3 mm at lead time 2 (15-27 h), -2.5 mm at lead time 3 (27-39 h), and became increasingly negative at longer lead times; the longest lead time (99-111 h) had the greatest median event-level error at -12.4 mm (Fig 4a). Event-level percent error followed a similar pattern of -9% at the 3-15 h lead time, -6% at 15-27 h, -14% at 27-39 h, and a maximum median event-level percent error of -70% at 99-111 h (Fig 4b). Median POD is between 0.8 and 1 for the first four lead times, then decreases to between 0 and 0.3 from lead times 5 (51-63 h) to 9 (99-111 h). The minimum POD is 0 for all lead times except the two shortest, where it is 0.2 and 0.3, respectively (Fig. 4c).

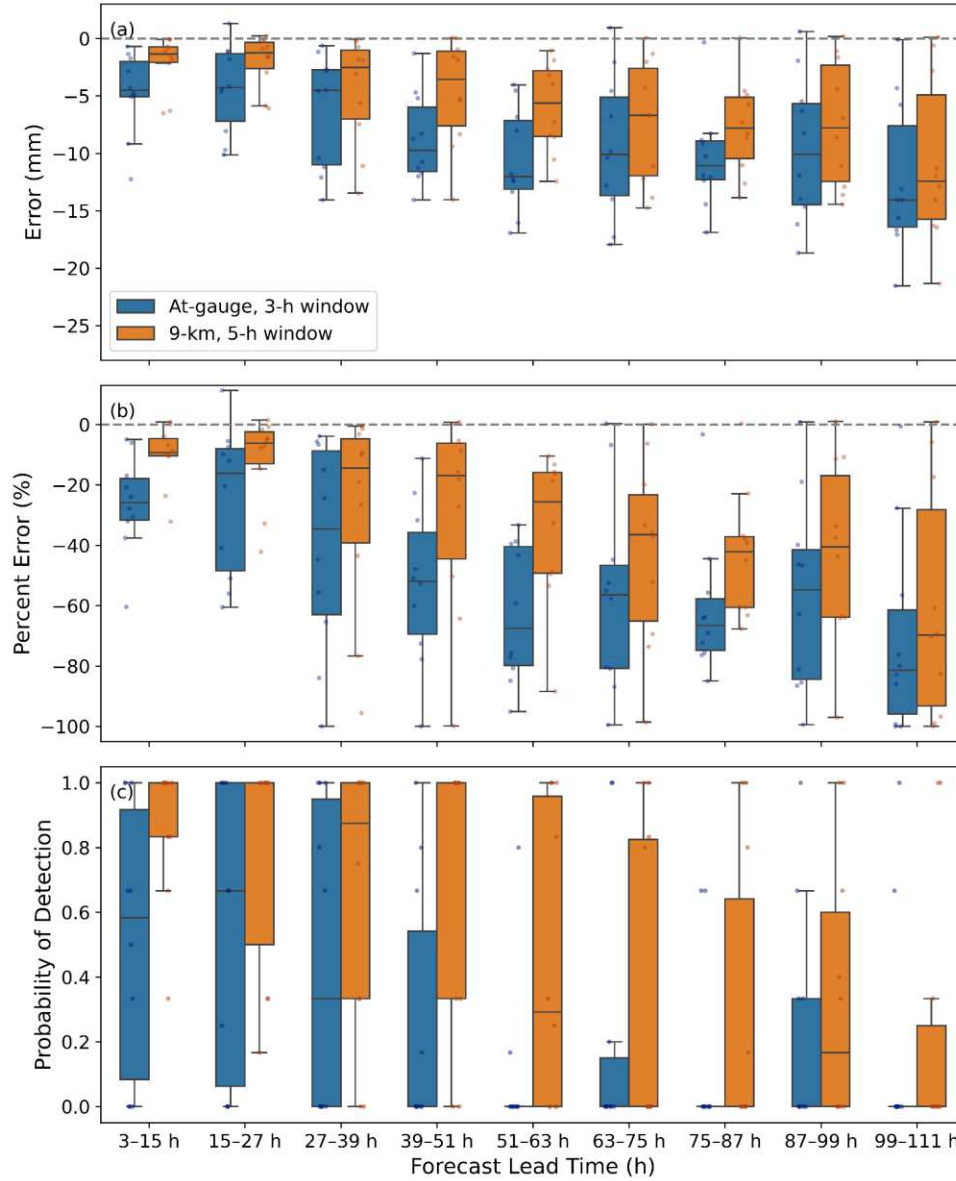


Fig. 4 Boxplots representing error (a) and percent error (b) and probability of detection (c) among stations in all 10 events for the WestWRF analysis. Blue box plots represent verification based on the grid cell in which the gauge is located and a 3 h time window. Orange box plots represent a 9 km neighborhood and 5 h time window. Whisker extent is 1.5 times the interquartile range. All three plots use the “best” fit value in the space/time window. The blue and orange dots represent the mean of error, percent error, and POD in their respective plots among gauges for each of the 10 events, with color corresponding to the spatial/temporal window used.

In this retrospective analysis, we considered the best fit in a neighborhood to objectively define how the model performed relative to the rain gauge observation. However, when forecasters are considering a forecast at a point, they do not yet know which nearby grid cell in the forecast model will perform “best.” They need to consider whether the forecast value at the location of

interest or some other statistic, such as maximum, minimum, or mean of adjacent grid cells, is the most likely outcome. Fig. 5 shows that maximum among grid cells tends to be closer to best fit than the mean or the minimum. However, there is potential for the maximum to be much greater than the best fit. At lead time 1 (3-15 h) the spread of the best and maximum are quite close (within a few mm/%) for event-level error, percent error, and POD.

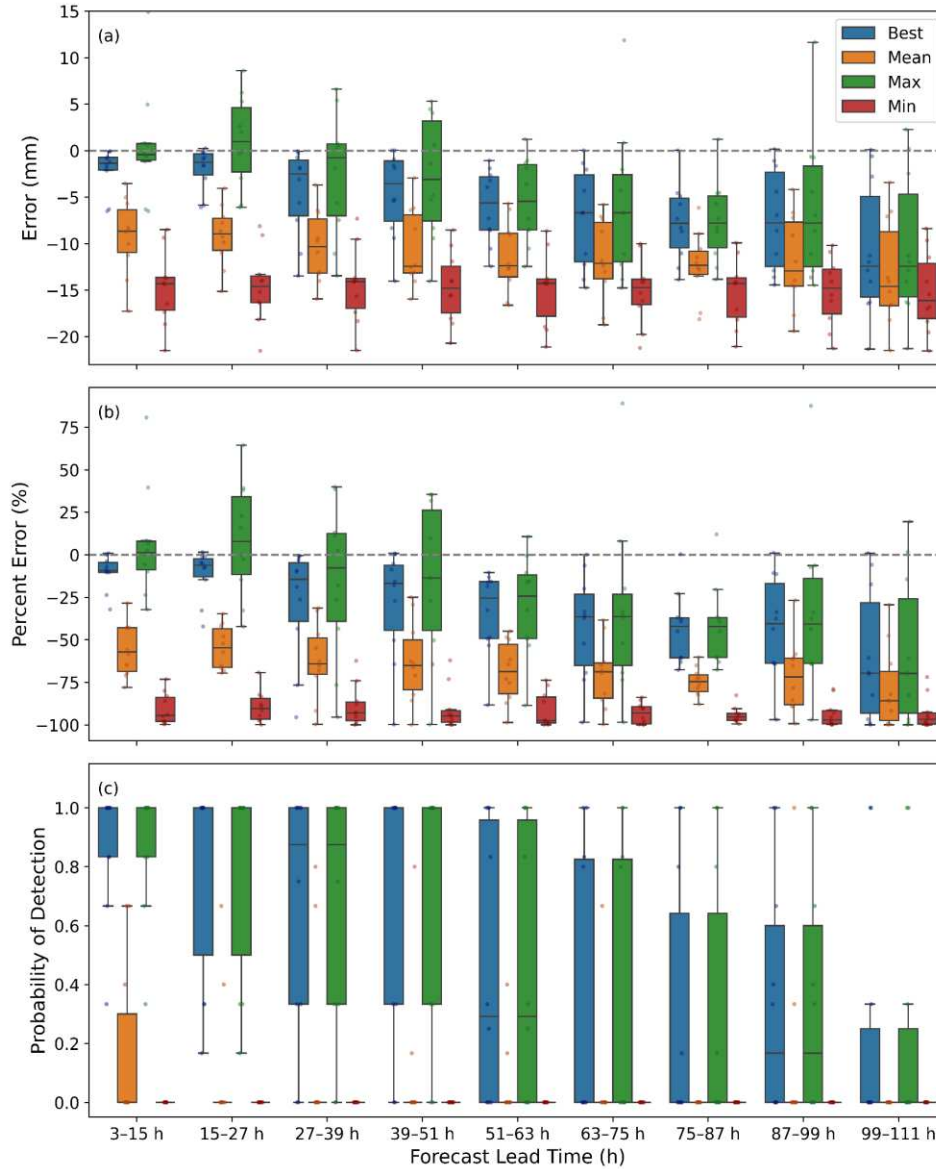


Fig. 5 Boxplots representing median error (a), median percent error (b), and median probability of detection (c) for the WestWRF model for the 10 PFDF events when using the 9 km neighborhood and 5 h time window. Blue bars represent results considering best fit value in the neighborhood (as in orange bars in Fig. 4). Orange, green, and red bars show model performance using the mean, maximum, and minimum of all values in the neighborhood/window, respectively. Whisker extent is 1.5 times the interquartile range. The dots represent the mean of error, percent error, and POD in their respective plots among gauges for each of the 10 events, with color corresponding to the metric considered.

Similar to the event composites (Fig. 4), error is typically smallest at the two shortest lead times for individual events. At lead time 1, all but two events (Carmel and Palisades) had at least two observations where the absolute forecast error was <2 mm. At lead time 2 (15-27 h), all but three events (Carmel, Dolan and Woolsey) had at least two observations where the absolute forecast error was <2 mm. Figures 6 and 7 provide examples of WestWRF forecast evolution for stations for the Dolan and Eaton Fire events; similar figures for the other eight events are presented in Supplementary Information C (Fig. S3-S10).

There were several cases where individual station forecasts are close to the observed value at a long lead time. For example, the forecast for station SANTPK in the Bond Fire was ≤ 1 mm in error at all lead times except 8 (91 h, -1.4 mm) and 9 (103 h, -6.5 mm) (Fig. S5). Two stations on the CZU Fire event (Fig. S3, BOULDER, BIG) and two on the Eaton Fire event (Fig. 7, F2444, HNGC1) had <1 mm of error from lead time 5 (~60 h) through the time of event.

In contrast, there were cases where forecast error was large at an individual station until the shortest lead time. The forecast for station LOWC1 in the Dolan Fire event (Fig. 6) is more than 12 mm too low at nearly all lead times excepting the shortest lead time. Likewise, for the Eaton Fire event, stations DVLC1 and SIRC1 both had large errors for most or all of the forecast period (Fig 7b). In both events, the observed rainfall at the noted stations is >20 mm h^{-1} and higher than other observations for those events, suggesting the model may struggle to represent higher intensities.

Another interesting case was the Bobcat event (Fig. S6) where the forecast was <1 mm in error for all stations at the 91 h lead time. Error for this event increased with decreasing lead time, with >10 mm in error at two stations at lead times 3 and 4 (31 h and 43 h). However, at the two shortest lead times (7 h and 19 h), the forecast improved such that all but one station observation was forecast within 1 mm of error.

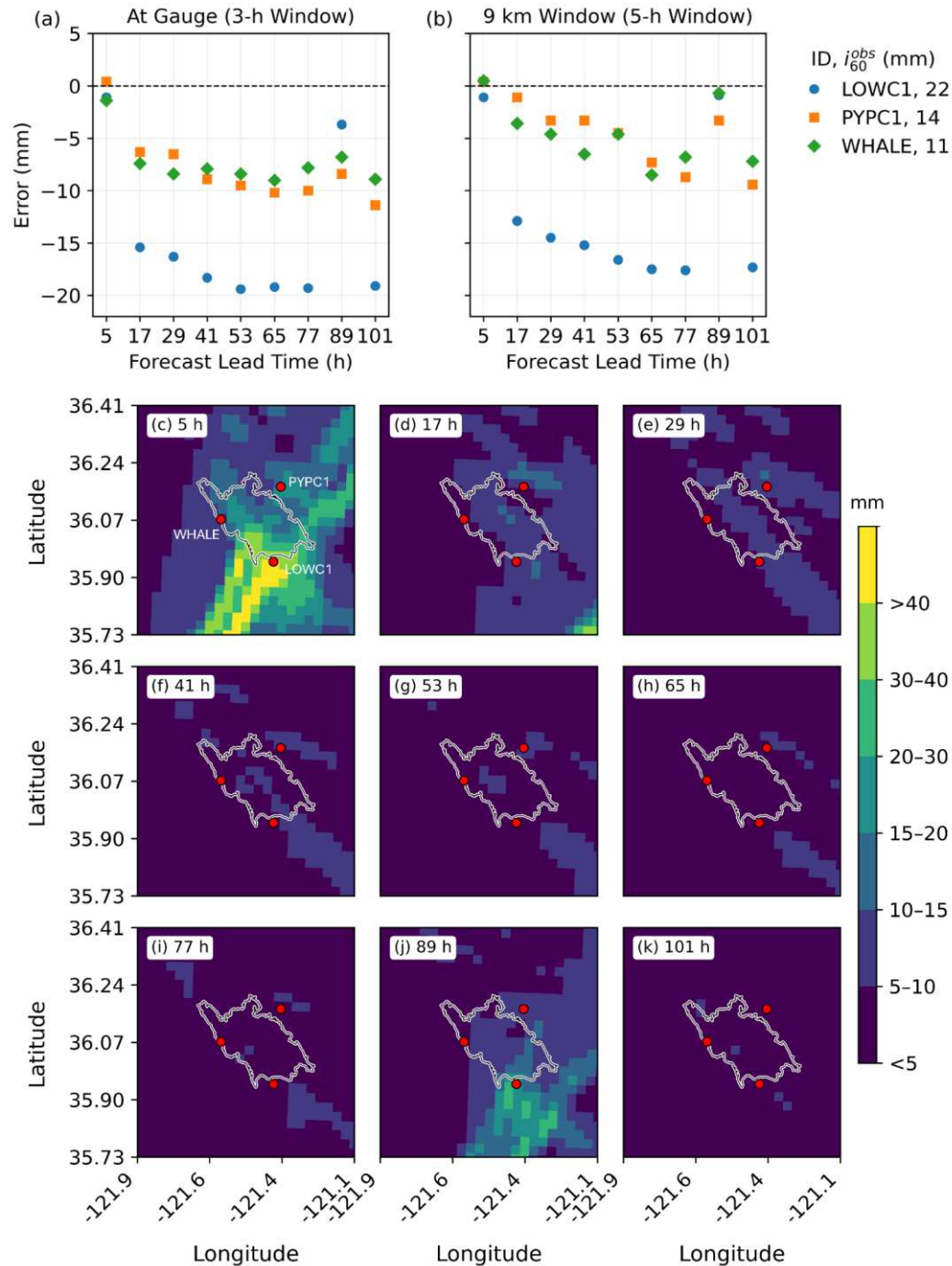


Fig. 6 Analysis of WestWRF for the Dolan Fire event, 27 January 2021. Forecast error (mm) for a) model grid cell associated with gauge location and 3 h temporal window, and b) a 9 km neighborhood of model grid cells and 5h temporal window. The forecast is compared to three rain gauge stations in the vicinity of the burn area. Station IDs and associated peak 1 h rainfall are given in the upper right of the figure. c–k) A spatial depiction of peak 1 h rainfall for the 3 h event window (Table 1) at each forecast lead time. Colorbar indicates the 1 h rainfall accumulation, in mm. The fire perimeter is indicated by a grey outline, and red dots indicate the location of each station evaluated.

437

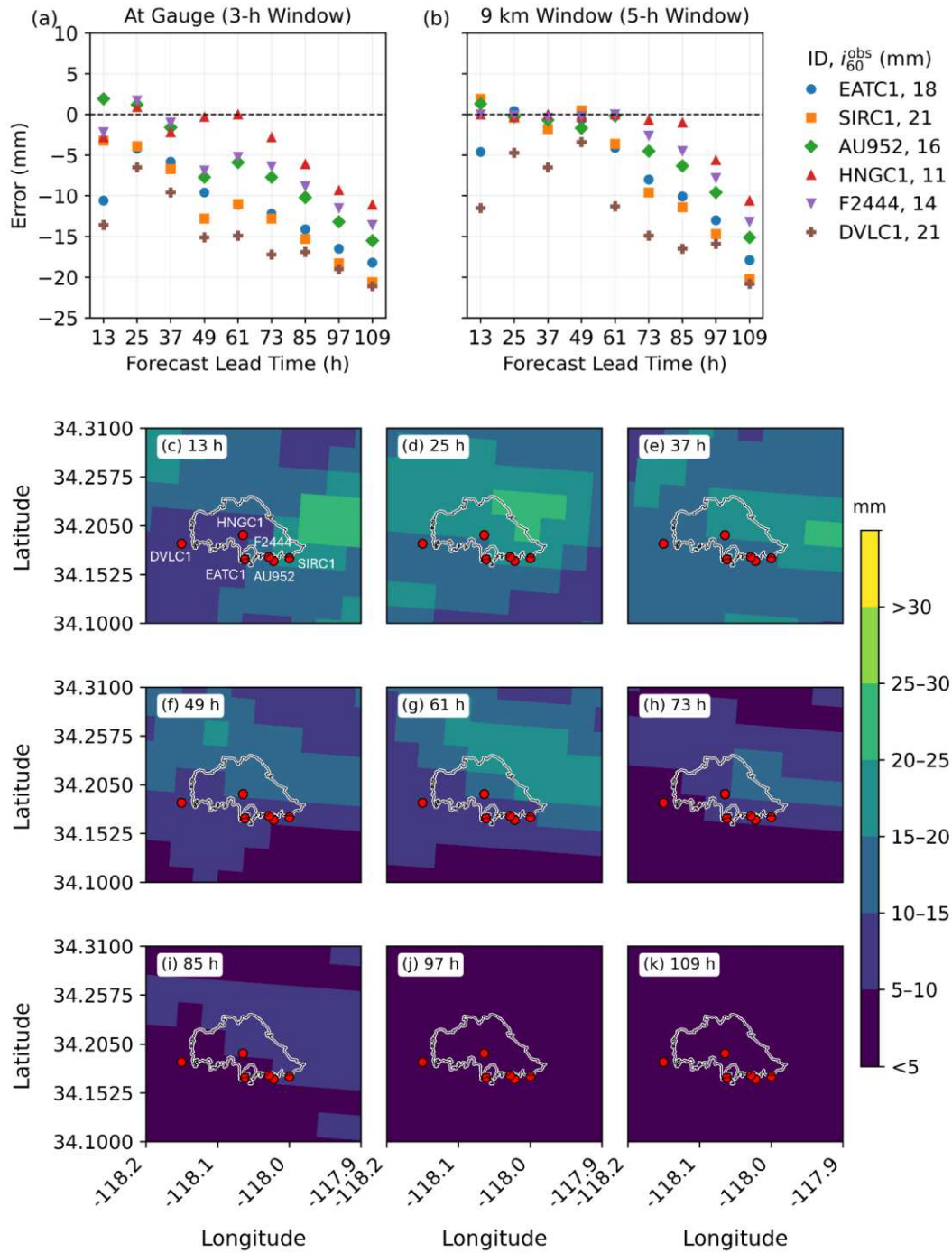


Fig. 7 As in Figure 6, but for WestWRF forecast of the 13 February 2025 Eaton Fire event.

3.3 High Resolution Rapid Refresh evaluation

HRRR forecast lead times ranged from 3-9 h at lead time 1 out to 39-45 h at lead time 7 at 6-hour intervals (Fig. 8). Similar to the WestWRF results, forecast evolution varied among events

and individual gauges within an event, though overall showed a general improvement in forecast with decreasing lead time.

Similar to WestWRF, the HRRR 1 h precipitation forecasts were generally less intense than the observed values across all lead times (Fig 8a,b). As with WestWRF, median error was not smallest at lead time 1 (3-9 h, -2.6 mm), but rather at lead time 3 (15-21 h, -2 mm). In contrast to WestWRF results, where median error was greatest at the longest lead time, the median error among events was greatest at lead times 4 and 6 (-3.8 mm at 21-27 h and 33-39 h, respectively) and slightly less at the longest lead time 7 (39-45 h, -3 mm; Fig 8a). The median of event-level percent error was lowest, -10.6%, at lead time 3 (15-21 h) and only slightly greater, -10.9%, at lead time 7 (39-45 h). The median of event-level percent error varied from -12.7 to -22.8% at other lead times (Fig. 8b).

The median of event-level POD varies from 0.8 to 1 at the four shortest lead times (3-9 h to 21-27 h), reduces to 0.7 and 0.5 at lead times 5 and 6 (27-33h, 33-39 h), respectively, and increases to 1 at lead time 7 (39-45 h). The minimum POD is 0 for all lead times except the shortest, where it is 0.2 (Fig. 8c).

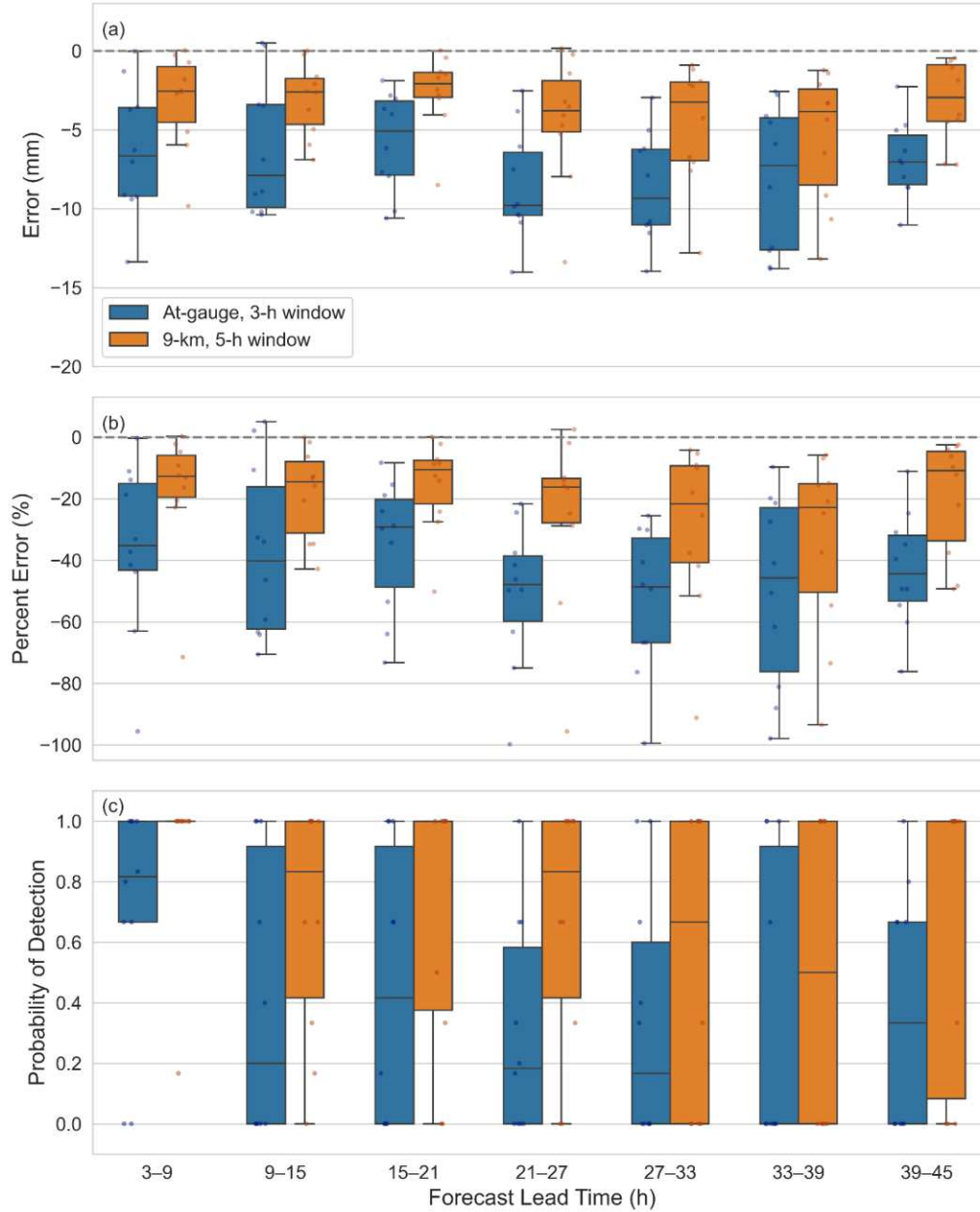


Fig. 8: Boxplots representing error (a) and percent error (b) and probability of detection (c) among all stations in all 10 events for the HRRR analysis. Blue box plots represent verification based on the grid cell in which the gauge is located and a 3h time window. Orange box plots represent a 9 km neighborhood and 5 h time window. Whisker extent is 1.5 times the interquartile range. Both plots use the “best” fit value in the space/time window. The blue and orange dots represent the mean of error, percent error, and POD in their respective plots among gauges for each of the 10 events, with color corresponding to the spatial/temporal window used.

Comparison of the event-level error, percent error, and POD using the maximum, minimum, mean, and best fit in the 9 km neighborhood and 5 h time window for HRRR yields similar results as with WestWRF. Event-level error using the maximum value within the neighborhood

is similar to that obtained using the best fit value. Event-level error and percent error computed using the maximum neighborhood intensity are more symmetrically distributed around zero compared to when the best fit value is used. This indicates that the best fit value tends to be an underestimate whereas the peak intensity provides a better balance of overestimation and underestimation across events.

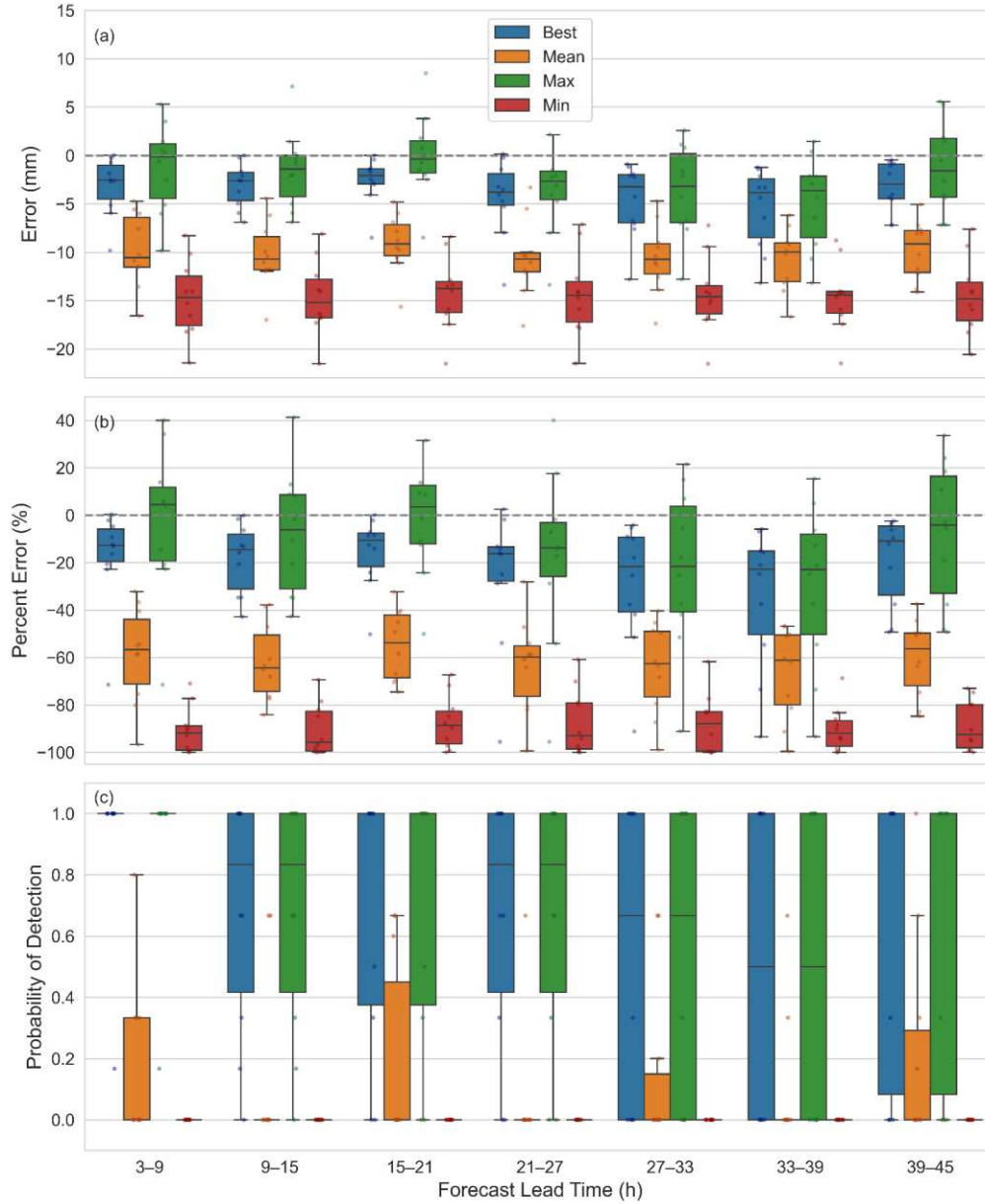


Fig. 9 Boxplots representing median error (a), median percent error (b), and median probability of detection (c) for the HRRR model for the 10 PFDF events when using the 9 km neighborhood and 5 h time window. Blue bars represent results considering best fit value in the neighborhood (as in orange bars in Fig. 4 above). Orange, green, and red bars show model performance using the mean, maximum, and minimum of all values in the neighborhood/window, respectively. Whisker extent is 1.5 times the interquartile range. The dots represent the mean of error, percent error, and POD in their respective plots among gauges for each of the 10 events, with color corresponding to the metric considered.

At the shortest lead time (3-9 h), all but three events (Carmel, Woolsey, and Bobcat) had at least two stations with <2 mm of error. Figures 10 and 11 provide examples of HRRR forecast evolution for gauges for the Dolan and Eaton Fire events; similar figures for the other eight events are presented in Supplementary Information C (Fig. S11-S18).

The HRRR results also had several cases where the forecast at an individual station is close to the observed value for the full forecast period, out to 39-45 h. For example, forecast error was less than 0.5 mm at all lead times for station SANTPK on the Bond Fire (Fig. S13) and station HNGC1 on the Eaton Fire (Fig. 11b). Station LPRC1 on the Carmel Fire has less than 0.5 mm of error until lead time 1, where the forecast is -1.5 mm in error (Fig. S12). The forecasts for the Record (Fig. S18), Dolan (Fig 10b), and Bond (Fig. S13) Fires displayed an interesting behavior. One or more rainfall observations for these events are forecast with little error at the longest lead time, 7. The forecast error then generally increases at lead times 2-6 before improving to <0.5 mm error at lead time 1.

Consistent with the WestWRF results, stations with the highest intensity precipitation observations, especially those over 20 mm h^{-1} , often had greater forecast error throughout the forecast period than stations that recorded lower intensities. For example, LOWC1 on the Dolan Fire is forecast 15 mm too low at several lead times, though <1 mm of error at lead time 1 (5 h; Fig. 10b). On the Eaton Fire, the forecast for DVLC1 is more than 5 mm too low at nearly all lead times, and still -4.6 mm too low at lead time 1 (7 h, Fig. 11b). During the same storm on the Palisades Fire, the forecast for station MBUC1, which recorded 32 mm h^{-1} , is more than 9 mm too low for the entire forecast period (Fig. S17). In other cases, the highest precipitation observations had variable forecast errors with lead time. Error magnitudes at stations BROOK, GREEN and SCOTT on the CZU Fire, which all reported over 20 mm h^{-1} , vary from 5-10 mm at some lead times to <1 mm at others (Fig. S11b).

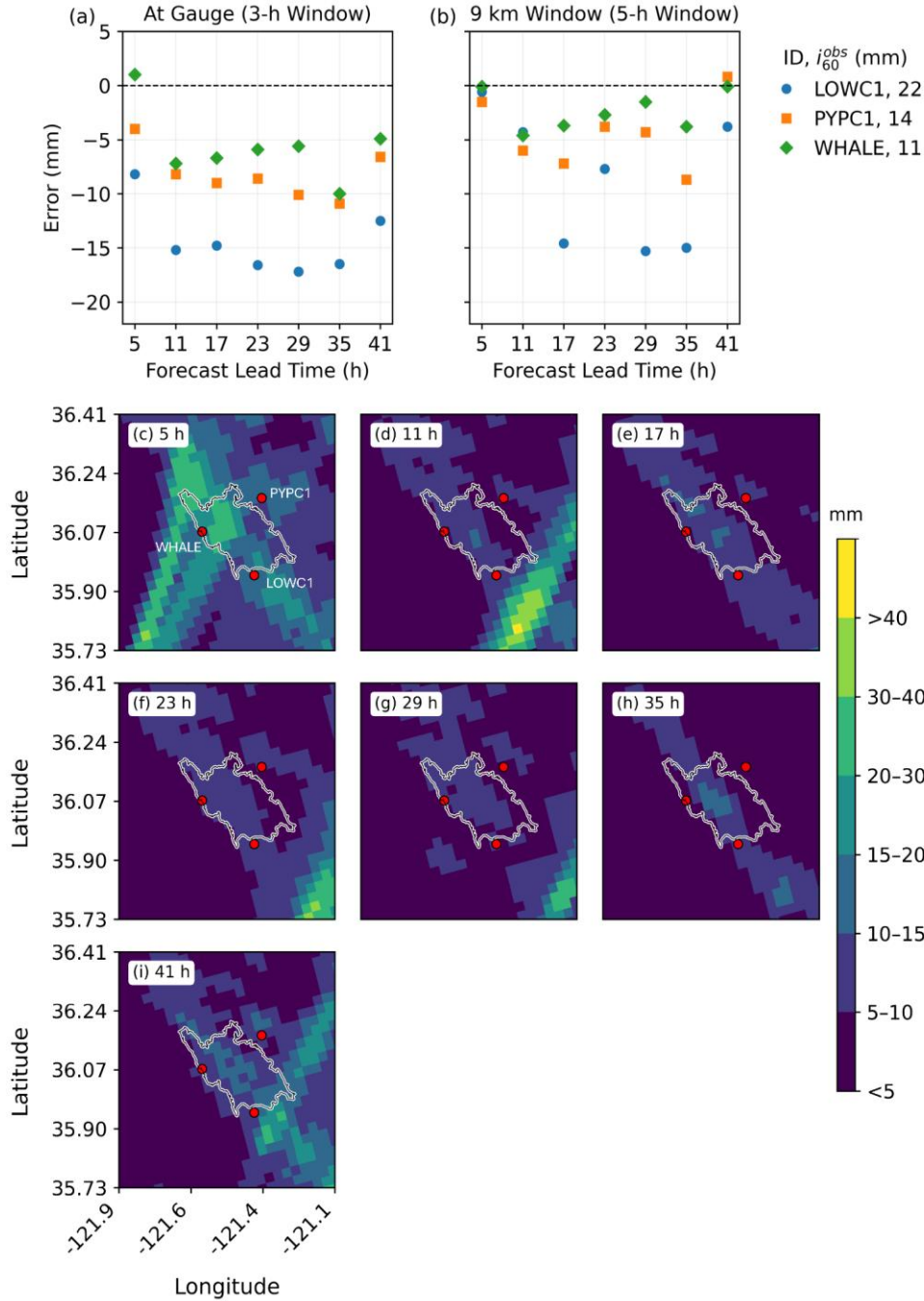


Fig. 10 Analysis of HRRR for the Dolan Fire event, 27 January 2021. Forecast error (mm) for a) model grid cell associated with gauge location and 3 h temporal window, and b) a 9 km neighborhood of model grid cells and 5 h temporal window. The forecast is compared to three rain gauge stations in the vicinity of the burn area. Station IDs and associated peak 1 h rainfall are given in the upper right of the figure. (c–i) A spatial depiction of peak 1 h rainfall for the 3 h event window (Table 1) at each forecast lead time. Colorbar indicates the 1 h rainfall accumulation, in mm. The fire perimeter is indicated by a grey outline, and red dots indicate the location of each station.

525

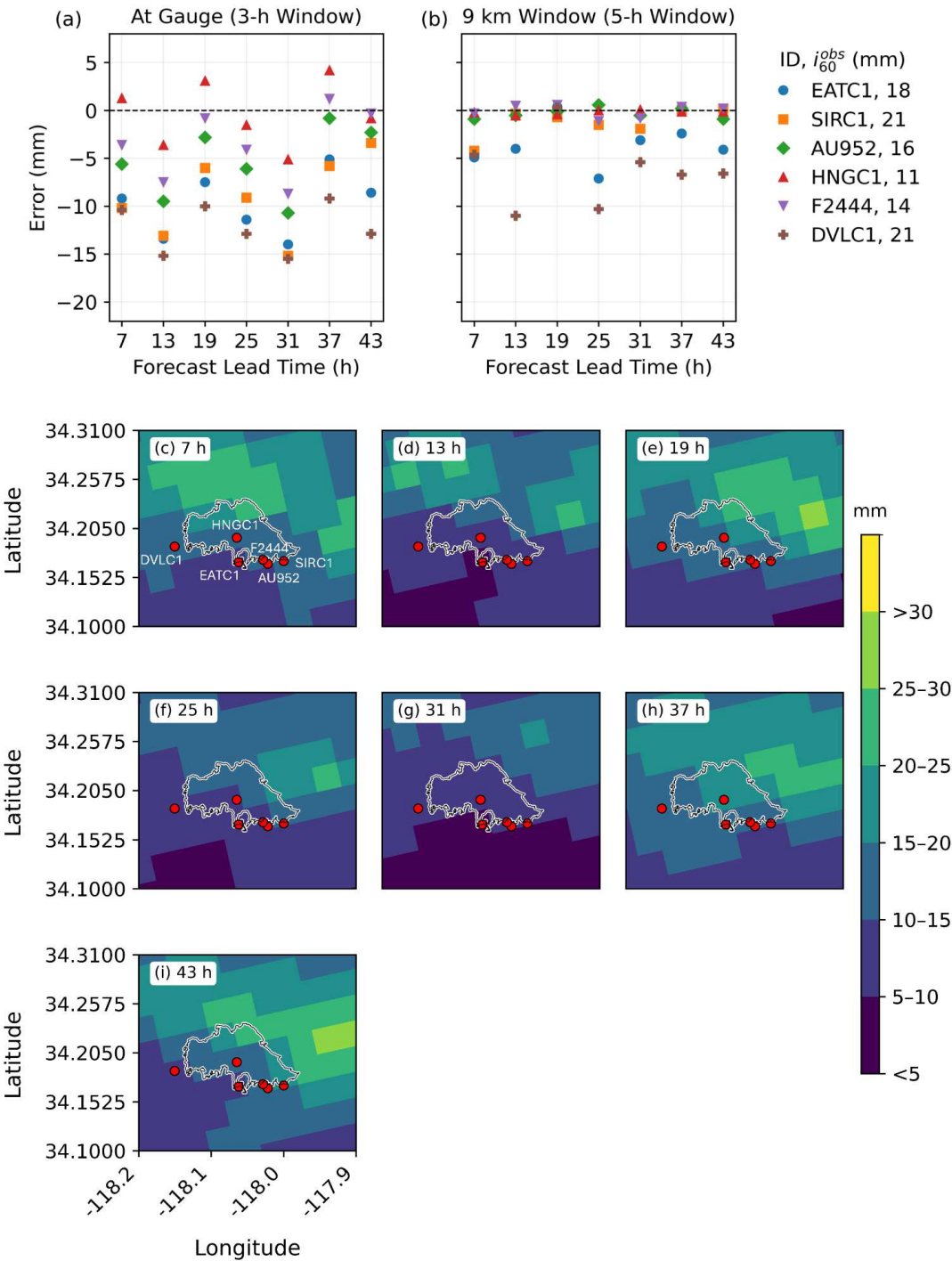


Fig. 11: As in Fig. 10, but for HRRR forecast of the 13 February 2025 Eaton Fire event.

3.4 Model performance across events

Using the event error index (mean among stations for each event for lead times <24h), we ranked the performance for each event for both the WestWRF and HRRR models (Figure 12). Results discussed focus on the 9 km neighborhood, 5 h window.

The sign, magnitude, and range of event error were similar between the two models. Error ranged from approximately +0.1 to -5.1 mm for WestWRF and -0.6 to -8.3 mm in the HRRR. The Woolsey event, which had the greatest error magnitude among events in the HRRR, had an event error index that is twice as large in absolute value relative to the second worst performing HRRR event. The only event with a positive event error index is the WestWRF forecast for the Bobcat (+0.1 mm).

The ranking of the event error index was different between the two models. The Bobcat event was the best performing for WestWRF (+0.1 mm error index) but ranked 7 out of 10 in the HRRR forecast (-3.6 mm error index). The Mosquito event had the smallest magnitude event error index in HRRR (-0.6 mm) but ranked 3rd for WestWRF (-0.7 mm).

The event error index also varied for PFDF events that occurred within the same storm. For example, the 27-28 January 2021 storm produced PFDFs on the CZU, Dolan, and Carmel fires, which are spread out over about 120 km (Fig. 2). In WestWRF, the event error index for the CZU event ranked 2nd among the 10 events, Dolan 7th, and Carmel 10th, with a range of 4.6 mm among them (Fig. 12a). In HRRR, CZU ranked 4, Carmel 6, and Dolan 9, with a range of 2.5 mm (Fig. 12b).

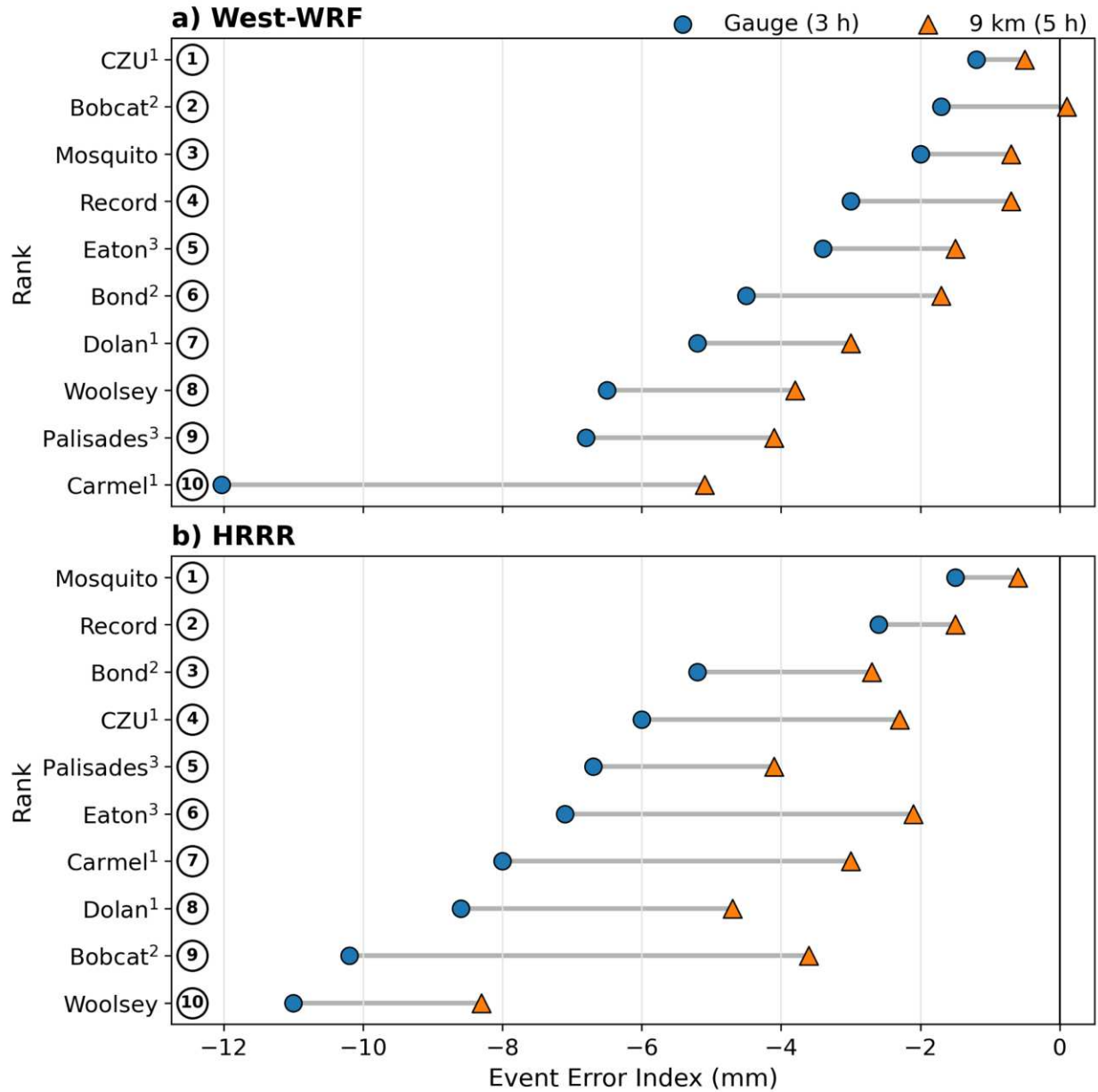


Fig. 12: Values and ranking of the event error index, defined as the mean forecast error among all stations for each event at lead times within approximately 24 hours for a) West-WRF and b) HRRR. Blue circles represent the at-gauge, 3 h time window analysis. Orange triangles represent the 9 km neighborhood, 5 h time window analysis. Events are ranked from smallest error magnitude (1) to greatest error magnitude (10), indicated by numbers on the lefthand side of the panel. Superscripts denote events that occurred within the same storm event. Events without superscripts did not occur on the same date as any other event.

4. Discussion

Forecast rainfall is a component of many landslide early warning systems, including one developed for PPDFs in Southern California (NOAA-USGS Debris Flow Task Force 2005; Guzzetti et al. 2020). In this warning system, rainfall triggering thresholds are developed by (1)

using regional thresholds (e.g., Staley et al. 2013), (2) using firewide average values derived from a watershed-scale debris-flow likelihood model (e.g., Staley et al. 2017; US Geological Survey 2026), (3) specified in Watershed Emergency Response Team reports based on site-specific conditions (CAL FIRE/CGS 2026), or some combination of the three. National Weather Service forecasters then use a variety of forecast tools, including mesoscale models, and their expertise to communicate to partners the potential for exceeding these thresholds and impacts. As the event nears, they use radar, rain and stream gauges, and real-time reports of flood and debris flows to determine if an event is imminent or occurring and issue a flash flood warning.

Our results show that WestWRF and HRRR can provide useful guidance on the potential for PFDF-triggering rainfall at lead times of 24 hours or less based on a high probability of detection (Fig. 4c, 8c). However, probability of detection declined substantially at greater lead times, particularly beyond 51 hours for WestWRF (Fig 4c). The evolution of forecast error with lead time and variability among events also highlights the challenges with using mesoscale models to identify burned areas likely to experience high-intensity rainfall multiple days in advance of a rainstorm. A strict comparison of forecast rainfall with thresholds could alternately identify a burned area as having a high PFDF threat at one lead time and then not at others, in some cases with little consistency (e.g., Fig. 6, 10, S6, S8, S18). This is not unexpected given the challenges associated with forecasting the smaller scale features responsible for delivering short-duration high-intensity rainfall. These results, however, provide illustrative examples that help demonstrate expected differences between forecasting PFDF events and other hazards associated with larger scale atmospheric phenomena.

Correctly forecasting the magnitude of hourly to sub-hourly rainfall intensity is also important for PFDF early warning. Both models struggled to accurately forecast 1-hour intensities ≥ 20 mm h^{-1} . Underestimates of high-intensity rainfall are of concern since the volume and, in turn, runout of debris flows increases with rainfall intensity (Gartner et al. 2014; Barnhart et al. 2026). These errors limit the reliability of PFDF models forced with forecast rainfall, as underestimates of rainfall result in underestimates of inundation depth and runout extent (e.g., Prescott et al. 2024). The magnitude of rainfall intensities is also valuable when considering threshold exceedances. For example, a year-1 60-minute fire-wide threshold in southern California is 13 mm h^{-1} for the 2025 Eaton Fire (CAL FIRE/CGS 2025). A forecast error of 2 mm represents 15% of this threshold. Median WestWRF event-scale error at lead times of 51 hours or more exceeded 5 mm, or 38% of that threshold (Fig. 4). Similar to our results, Bytheway et al. (2025) evaluated 1-hour rainfall intensity forecasts in the warm season from the High Resolution Ensemble Forecast (HREF) system. The study included a western US region. They found that ensemble members were more skillful in predicting exceedances of lower rainfall thresholds than more extreme rainfall thresholds.

Model performance improves substantially when evaluated over a 9 km spatial neighborhood relative to a single model grid cell (Fig. 4, 8). As these deterministic mesoscale model forecasts

verify much better using the neighborhood approach, they are most useful for assessing the potential for PFDF initiation within a fire perimeter rather than identifying specific watersheds most likely to generate debris flows. Forecasting the intersection of susceptible watersheds within burned areas and high intensity rainfall would require models to correctly place the location of high-intensity rainfall in space. Ensemble mesoscale forecast models could be better suited for identifying the likelihood that rainfall over individual watersheds will exceed rainfall triggering thresholds and the potential magnitude of the resulting PFDF (Oakley et al. 2023).

There was not sufficient diversity among the synoptic characteristics of the storm events triggering these PFDF events to evaluate how that factor may have affected model performance (Table 1). Of the six unique storms associated with the 10 PFDF events, five were associated with a weak-to-moderate atmospheric river and the development of a squall line and high intensity rainfall in the vicinity of the storm's cold front. The Woolsey event was distinct from the others in that it was associated with an upper-level closed low-pressure system and scattered thunderstorms. This helps explain its low-ranking event error index for both models (Fig. 12). The track of closed and cutoff lows is notoriously difficult to forecast (e.g. Oakley et al. 2011; Lupo and Schwartz 2025), and distinct from the other events, the high-intensity rainfall is not tied to a frontal passage, which may offer greater predictability (e.g., Cannon et al. 2021).

5. Limitations and future work

A limitation of this analysis is the small sample size of PFDF events and storms. Analyses across a broader range of storms, seasons, and mesoscale phenomena producing high-intensity rainfall are needed to determine the extent to which these findings generalize. Still, these results provide new insight into how forecast skill changes with lead time for PFDF-triggering rainfall and document the increasing error in predicted intensities with lead time, which has been anecdotally observed but not quantified.

Because our analysis is station-based and focuses on events with observed PFDFs, we cannot evaluate false alarms or whether the model correctly predicts high-intensity rainfall that occurs outside the evaluated neighborhood. This limits our ability to calculate many standard forecast verification metrics. To make these assessments, we would need a gridded QPE product that reliably represents short-duration, high-intensity rainfall events in areas of complex terrain and limited radar coverage.

WestWRF and HRRR are both implementations of the Weather Research and Forecasting model and are forced with the GFS (via RAP for HRRR), though use different data assimilation methods and physics schemes. This is likely why we see similar results between the two models, with some differences in the magnitude of error across individual stations and events. Future work could consider additional and new operational models, such as the RRFS. Additionally, ensemble forecasts that provide probabilistic guidance may be more appropriate for assessing the

likelihood of short-duration, high-intensity rainfall events over burn areas (e.g., Oakley et al. 2023) and could also be evaluated.

This study uses data primarily from unshielded tipping bucket rain gauges. These gauges may underestimate rainfall by 5-40%, with wind-induced undercatch one of the primary drivers of error (Sypka 2019; Dunn et al. 2025). Given that forecast rainfall was generally lower than observed, gauge undercatch could mean that the model dry bias is greater than indicated by our results. Additional information on the siting, maintenance, and equipment used on all gauges would help assess gauge observation errors.

We assessed hourly rainfall, but PFDF initiation is often better predicted by subhourly rainfall intensities. Mesoscale model output and archives are not widely available at the 15-minute duration. While 15-minute precipitation output is available for the HRRR, it is not typically incorporated into model output visualization tools. Increasing the availability of 15-minute precipitation output and viewing tools for the output, as well as model verification at sub-hourly durations will benefit PFDF forecasting.

6. Conclusion

Understanding weather forecast model performance for short-duration, high-intensity rainfall conducive to postfire flash flood and debris flow events (PFDF) is important for both forecasting and communicating uncertainty about these hazards. This topic has been understudied, especially in the western US where PFDFs are common. Here, we evaluate two deterministic mesoscale weather forecast models, the NOAA HRRR and CW3E WestWRF, against rain gauge observations for 10 PFDF events in California during six cool-season storms. We evaluate performance for both an at-gauge, 3 h window and a 9 km neighborhood, 5 h window and calculate error, percent error, and probability of detection.

We find that both models exhibit a dry bias relative to station observations. At lead times within 24 h, for the 9 km, 5 h analysis, median error among events ranges from -1.3 to -3.8 mm and percent error ranges from -6% to -16%. Forecast error magnitude and underestimates of observed rainfall intensities increased substantially beyond the 24-30 h window. Median probability of detection (POD, for exceeding 10 mm h^{-1}) is at or close to 1 at lead times ≤ 24 h. Median POD is ≥ 0.5 for the full forecast in HRRR (out to 39-45 h). In WestWRF, median POD is > 0.8 out to 39-51h, and beyond, ranges from 0 to 0.3. The highest observed hourly intensities at individual gauges in our analysis, in the 20-30 mm h^{-1} range, typically had the lowest forecast skill. This relatively low forecast skill often persisted throughout the duration of the forecast, even into the < 24 h window.

This assessment for a small number of PFDF events quantifies previously anecdotal/experiential information noted by forecasters around mesoscale model performance for short-duration, high-intensity rainfall events. This analysis can be used to inform emergency managers and others

making decisions around forecasts of postfire hazards about the benefits and limitations of using mesoscale models to forecast rainfall conducive to PFDF hazards, notably the challenges at longer lead times. Future work could assess performance of emerging ensemble forecast models and explore model development to improve forecasts of rainfall events conducive to PFDF at longer lead times. This would allow for more effective integration with hydrologic and geomorphic models to evaluate flow runout and inundation potential.

Author Contributions:

NO and LM contributed equally to this work. NO and LM led the study conceptualization and methodology development, with contributions from DN. LM led the data analysis and visualization, with data acquisition, analysis, and visualization support from both NO and DN. NO wrote the first draft of the manuscript and all authors commented on previous versions of the manuscript. All authors read and approved of the final manuscript.

Funding:

This study was supported by the Atmospheric River Program Phases III, IV, and V awarded by the California Department of Water Resources (Contracts 4600014294, 4600014942, and 4600015671) to University of California, San Diego.

Competing Interests:

The authors declare no competing interests.

Acknowledgements:

The authors thank Daniel Steinhoff and Matthew Simpson for providing archived WestWRF data. The authors would also like to thank Ryan Kittell, Jayme Laber, and Robert Munroe at the National Weather Service Los Angeles/Oxnard Office for conversations and feedback that helped improve this manuscript.

Data Availability:

The West-WRF output is too large to be publicly archived with available resources. Access to archived model output is available by contacting Daniel Steinhoff at CW3E/UCSD (dsteinhoff@ucsd.edu). HRRR output is available as indicated in Section 2.3. Precipitation data are available through the resources provided in Supplementary Information A and B.

References:

- Banos IH, Mayfield WD, Ge G, Sapucci LF, Carley JR, Nance L (2022) Assessment of the data assimilation framework for the Rapid Refresh Forecast System v0.1 and impacts on forecasts of a convective storm case study. *Geosci Model Dev* 15:6891–6917. <https://doi.org/10.5194/gmd-15-6891-2022>
- Barnhart KR, Kean JW, Lindsay DN, Bilderback EL (2026) Multi-site evaluation of a postfire debris-flow runout forecast method. *Earth's Future* 14:e2025EF007004. <https://doi.org/10.1029/2025EF007004>
- Benjamin SG, Weygandt SS, Brown JM, Hu M, Alexander CR, Smirnova TG, Olson JB, James EP, Dowell DC, Grell GA, Lin H, Peckham SE, Smith TL, Moninger WR, Kenyon JS, Manikin GS (2016) A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh. *Mon Weather Rev* 144:1669–1694. <https://doi.org/10.1175/MWR-D-15-0242.1>
- Blaylock BK (2026) Herbie: Retrieve Numerical Weather Prediction Model Data (Version 2026.1.1) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.18471944>
- Bytheway JL, Abel MR, Mahoney K, Cifelli R (2020) On the uncertainty of high-resolution hourly quantitative precipitation estimates in California. *J Hydrometeorol* 21:865–879. <https://doi.org/10.1175/JHM-D-19-0160.1>
- Bytheway JL, Abel MR, Cifelli R, Mahoney K, English JM (2022) Demonstrating a Probabilistic Quantitative Precipitation Estimate for Evaluating Precipitation Forecasts in Complex Terrain. *Weather Forecast* 37:45–64. <https://doi.org/10.1175/WAF-D-21-0074.1>
- Bytheway JL, Jesberg DR, Trojnia S, Mahoney KM, Correia J, Moore B (2025) Analysis of 3 Years of Summertime Extreme Precipitation Forecasts from the High-Resolution Ensemble Forecast System. *Weather Forecast* 40:2111–2136. <https://doi.org/10.1175/WAF-D-24-0059.1>
- CAL FIRE (2026) California Historical Fire Perimeters. Available at: https://services1.arcgis.com/jUJYIo9tSA7EHvfZ/arcgis/rest/services/California_Historic_Fire_Perimeters/FeatureServer. Accessed 27 Jul 2026
- CAL FIRE/California Geological Survey (CGS) (2025). Eaton Fire Watershed Emergency Response Team (WERT) Report. Available at: <https://www.conservation.ca.gov/cgs/documents/publications/wert/EATON%20WERT%20FINAL%202042025%20ADA.pdf>. Accessed 12 Sep 2026
- CAL FIRE/California Geological Survey (CGS) (2026). Watershed Emergency Response Team (WERT) Reports. Available at: <https://www.conservation.ca.gov/cgs/documents/publications/wert/>. Accessed 29 Sep 2026

- California Geological Survey (2026) Observed postfire debris flows in California, 2000–2025. Available at: <https://cadoc.maps.arcgis.com/apps/dashboards/f165a149061d4854830a7aa0c1e19226> Accessed 13 Jul 2026
- California-Nevada River Forecast Center (CNRFC) (2026) Radar Archive. Available at: <https://www.cnrfc.noaa.gov/radarArchive.php>. Accessed 17 Jul 2026
- Cannon F, Oakley NS, Hecht CW, Michaelis A, Cordeira JM, Kawzenuk B, Demirdjian R, Weihs R, Fish MA, Wilson AM, Ralph FM (2020) Observations and Predictability of a High-Impact Narrow Cold-Frontal Rainband over Southern California on 2 February 2019. *Weather Forecast* 35:2083–2097. <https://doi.org/10.1175/WAF-D-20-0012.1>
- Cannon SH, DeGraff J (2009) The increasing wildfire and post-fire debris-flow threat in western USA, and implications for consequences of climate change. In: Sassa K, Canuti P (eds) *Landslides – Disaster Risk Reduction*. Springer, Berlin, Heidelberg, pp 177–190. https://doi.org/10.1007/978-3-540-69970-5_9
- Center for Western Weather and Water Extremes (CW3E) (2022) CW3E Event Summary 22 December 2021–1 January 2022. Available at: <https://cw3e.ucsd.edu/cw3e-event-summary-22-december-2021-1-january-2022/>. Accessed 17 Jul 2026
- Center for Western Weather and Water Extremes (CW3E) (2025a) CW3E Event Summary 12–14 February 2025. Available at: <https://cw3e.ucsd.edu/cw3e-event-summary-12-14-february-2025/>. Accessed 17 Jul 2026
- Center for Western Weather and Water Extremes (CW3E) (2025b) CW3E Event Summary 12–17 March 2025. Available at: <https://cw3e.ucsd.edu/cw3e-event-summary-12-17-march-2025/>. Accessed 17 Jul 2026
- Collins BD, Oakley NS, Perkins JP, East AE, Corbett SC, Hatchett BJ (2020) Linking Mesoscale Meteorology With Extreme Landscape Response: Effects of Narrow Cold Frontal Rainbands (NCFR). *J Geophys Res Earth Surf* 125:e2020JF005675. <https://doi.org/10.1029/2020JF005675>
- Delle Monache L, Steinhoff DF, Weihs R, Simpson M, Ghazvinian M, Afzali Goroooh VA, Lupo KM, Mulrooney P, Papadopoulos C, Ralph FM (2025) CW3E’s West-WRF 200-Member Ensemble. *Mon Weather Rev* 153:1401–1429. <https://doi.org/10.1175/MWR-D-24-0136.1>
- Dowell DC, Alexander CR, James EP, Weygandt SS, Benjamin SG, Manikin GS, Blake BT, Brown JM, Olson JB, Hu M, Smirnova TG, Ladwig T, Kenyon JS, Ahmadov R, Turner DD, Duda JD, Alcott TI (2022) The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part I: Motivation and System Description. *Weather Forecast* 37:1371–1395. <https://doi.org/10.1175/WAF-D-21-0151.1>

- Dunn RE, Fowler HJ, Green AC, Lewis E (2025) Tipping-bucket rain gauges: a review of the undercatch phenomenon, and methods for its reduction and correction. *Weather* 80:196–205. <https://doi.org/10.1002/wea.7736>
- Ebert EE (2008) Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorol Appl* 15:51–64. <https://doi.org/10.1002/met.25>
- English JM, Turner DD, Alcott TI, Moninger WR, Bytheway JL, Cifelli R, Marquis M (2021) Evaluating Operational and Experimental HRRR Model Forecasts of Atmospheric River Events in California. *Weather Forecast.* 36:1925–1944. <https://doi.org/10.1175/WAF-D-21-0081.1>
- Gartner JE, Cannon SH, Santi PM (2014) Empirical models for predicting volumes of sediment deposited by debris flows and sediment-laden floods in the transverse ranges of southern California. *Engineering Geology* 176:45–56. <https://doi.org/10.1016/j.enggeo.2014.04.008>
- Gaudet LC, Sulia KJ, Torn RD, Bassill NP (2024) Verification of the Global Forecast System, North American Mesoscale Forecast System, and High-Resolution Rapid Refresh Model Near-Surface Forecasts by Use of the New York State Mesonet. *Weather Forecast.* 39:369–386. <https://doi.org/10.1175/WAF-D-23-0094.1>
- Goto EA, Gray S, Keller E, Clarke K (2021) Evacuation choice before and after major debris flows: the case of Montecito, CA. *Int J Disaster Risk Reduct* 62:102400. <https://doi.org/10.1016/j.ijdrr.2021.102400>
- Guan B, Waliser DE (2024) A regionally refined quarter-degree global atmospheric rivers database based on ERA5. *Sci Data* 11:440. <https://doi.org/10.1038/s41597-024-03258-4>
- Guzzetti F, Gariano SL, Peruccacci S, Brunetti MT, Marchesini I, Rossi M, Melillo M (2020) Geographical landslide early warning systems. *Earth-Science Reviews* 200:102973. <https://doi.org/10.1016/j.earscirev.2019.102973>
- Hayden T (2018) Internal Records Reveal Mixed Messages Missed Opportunities Before 1/9 Debris Flow. Santa Barbara Independent. Available at: <https://www.independent.com/2018/05/24/internal-records-reveal-mixed-messages-missed-opportunities-before-1-9-debris-flow/>. Accessed 27 Jul 2026
- Houze RA Jr (2014) *Cloud dynamics*, 2nd edn. Academic Press, Oxford.
- Iowa Environmental Mesonet (IEM) (2026) IEM NWS Text Product Finder. Available at: <https://mesonet.agron.iastate.edu/wx/afos/>. Accessed 13 Jul 2026
- James EP, Alexander CR, Dowell DC, Weygandt SS, Benjamin SG, Manikin GS, Brown JM, Olson JB, Hu M, Smirnova TG, Ladwig T, Kenyon JS, Turner DD (2022) The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part II: Forecast Performance. *Weather Forecast* 37:1397–1417. <https://doi.org/10.1175/WAF-D-21-0130.1>

- Kean JW, Staley DM, Cannon SH (2011) In situ measurements of post-fire debris flows in southern California: Comparisons of the timing and magnitude of 24 debris-flow events with rainfall and soil moisture conditions. *J Geophys Res Earth Surf* 116:F04019. <https://doi.org/10.1029/2011JF002005>
- Kean JW, Staley DM (2021) Forecasting the frequency and magnitude of postfire debris flows across southern California. *Earth's Future* 9:e2020EF001735. <https://doi.org/10.1029/2020EF001735>
- Kieu CQ, Moon Z (2016) Hurricane Intensity Predictability. *Bull Amer Meteor Soc* 97:1847–1857. <https://doi.org/10.1175/BAMS-D-15-00168.1>
- Landsea CW, Cangialosi JP (2018) Have We Reached the Limits of Predictability for Tropical Cyclone Track Forecasting? *Bull Amer Meteor Soc* 99:2237–2243. <https://doi.org/10.1175/BAMS-D-17-0136.1>
- Lin YL (2007) Mesoscale dynamics. Cambridge University Press.
- Lupo KM, Schwartz CS (2025) Environmental Conditions and Forecast Error Evolution Associated with Synoptic Progressiveness Errors in GFS Forecasts. *Weather Forecast* 40:1103–1124. <https://doi.org/10.1175/WAF-D-24-0137.1>
- Markowski P, Richardson Y (2010) Mesoscale Meteorology in Midlatitudes. Wiley-Blackwell, Chichester. <https://doi.org/10.1002/9780470682104>
- Martin A, Ralph FM, Demirdjian R, DeHaan L, Weihs R, Helly J, Reynolds D, Iacobellis S (2018) Evaluation of Atmospheric River Predictions by the WRF Model Using Aircraft and Regional Mesonet Observations of Orographic Precipitation and Its Forcing. *J Hydrometeorol* 19:1097–1113. <https://doi.org/10.1175/JHM-D-17-0098.1>
- Mascioli NR, Ghazvinian M, Weihs R, Papadopoulos C, Steinhoff DF, Simpson M, Della Monache L (2026) Ensemble Design for Forecasting Extreme Orographic Precipitation: Finer Horizontal Grid Spacing with Fewer Members? *Weather Forecast* 41:1275–1288. <https://doi.org/10.1175/WAF-D-25-0177.1>
- McGuire LA, Ebel BA, Rengers FK, Vieira DCS, Nyman P (2024) Fire effects on geomorphic processes. *Nat Rev Earth Environ* 5:486–503. <https://doi.org/10.1038/s43017-024-00557-7>
- NOAA Rapid Refresh Forecast System (RRFS) Prototype (2026) Available at: <https://registry.opendata.aws/noaa-rrfs>. Accessed 15 Feb 2026
- NOAA Weather Prediction Center (WPC) (2026) WPC's Surface Analysis Archive. Available at: https://www.wpc.ncep.noaa.gov/archives/web_pages/sfc/sfc_archive.php. Accessed 13 Jul 2026
- NOAA-USGS Debris Flow Task Force (2005) NOAA-USGS debris flow warning system- Final report. U.S. Geological Survey Circular 1283. Available at <http://pubs.usgs.gov/circ/2005/1283/pdf/Circular1283.pdf>. Accessed 12 Sep 2026

- Oakley NS and Redmond KT (2014) A Climatology of 500-hPa Closed Lows in the Northeastern Pacific Ocean, 1948–2011. *J. Appl. Meteorol. Climatol.* 53-6:1578-1592. <https://doi.org/10.1175/JAMC-D-13-0223.1>
- Oakley, NS, Lancaster, JT, Kaplan ML, Ralph, FM (2017) Synoptic conditions associated with cool season post-fire debris flows in the Transverse Ranges of southern California. *Nat Hazards* 88:327–354. <https://doi.org/10.1007/s11069-017-2867-6>
- Oakley NS, Liu T, McGuire LA, Simpson M, Hatchett BJ, Tardy A, Kean JW, Castellano C, Laber JL, Steinhoff D (2023) Toward probabilistic post-fire debris-flow hazard decision support. *Bull Amer Meteor Soc* 104:E1587–E1605. <https://doi.org/10.1175/BAMS-D-22-0188.1>
- Oakley NS, Cheung DJ, Lindsay DN, Nash D (2025) Insights from a 25-year database of post-fire debris flows in California. *Int J Wildland Fire* 34:WF25136. <https://doi.org/10.1071/WF25136>
- Orlanski I (1975) A rational subdivision of scales for atmospheric processes. *Bull. Am. Meteorol. Soc.* 56-5:527-530.
- Prescott AB, McGuire LA, Jun K-S, Barnhart KR, Oakley NS (2024) Probabilistic assessment of postfire debris-flow inundation in response to forecast rainfall. *Nat Hazards Earth Syst Sci* 24:2359–2374. <https://doi.org/10.5194/nhess-24-2359-2024>
- Ralph FM, Rutz JJ, Cordeira JM, Dettinger MD, Anderson M, Reynolds D, Schick LJ, Smallcomb C (2019) A Scale to Characterize the Strength and Impacts of Atmospheric Rivers. *Bull Amer Meteor Soc* 100:269–289. <https://doi.org/10.1175/BAMS-D-18-0023.1>
- Roebber PJ (2009) Visualizing Multiple Measures of Forecast Quality. *Weather Forecast* 24:601–608. <https://doi.org/10.1175/2008WAF2222159.1>
- Sawicki, E (2021) Winter storm closes sections of PCH. Santa Monica Daily Press. Available at <https://www.smdp.com/winter-storm-closes-sections-of-pch/>. Accessed 29 Sep 2026
- Schwartz JY, Oakley NS, Alessio P (2021) Assessment of a Post-Fire Debris Flow Impacting El Capitan Watershed, Santa Barbara County, California, U.S.A. *Environmental and Engineering Geoscience* 27:423–437. <https://doi.org/10.2113/EEG-D-21-00008>
- Sinclair AE, Oakley NS, West JA, Clark MK, Horton DE (2026) Uncertainty in precipitation estimates warrants greater consideration in debris-flow research. *J. Geophys. Res. Earth Surface*
- Skamarock WC, Klemp JB, Dudhia J, Gill DO, Liu Z, Berner J, Wang W, Powers JG, Duda MG, Barker DM, Huang X-Y (2019) A description of the Advanced Research WRF version 4. NCAR Technical Note NCAR/TN-556+STR, 145 pp. <https://doi.org/10.5065/1dfh-6p97>

- Staley DM, Kean JW, Cannon SH, Schmidt KM, Laber, JL (2013) Objective definition of rainfall intensity–duration thresholds for the initiation of post-fire debris flows in southern California. *Landslides* 10:547–562. <https://doi.org/10.1007/s10346-012-0341-9>
- Staley DM, Negri JA, Kean JW, Laber JL, Tillery AC, Youberg AM (2017) Prediction of spatially explicit rainfall intensity–duration thresholds for post-fire debris-flow generation in the western United States. *Geomorphology* 278:149–162. <https://doi.org/10.1016/j.geomorph.2016.10.019>
- Staley DM, Kean JW, Rengers FK (2020) The recurrence interval of post-fire debris-flow generating rainfall in the southwestern United States. *Geomorphology* 370:107392. <https://doi.org/10.1016/j.geomorph.2020.107392>
- Su L, Cao Q, Shukla S, Pan M, Lettenmaier DP (2023) Evaluation of Subseasonal Drought Forecast Skill over the Coastal Western United States. *J Hydrometeorol* 24:709–726. <https://doi.org/10.1175/JHM-D-22-0103.1>
- Sypka P (2019) Dynamic real-time volumetric correction for tipping-bucket rain gauges. *Agric For Meteorol* 271:158–167. <https://doi.org/10.1016/j.agrformet.2019.02.044>
- U.S. Geological Survey (2026) USGS Postfire Debris Flow Hazard Assessment Viewer. Available at: <https://usgs.maps.arcgis.com/apps/dashboards/c09fa874362e48a9afe79432f2efe6fe>. Accessed 15 Jun 2026
- White P, Nelson PA (2024) Evaluation of Sub-Hourly MRMS Quantitative Precipitation Estimates in Mountainous Terrain Using Machine Learning. *Water Resour Res* 60:e2024WR037437. <https://doi.org/10.1029/2024WR037437>
- Wick GA, Neiman PJ, Ralph FM, Hamill TM (2013) Evaluation of Forecasts of the Water Vapor Signature of Atmospheric Rivers in Operational Numerical Weather Prediction Models. *Weather Forecast* 28:1337–1352. <https://doi.org/10.1175/WAF-D-13-00025.1>
- Zou X, Cordeira JM, Bartlett SM, Kawzenuk B, Roj S, Castellano C, et al. (2023) Mesoscale and synoptic scale analysis of narrow cold frontal rainband during a landfalling atmospheric river in California during January 2021. *J Geophys Res Atmos* 128:e2023JD039426. <https://doi.org/10.1029/2023JD039426>

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementAstationlist.xlsx](#)
- [SupplementBBWGRG1.csv](#)
- [SupplementCadditionalfigures.docx](#)