Improving Confidence in Model-Based Probable Maximum Precipitation: How Important is Model Uncertainty in Storm Reconstruction and Maximization?

EMILIE TAROUILLY,[a] FOREST CANNON,[b] AND DENNIS P. LETTENMAIER[c]

[a] *Department of Civil and Environmental Engineering, University of California, Los Angeles, Los Angeles, California*
[b] *Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California, San Diego, San Diego, California*
[c] *Department of Geography, University of California, Los Angeles, Los Angeles, California*

(Manuscript received 31 March 2022, in final form 14 October 2022)

ABSTRACT: We analyze uncertainty in model-based estimates of probable maximum precipitation (PMP) as used in dam spillway design. Our focus is on model-based PMP derived from Weather Research and Forecasting (WRF) Model reconstructions of severe historical storms, amplified by the addition of moisture in the boundary conditions [so-called relative humidity maximization (RHM)]. By scaling moisture and predicting the resulting precipitation, the model-based approach arguably is more realistic than currently used techniques [documented in NOAA's Hydrometeorological Reports (HMRs)], which assume that precipitation scales linearly with moisture. Despite the important improvement this represents, model-based PMP is subject to several sources of uncertainty that have slowed adoption in operational settings. We analyze an ensemble of PMP simulations that reflect recognized sources of uncertainty including the following: 1) initial condition error, 2) choice of physics parameterizations, and 3) upscale propagating model errors. We apply this ensemble approach to the Feather River watershed (Oroville Dam) in California for the storms of February 1986 and January 1997, which produced some of the largest floods on record at that location, after carrying out in-depth evaluations of model reconstructions. Differences in the maximized 72-h precipitation totals across the 56 ensemble members we produced for each storm are modest, ranging from ±7% of ensemble mean. Our results suggest that while model-based PMP estimates should be interpreted as a range of values, model uncertainty appears to be relatively small for the major atmospheric river–driven flood events we investigated.

KEYWORDS: Atmospheric river; Extreme events; Ensembles

1. Introduction

California is especially prone to hydrologic extremes, including droughts and floods, such that it relies on a large network of dams for both water supply and flood control. The integrity of these dams is ensured by spillways that are sized to pass the most severe flood that could occur, which is termed the probable maximum flood (PMF). The PMF is additionally used to ensure the safety of other high-risk structures, such as nuclear power plants. The PMF estimate is derived using the probable maximum precipitation (PMP) as an input (WMO 2009). The PMP is defined as "theoretically the greatest depth of precipitation for a given duration that is physically possible." For the western United States, guidance on how to obtain the PMP using a storm maximization approach is provided in NOAA Hydrometeorological Report (HMR) 59 (Corrigan et al. 1999).

The storm maximization approach as described in the HMRs (hereafter referred to as "HMR PMP") has changed little since it was first introduced for California in HMR 36 (USWB 1961). HMR PMP is obtained by scaling the precipitation of an extreme historical storm. The rationale is that storms with very high precipitation efficiency have occurred in the historical record but may have been moisture limited. Therefore, scaling precipitation, typically by the ratio of atmospheric moisture during an extreme storm to the climatological maximum, should approximate the largest precipitation depth that could occur. While the HMR PMP estimates contain some useful information, the methodology relies on many simplifying assumptions. Among them are that precipitation scales linearly with moisture, the reliance on historical storms in a changing climate, and that maximum efficiency has been achieved by historical storms. All of these assumptions have been questioned (Abbs 1999; Chen and Bradley 2006). It is now widely accepted that leveraging state-of-the-art numerical weather prediction models could address some of these assumptions and produce more robust PMP estimates (Chen and Hossain 2016). These developmentsare prompting a reevaluation of PMP estimates, starting with the use of model-derived precipitation to fill-in data gaps and apply HMR PMP methods to model output when observations are limited (e.g., Mahoney et al. 2021 for Colorado and New Mexico).

New methods to estimate PMP directly using the Weather Research and Forecasting (WRF) Model (Skamarock and Klemp 2008) have also been developed over the last decade. These methods (hereafter "model-based PMP" as opposed to "HMR PMP") follow the same logic as the HMRs. The most widely used method, termed relative humidity maximization (RHM) (Ohara et al. 2011; Ishida et al. 2015a,b), consists of

Corresponding author: Dennis Lettenmaier, dlettenm@ucla.edu

Publisher's Note: This article was revised on 23 February 2022 to correct the affiliation of the second author, whose affiliation was not correct when originally published.

DOI: 10.1175/JHM-D-22-0044.1

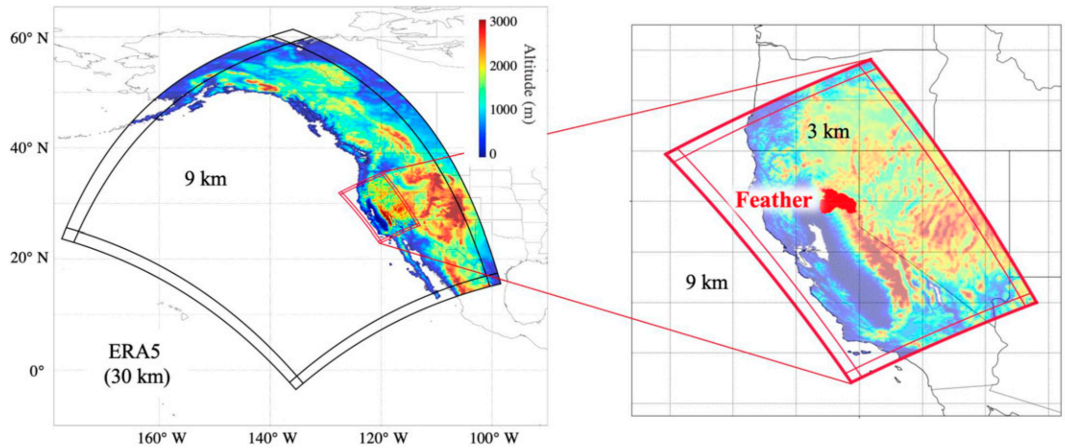

FIG. 1. Location of the Feather River watershed, on the western slopes of the Sierra Nevada, and the two nested 9- and 3-km modeling domains.

reconstructing and amplifying a severe historical storm. This is achieved by modifying the forcing dataset such that relative humidity is 100% at the model boundaries, i.e., by saturating the atmospheric column. This increased moisture at the model boundaries is advected into the model domain and generally increases precipitation rates relative to the historical storm. The largest accumulated maximized precipitation over the basin (typically over 72 h) is then retained as the PMP estimate. This model-based approach to PMP arguably is more physically realistic than HMR methods as the increased precipitation is produced in accordance with the model's representation of storm physics, rather than linearly scaling precipitation with moisture.

The development of model-based PMP nevertheless creates new challenges that are only beginning to be recognized (Mahoney et al. 2021). In particular, the choice of model parameterizations, errors in the initial conditions and model errors (upscale propagation of unresolved subgrid processes) can create substantial differences in the reconstructed storm totals and hence in PMP estimates. Despite a growing body of work applying model-based PMP (e.g., Toride et al. 2019; Gangrade et al. 2019 in addition to the above references), their uncertainty to our knowledge has not been assessed. Some of the aforementioned PMP studies (Ohara et al. 2011; Ishida et al. 2015a,b) may have used different models (including MM5, a predecessor of the WRF Model) and setups (domains, resolution, physics options), the precise impact of which remains to be quantified. Our approach here is to design an ensemble of simulations that captures those important sources of uncertainty in model-based PMP such that their impact on PMP estimates can be assessed.

Additional motivation for representing uncertainty in model-based PMP is the growing interest in risk-based rather than deterministic approaches to flood preparedness. Risk-based approaches typically involve the generation of a large number of severe storms (which arise from different combinations of conditions) with different associated probabilities. This approach has been used to quantify uncertainty in HMR PMP (Micovic et al. 2015) and in stochastic flood modeling studies (e.g., MGS Engineering Consultants, Inc. 2005). Considering a number of plausible extreme values rather than a single storm as in risk-based approaches is valuable because though PMP is a deterministic concept (whether HMR or model-based), it is well known that the tail end of rainfall frequency distributions, where PMP lies, is highly uncertain (Smith and Baeck 2015; Enzel et al. 1993; O'Connor et al. 2002). Therefore, delineating the range (due to model uncertainty) of possible PMP estimates is not only a first step in improving confidence in model-based PMP but will also help generate an understanding of extreme storms that can support further development of risk-based frameworks.

We seek to improve the robustness and utility of model-based PMP by identifying key sources of model error and uncertainty and quantifying their impact on the range of possible PMP values. To do so, we first assess the performance of a single-configuration storm reconstruction, and then produce an ensemble of PMP values (rather than a single estimate) in order to adequately reflect uncertainty. The science questions we address are as follows:

1) What is the impact on the PMP estimate of the quality of model reconstructions of precipitation?
2) What is the overall impact on the PMP estimate of known sources of uncertainty (initial condition error, choice of model parameterizations, and model errors)?
3) How important is this uncertainty relative to the size of the maximization signal?

2. Methods

a. Study area and storms of interest

Our study area is the Feather River watershed (3600 mi^2) upstream of Oroville Dam, California (Fig. 1). Our choice of this location is guided by the existence of earlier PMP estimation work for the Feather River and adjacent basins (Yuba and American Rivers) (e.g., Ishida et al. 2015a,b; Ohara et al.

TABLE 1. West-WRF physics parameterizations used as the baseline run against which ensemble members (with different combinations of physics options and/or perturbations) are compared (Martin et al. 2018).

WRF option	Scheme name
Microphysics	Thompson (Thompson et al. 2008)
Cumulus scheme	Grell–Freitas (Grell and Dévényi 2002)
Boundary layer scheme	Yonsei University (Hong et al. 2006)
Longwave physics	RRTMG (Mlawer et al. 1997)
Shortwave physics	RRTMG (Mlawer et al. 1997)
Surface layer physics	Revised MM5 (Jiménez et al. 2012)
Land surface physics	Unified Noah Land Surface Model (Niu et al. 2011)

2011). The Feather River, which is located on the western slopes of the Sierra Nevada, makes for an ideal environment for improving PMP in that the dominant precipitation mechanism is atmospheric rivers (ARs). These synoptic-scale weather systems are inherently more predictable and better represented by numerical models than e.g., small-scale convective storms. In addition, topography plays a role in producing more constrained simulations (Mahoney et al. 2021). We performed model reconstruction and maximization of two extreme historical events: the storms of February 1986 and January 1997. The aforementioned studies have identified those two storms as producing some of the largest precipitation totals both in the historical record and after maximization. They additionally have different dynamics (amounts of moisture, convection) (Leung and Qian 2009), which may allow us to capture different reconstruction performance and responses to maximization, if any. We considered other storms such as December 1964 and February 2017 (associated with the Oroville Dam spillway incident) but these storms were not retained because of limited hourly observations available to evaluate model reconstructions and relatively low 72-h precipitation totals, respectively.

b. *Baseline reconstruction of historical storms: Model setup*

We first produced single-configuration WRF reconstructions of the February 1986 and January 1997 storms. These baseline storm reconstructions allowed us to assess model performance before performing maximization (section 2d) and subsequently including other configurations in the ensemble experiments for both reconstructed and maximized versions of the storms (section 2e).

We used the WRF Model version 3.7.1 (Skamarock and Klemp 2008). We followed the "West-WRF" model configuration for our baseline runs. West-WRF is used by the Center for Western Weather and Water Extremes as its operational forecast model and is tailored for extreme precipitation associated with atmospheric rivers along the U.S. West Coast. The West-WRF configuration is described by Martin et al. (2018) and parameterization schemes are summarized in Table 1. The WRF Model was set up with two nested domains with resolutions of 9 and 3 km over the coastal western United States (Fig. 1). The cumulus scheme was turned off in the inner domain. The wide area covered by the outer domain allowed to capture storm tracks and the Aleutian low, which plays an important role in steering ARs toward the U.S. West Coast.

We took the initial and boundary conditions for the WRF simulations from ERA5 Reanalysis (Hersbach et al. 2020) at 30-km spatial resolution. We chose ERA5 because of its ability to reproduce IVT (IVT being an important factor in extreme precipitation in this region), advanced data assimilation scheme and high resolution (Cobb et al. 2021). Our WRF Model runs are for periods of roughly six days which allows for at least one day for spin up prior to the storm as well as capturing three days of storm conditions for which maximized storm totals are calculated.

c. *Evaluation of model reconstructions*

We evaluated the WRF reconstructions of historical storms through comparisons with the Cao et al. (2019) hourly gridded (1/32°) precipitation dataset. The gridded precipitation was obtained using the Mountain Mapper method using hourly and daily data from NOAA's Cooperative Observer Program (COOP) network, Remote Automatic Weather Stations (RAWS), the Automated Surface Observing System (ASOS), the NOAA Hydrometeorological Automated Data System (HADS), the California Data Exchange Center (CDEC), and NOAA's Hydrometeorology Testbed (HMT). In total, 60 stations were available for the February 1986 storm and 85 stations for January 1997. This dataset was chosen among other gridded products (which also use variations of the Mountain Mapper method) as it is the only dataset that is hourly and available during both storms for this location.

d. *Baseline storm maximization for PMP estimation*

The maximized storm simulations were produced using a technique developed by Ishida et al. (2015b) called relative humidity maximization (RHM). It is currently the most widely used model-based PMP technique (see e.g., Gangrade et al. 2019). While other methods are being developed (see Toride et al. 2019), the primary goal of this work is not to further refine the model-based PMP technique but rather to evaluate uncertainty in an approach that currently exists. The logic of RHM is to amplify a severe historical storm (as reconstructed by the WRF Model, see section 2b above) by providing it with additional moisture at the domain boundaries. This is done by setting relative humidity to 100% at all boundary grid cells and all vertical levels in the forcing dataset (i.e., saturating the boundary conditions) so that additional moisture flows into the model domain. The basin-averaged 72-h total precipitation (for the same window that produced the highest observed totals) obtained in the moisture-maximized simulation is calculated for each storm (February 1986 and January 1997). We emphasize that our goal here is not to obtain a PMP estimate but rather to assess uncertainties in maximized storm totals that have the potential to yield the PMP estimate, therefore we continue

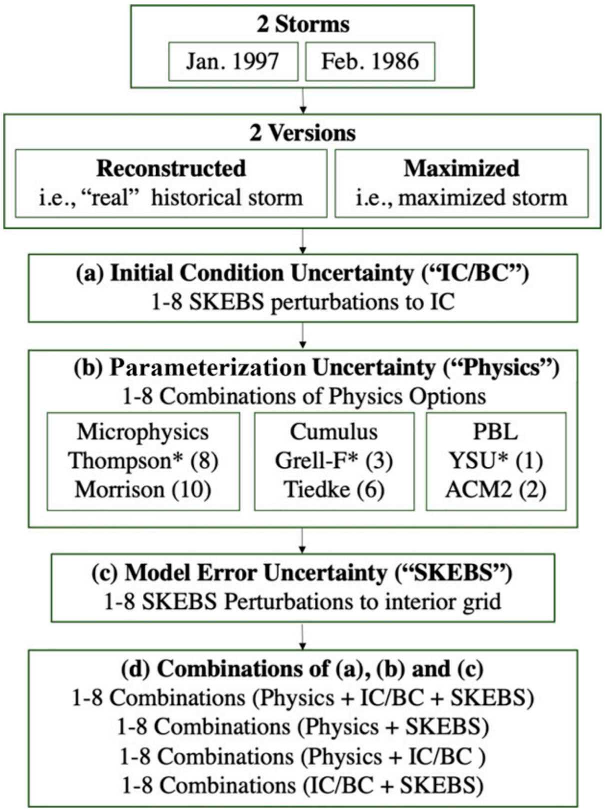

FIG. 2. (a)–(d) Summary of the ensembles for the reconstructed and maximized sets of simulations for both the storms of February 1986 and January 1997. The ensembles were designed to sample known sources of uncertainty that affect precipitation, hence PMP estimates. In (b), the asterisks denote the physics options used in the West-WRF setup and the numbers in parentheses refer to the WRF parameterization code.

to work with both storms throughout this paper and refer to "maximized" rather than "PMP" totals.

e. Uncertainty experiments

In addition to the baseline runs described above, we formed ensembles of WRF simulations for both versions (reconstructed and maximized) of both storms (February 1986 and January 1997) that capture key sources of uncertainty in modeled precipitation. We focused on three of the most widely acknowledged sources of uncertainty that affect modeled precipitation: boundary and initial condition errors, model parameterization and upscale propagating errors (Berner et al. 2015). The remainder of this section describes the design of the ensembles, which aims to balance computational cost with the generation of realistic spread. Figure 2 provides an overview of all experiments, which resulted in 56 ensemble members for each version (reconstructed and maximized) of each storm.

Uncertainty due to the choice of physics parameterization was assessed by testing different combinations of possible microphysics, cumulus and boundary layer schemes. These are the parameterizations that exert the most control on WRF precipitation (Michaelis et al. 2021; Martin, et al. 2018), hence on PMP estimates. Besides the Thompson (Thompson et al. 2008), Grell–Freitas (Grell and Dévényi 2002), and YSU (Hong et al. 2006) combination of the West-WRF setup (which we use as our baseline), the following alternative schemes were used: Morrison (Morrison et al. 2009) for microphysics, Tiedtke (Tiedtke 1989) for cumulus, and ACM2 (Pleim et al. 2007) for PBL. This produced eight different combinations of those possible schemes (Fig. 2b). This choice of schemes is guided by F. Cannon et al. (2022, unpublished manuscript), who identified schemes that differ from each other (in order to generate representative spread) among those found to be appropriate for the reconstruction of AR storms in this region.

Besides parameterization, the influence of upscale propagating errors arising due to initial conditions and model formulation was addressed using WRF's stochastic energy backscatter scheme (SKEBS; Shutts et al. 2011; Palmer et al. 2009; Shutts 2005; Berner et al. 2009). SKEBS as implemented in WRF represents the upscale transfer of kinetic energy by generating streamfunction perturbations, which perturb the rotational wind. SKEBS was used with default WRF values as recommended by Berner et al. (2011). We applied SKEBS perturbations to initial and boundary conditions in eight different WRF runs (which we term "IC/BC" runs), and to the interior grid in another eight runs (which we term "SKEBS" runs) (Figs. 2a,c, respectively).

In addition to ensemble members that consist of either different physics parameterization, or initial/boundary condition perturbations, or interior grid perturbations alone, we produced runs that consist of various combinations of the experiments above (Fig. 2d). We therefore obtain a 56-member ensemble for each of the "reconstructed" and "maximized" version of each storm (i.e., 112 members for each storm). Our goal in creating this ensemble was not to attribute uncertainty to either of these possible causes (boundary and initial condition errors, model parameterization, upscale propagating errors or a combination) but rather to assess the magnitude of the uncertainty produced by all of those different sources together.

3. Results

a. Baseline storm reconstructions

1) MODEL EVALUATION

We first assessed the performance of the baseline storm reconstructions (West-WRF, no perturbations) against which we subsequently (section 3d) compared the full ensembles. Our baseline precipitation reconstructions match observations well for January 1997, but less so for February 1986. The timing of precipitation peaks is closely reproduced for both storms (Fig. 3d, bottom row). The location of precipitation centers (Figs. 3a,b, top row) is also well represented, but the model has a wet bias at midelevations (Fig. 3c, second row), which exists in both storms but is more pronounced in February 1986. Correlations between observed and reconstructed precipitation (across all pixels within the basin and

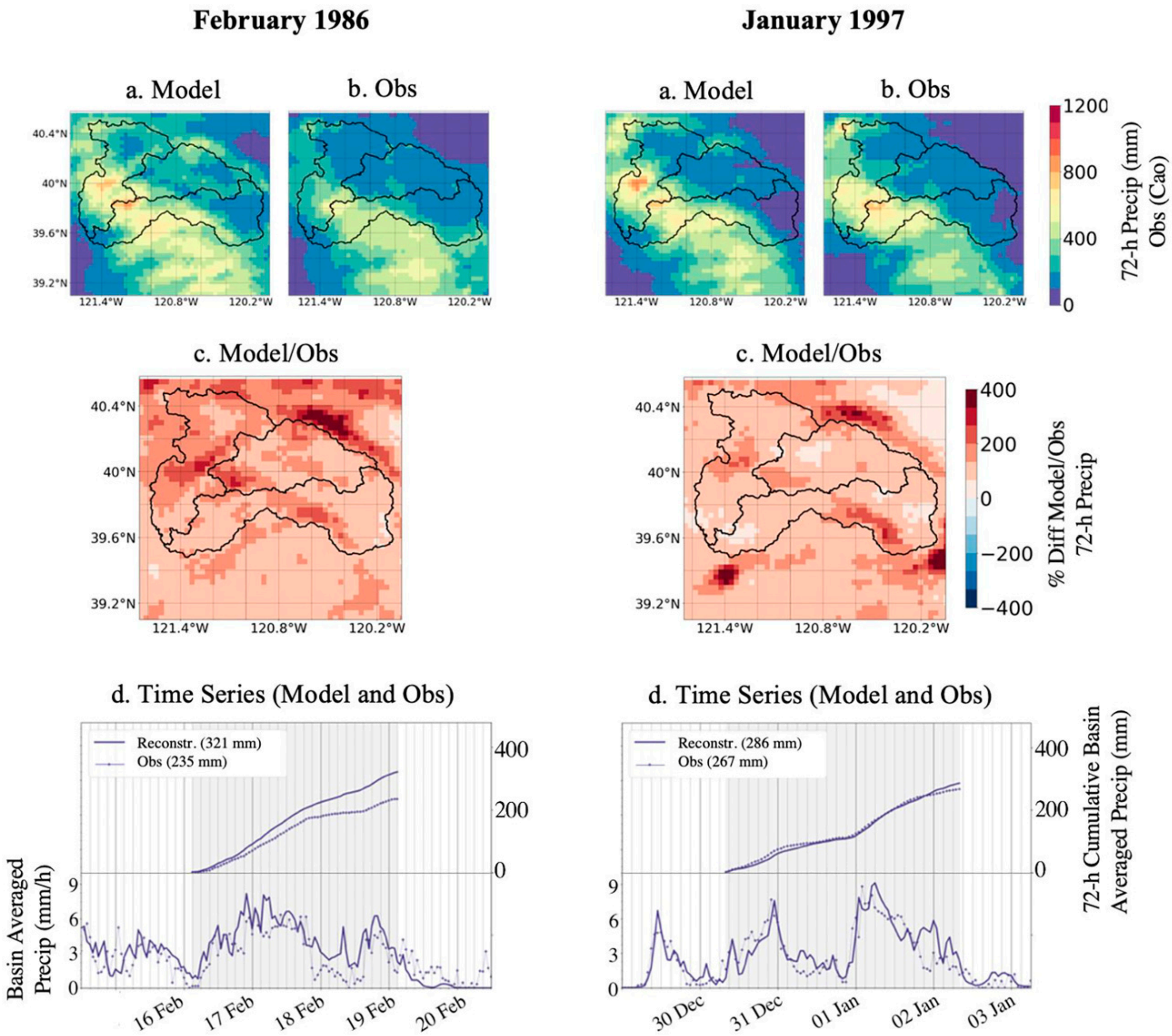

FIG. 3. Evaluation of historical storm reconstructions: the spatial and temporal pattern of precipitation match observations well for both the February 1986 and January 1997 storms, despite wet biases in basin-average 72-h total precipitation. (a),(b) Maps of modeled and observed 72-h total precipitation, respectively. (c) Maps of percentage difference (modeled/observed). (d) Precipitation time series (modeled and observed) with 72-h precipitation totals shown in the legend.

all 72 h during the storm) exceed 0.7 for February 1986 and 0.8 for January 1997, as has been pointed out before. For example, Ishida et al. (2015a) similarly report a correlation of 0.78 averaged across 61 extreme storms in the Feather River watershed. Overall, our simulations perform as well as such previous WRF applications to the same domain.

We draw attention to the 72-h precipitation totals, which are arguably an important characteristic to reproduce for PMP estimation. The reconstructed basin-averaged 72-h total precipitation for February 1986 has a wet bias which amounts to 37% of the observed total (235 mm), in large part due to excessive modeled precipitation during the first half of February 18. On the other hand, the reconstructed basin-averaged 72-h total precipitation for the January 1997 storm (286 mm) has a smaller wet bias of 7% of the observed total (267 mm). Such wet biases and the lower model performance in February 1986 compared to January 1997 have been noted in other PMP studies in the Sierras, including Ishida et al. (2015a) and Ohara et al. (2011, 2017). Yet this is the first time that the magnitude of these biases in reconstructed precipitation totals are examined in the context of PMP estimation, the importance of which we further assess in section 3b in light of the magnitude of the maximized totals.

The performance of our reconstructions reflects the current capabilities and limitations of mesoscale precipitation modeling. The wet bias we observe in the Sierra Nevada is acknowledged by a large body of work on WRF performance (e.g., F. Cannon et al. 2022, unpublished manuscript; English et al. 2021; Caldwell et al. 2009; Jankov et al. 2009). Further

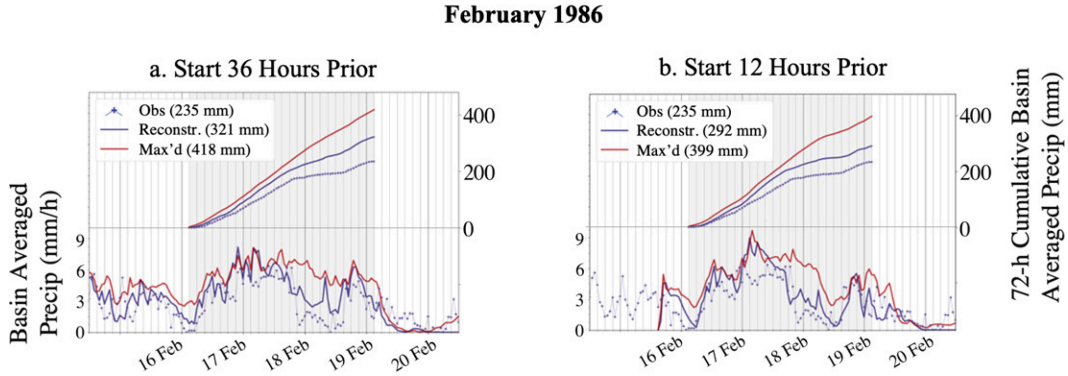

FIG. 4. Precipitation time series for reconstructed and maximized runs precipitation (with 72-h precipitation totals shown in the legend) started (a) 36 h prior (as in the rest of this study, as in Fig. 3d, shown again here for comparison) and (b) 12 h prior to the onset of the February 1986 storm. Late-start runs show a slight improvement in reconstructed precipitation and result in a lower maximized precipitation total compared to early-start simulations.

investigating the causes of the wet bias, and specifically condition-dependent performance between these two storms, could improve confidence in PMP. Precipitation performance is the topic of recent and ongoing research (such as the work cited above on WRF performance) which is targeted to errors in the modeled temperature profile, integrated water vapor, wind speed and direction. Given the complexity of these processes, we do not try to further explain precipitation performance but instead focus on evaluating the uncertainty that arises from the resulting errors. Conversely, discrepancies between observed and modeled precipitation could stem from errors in observations. To address observation errors, Newman et al. (2015) suggest that an ensemble of precipitation data should be used from which uncertainty can be estimated, but no such dataset is available for the Feather River basin at the hourly time scale. While some strategies to address model performance (e.g., retaining only storms that are well reconstructed or using bias correction designed to leverage ensemble model output, see Chapman et al. 2019) hold some promise, these strategies do not seem appropriate given the currently limited understanding of observation uncertainty. Thus, at this stage using all storms without correction appears to be the best strategy, so long as the possible errors are known and acknowledged.

2) IMPORTANCE OF SIMULATION START TIME RELATIVE TO STORM ONSET

Given the performance issues noted in the February 1986 reconstruction, we investigated whether starting the simulation closer to the beginning of the 72-h storm period would improve modeled precipitation, and whether this would have an impact on the maximized total. While the model does require some spinup, earlier start times are expected to accumulate more errors. Therefore, a choice about when to start the simulations is made, which represents an additional source of variability in PMP estimates. The wet bias in the reconstructed 72-h basin-averaged total decreases from 37% of observed totals (starting 36 h prior as in the rest of this study) to 24% (starting 12 h prior). The improvement is mainly due to the lower precipitation during the first half of February 18th being better represented in the simulation with the shorter lead time (Fig. 4). The maximized 72-h totals as a result decrease from 418 to 399 mm. Our goal here is not to determine the optimal time to start simulations, but rather to point out that simulation start time is another, albeit small, source of uncertainty in PMP estimates.

b. *Baseline storm maximization*

Next, we performed the maximized simulations using the RHM technique of Ishida et al. (2015b) for the same two storms (February 1986 and January 1997) we reconstructed above. The setup for maximized simulations is identical to the reconstructions with the exception that additional moisture is provided at the model boundaries. As expected, the RHM technique produces a stronger AR with increased integrated vapor transport (IVT) in comparison with the reconstruction (Figs. 5a,b, top row). The precipitation increases produce basin-averaged maximized totals (418 and 399 mm, respectively) that represent 130% and 140%, respectively, of reconstructed totals over the Feather, for the storms of February 1986 and January 1997 (Figs. 5c–e, middle and bottom rows). We refer to these precipitation increases as the maximization signal. The maximization signal is smaller for February 1986, possibly because that storm had a larger precipitation total before maximization than January 1997. We point out that the wet biases noted in section 3a above in the reconstructions represent 21% and 5% (February 1986 and January 1997, respectively) of the maximized totals; therefore, we reiterate that the February 1986 bias is not completely negligible.

c. *Ensemble runs (reconstructed and maximized) and assessment of uncertainty*

This section examines the two ensemble (56-member each) reconstructions of the February 1986 and January 1997 storms (shaded blue areas on Fig. 6), keeping in mind the issues with the baseline reconstruction of February 1986 (section 3a). We find that the February 1986 reconstructed

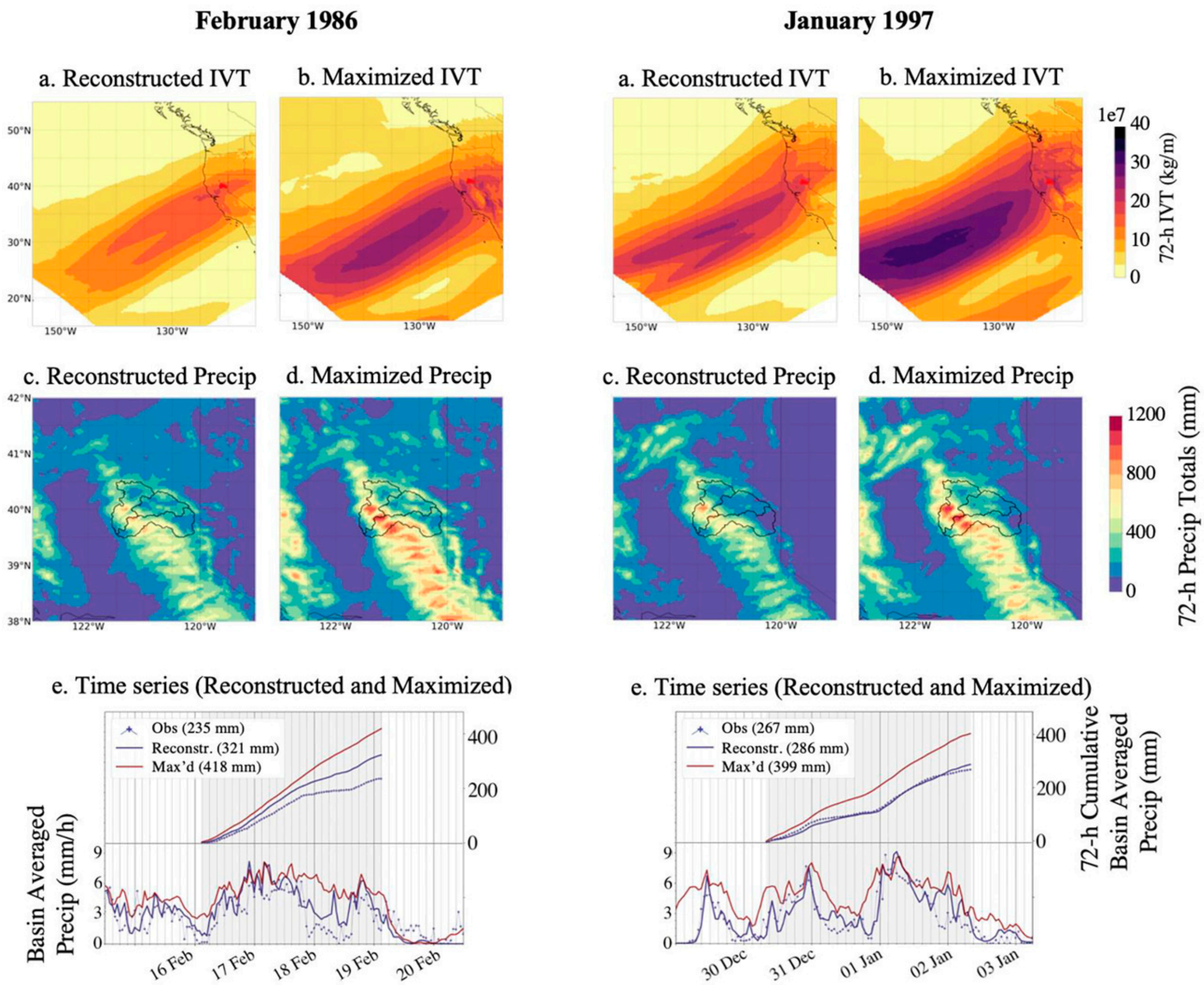

FIG. 5. Maximized simulations for the February 1986 and January 1997 storms show increased precipitation totals compared to reconstructed simulations. (a),(b) Maps of reconstructed and maximized 72-h total integrated vapor transport (IVT), respectively. (c),(d) Maps of reconstructed and maximized 72-h total precipitation, respectively. (e) Precipitation time series (observed, reconstructed, and maximized) with 72-h precipitation totals shown in the legend.

ensemble encounters the same issues as the baseline run (dashed blue lines): the majority of the observations (blue dots) are missed (Fig. 6). Even the lowest ensemble members overestimate precipitation during the trough on the first half of 18 February, which caused the large bias in the baseline reconstruction. This is despite the fact that the February 1986 ensemble has twice the amount of spread (118 mm) compared to January 1997 (55 mm). Expanding the ensemble to represent additional sources of uncertainty, e.g., attributable to parameterization schemes (see stochastic parameter perturbation, or SPPT: Palmer et al. 2009) in addition to SKEBS (which simulates upscale propagating errors), may explain why the current ensemble is missing observations. The reconstructed ensemble for January 1997, on the other hand, performs much better: it has less spread but encapsulates almost all of the precipitation observations. Irrespective of the storm, we note that the ensemble mean (solid blue lines) captures the temporal evolution of precipitation better than the baseline run (dashed blue lines). The fact that the ensemble mean performs better than even an extensively tested single configuration such as West-WRF highlights the value of such an ensemble in producing more robust PMP simulations.

We next examine the two maximized ensembles (shaded red areas on Fig. 6) and what they tell us about the robustness of the PMP estimates. The maximized ensembles are very similar between the two storms (unlike the reconstructed ensembles): there is almost the same spread among maximized runs for February 1986 (99 mm) and January 1997 (89 mm). The spread among maximized ensemble members amounts to ±7% of the ensemble mean estimate for both February 1986 and January 1997. The smaller maximization signal causes the maximized ensemble to overlap with reconstructions in February 1986, while they are clearly distinct in January 1997. The overlap raises questions as to the suitability of the PMP estimate obtained from the

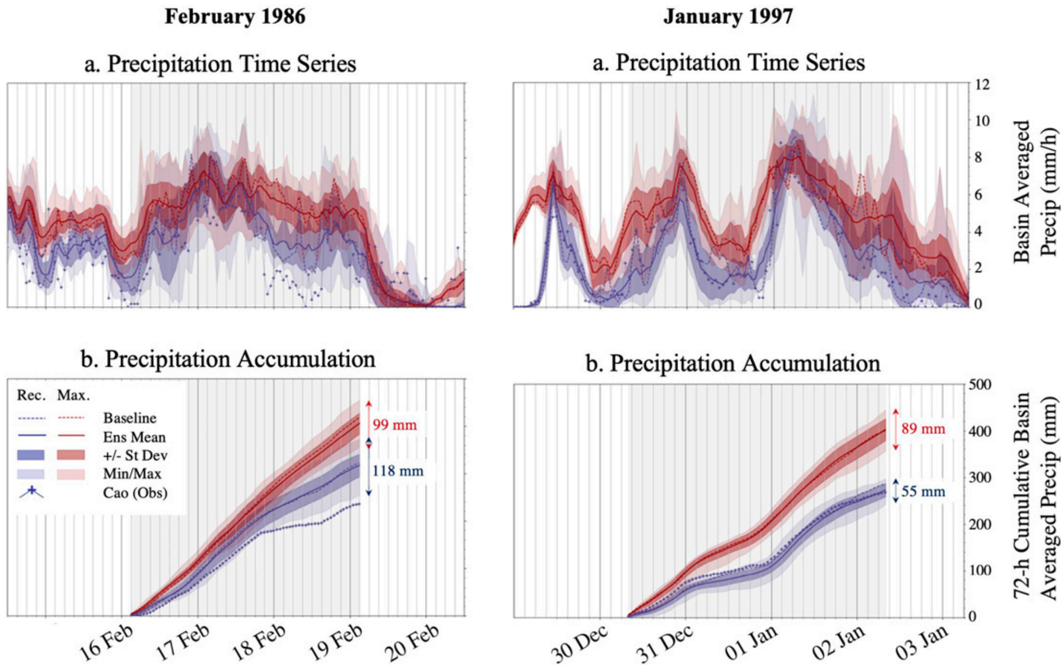

FIG. 6. Reconstructed (blue) and maximized (red) precipitation ensembles for the storms of February 1986 and January 1997: magnitude and spread of maximized precipitation totals are comparable for the two storms. (a) Precipitation time series. (b) Cumulative precipitation over the 72-h peak storm period. The arrows and numbers indicate the spread between the lowest and highest ensemble members.

February 1986 storm. That said, at this stage we recommend retaining storms like February 1986 for two reasons. First, this study applied the RHM technique only (additional storm maximization techniques are under development, see Toride et al. 2019), so it is possible that the PMP is in fact larger for both storms. Furthermore, the comparable behavior of the maximized simulations (magnitude of the totals and spread) in February 1986 and January 1997 suggests that most maximized storms for a given location may be similar, irrespective of which reconstruction they originate from. If confirmed by a larger sample of storms, this would indicate that so long as uncertainty is described, the PMP estimate is robust.

We conclude this section by giving an example of how our maximized precipitation ensembles can be interpreted to better understand the implications of model-based PMP uncertainty for dam safety. In Fig. 7, we show the ensemble mean and 90th percentile on histograms of 72-h precipitation resulting from maximization of the two storms, together with the single-value (West-WRF configuration, no perturbations) estimate. Our ensembles suggest that PMP is not likely to be much greater than the ensemble mean: the 90th percentile values are 432 mm for both storms, i.e., 107% of the ensemble mean for both storms. While there are many different ways to produce the ensembles that may yield various amounts of spread, this first assessment suggests that model uncertainty, often seen as a major barrier to the development of model-based PMP, may only have a modest impact on the PMP estimate.

4. Conclusions

The use of numerical weather models to estimate model-based PMP is an important advance over traditional PMP guidance (e.g., NOAA Hydrometeorological Reports, or HMRs) as it incorporates current understanding of the processes that control extreme precipitation. The model-based approach does, however, introduce new sources of uncertainty associated with initial conditions, upscale propagating errors, and the selection of alternative physics options in weather models like the Weather Research and Forecasting (WRF) Model used here. We evaluated these uncertainties in reconstruction and maximization of two very severe storms (February 1986 and January 1997) over the Feather River basin upstream of Oroville Dam. We first assessed the performance of a baseline (West-WRF, no perturbations) model configuration in reconstructing observed precipitation during these two events, and the impacts of any biases on the baseline maximized precipitation totals used for PMP estimation. We then designed ensembles of 56 reconstructed and 56 maximized simulations that captured model uncertainty using alternative model physics and stochastic perturbations for each of those two storms. The goal of our experiments was to describe how much the model-based PMP could be expected to vary due to the quality of the storm reconstructions and model uncertainties.

These analyses, the result of which we summarize here, allowed us to better characterize the robustness of model-based PMP estimates. We found that reconstructions of the storms of February 1986 and January 1997 generally match the

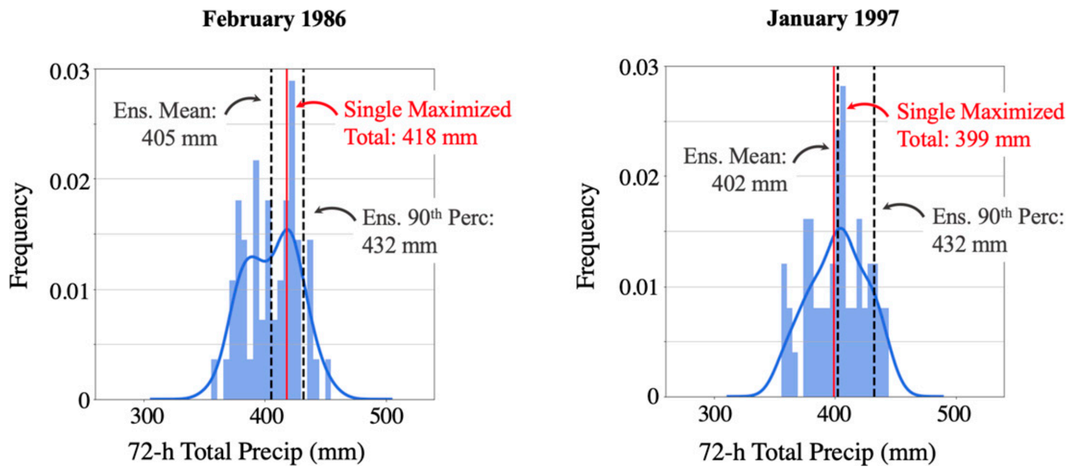

FIG. 7. Maximized precipitation totals histograms for the Feather River basin for the storms of February 1986 and January 1997. The histograms are produced using the 56 maximized simulations for each storm. The single-value (baseline, i.e., West-WRF without perturbations) maximized precipitation estimate is shown in red and the ensemble mean and 90th percentile values are shown in black.

spatial structure of observations well, but a wet bias exists that can be large (5% of maximized 72-h total precipitation for the 1997 storm but closer to 21% for the 1986 storm). Our maximized precipitation totals for both storms are similar in magnitude and uncertainty, i.e., they do not appear to be affected much by the characteristics of the historical storm or the quality of its reconstruction. In addition, uncertainty associated with initial conditions, upscale propagating errors and alternative physics options is modest, on the order of ±7% of the ensemble-mean model-based maximized estimate. Below are the main conclusions we draw from this analysis.

Our findings confirm that the quality of WRF precipitation reconstructions in the western United States is generally adequate for PMP estimation. This is especially true of atmospheric river (AR) storms interacting with topography, which we and earlier studies (Ishida et al. 2015a,b; Ohara et al. 2011; Toride et al. 2019) have shown can be reproduced well. Ongoing work addressing known issues with the precipitation efficiency of WRF microphysics schemes in the coastal western United States (English et al. 2021) as well as bias correction techniques (Chapman et al. 2019) is expected to bring further improvements. In fact, Lundquist et al. (2019) have shown that modeled precipitation can surpass the skill of observations in some cases. Therefore, biases in storm reconstructions should in most cases not be a greater concern than observation errors which may affect HMR PMP estimates.

We additionally believe that concerns about model uncertainty should not be a barrier to further development of model-based PMP. We have not attempted to decide on an acceptable amount of uncertainty as such a threshold would be context dependent. Instead, we suggest that characterization of uncertainty should become part of the PMP estimation process. This can be accomplished by use of ensemble methods such that the uncertainty range is known, and a single value, if needed, can be selected by the user according to the appropriate level of risk for a given application. While the currently in-use PMP methodology (Corrigan et al. 1999) was meant to be deterministic and extremely conservative in the absence of means to estimate its uncertainty, we believe that the ensemble approach we propose provides a first step toward re-evaluating PMP estimates in light of new information afforded by NWP modeling.

Although modest, the biases and uncertainties we identify highlight the importance of working with a large sample of historical storms. We saw that the quality of reconstructions, magnitude of the maximization signals and amounts of uncertainty differed among only two storms. Understanding how exactly the enforced moisture increases control changes in precipitation, depending on physical differences between storm systems, will also be needed to ensure the characteristics of uncertainty are captured. That said, an important insight from this study is that the historical storm reconstruction, whose moisture is “filled in” by the moisture maximization technique, has a limited impact on the maximized precipitation totals. If this pattern is confirmed across a larger sample of storms, it would imply that consistent model-based PMP estimates can be produced that may be approaching a physical limit, irrespective of the storm they are obtained from.

Importantly, model uncertainty is not the only source of uncertainty that needs to be evaluated in the model-based approach. Scaling moisture (rather than directly scaling precipitation as in the HMR approach) requires decisions as to how much moisture should be added, where, and for how long. These decisions and the variability they introduce in the maximized totals could also be represented as alternate ensemble members. Producing PMP ensembles that reflect both model (as in this analysis) as well as maximization uncertainty would provide a more complete picture of the robustness of model-based PMP. Characterizing model-based PMP uncertainty will help assess the extent to which model-based PMP can become a physically based alternative to HMR guidelines and at the same time support continued developments in the modeling of extreme precipitation events and flood risk analysis.

Acknowledgments. This work was funded by NASA FINESST Grant 80NSSC20K1663 to the University of California, Los Angeles (UCLA) and the first author. We additionally acknowledge support to UCLA from the Center for Western Weather and Water Extremes (CW3E) at the Scripps Institution of Oceanography UC San Diego via AR Program Phase 2, Grant 4600013361, sponsored by the California Department of Water Resources. We thank Zoran Micovic, John England, Kelly Mahoney, Michael Anderson, and Gary Estes for discussions that helped frame this study and providing comments on the manuscript; Daniel Steinhoff for recommendations on WRF configuration and the design of the ensembles; and Xiaodong Chen, Blake Allen, and Stefan Rahimi for guidance on the moisture maximization technique. This work used the COMET supercomputer, which was made available by the Atmospheric River Program Phase 2 and 3 supported by the California Department of Water Resources (Awards 4600013361 and 4600014294, respectively) and the Forecast Informed Reservoir Operations Program supported by the U.S. Army Corps of Engineers Engineer Research and Development Center (Award USACE W912HZ-15-2-0019).

Data availability statement. The ERA5 dataset used to force the WRF Model runs was obtained through the NCAR/UCAR Research Data Archive.

REFERENCES

Abbs, D. J., 1999: A numerical modeling study to investigate the assumptions used in the calculation of probable maximum precipitation. *Water Resour. Res.*, **35**, 785–796, https://doi.org/10.1029/1998WR900013.

Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626, https://doi.org/10.1175/2008JAS2677.1.

——, S.-Y. Ha, J. P. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139**, 1972–1995, https://doi.org/10.1175/2010MWR3595.1.

——, K. R. Fossell, S.-Y. Ha, J. P. Hacker, and C. Snyder, 2015: Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Mon. Wea. Rev.*, **143**, 1295–1320, https://doi.org/10.1175/MWR-D-14-00091.1.

Caldwell, P., H.-N. S. Chin, D. C. Bader, and G. Bala, 2009: Evaluation of a WRF dynamical downscaling simulation over California. *Climatic Change*, **95**, 499–521, https://doi.org/10.1007/s10584-009-9583-5.

Cao, Q., A. Mehran, F. M. Ralph, and D. P. Lettenmaier, 2019: The role of hydrological initial conditions on atmospheric river floods in the Russian River basin. *J. Hydrometeor.*, **20**, 1667–1686, https://doi.org/10.1175/JHM-D-19-0030.1.

Chapman, W. E., A. C. Subramanian, L. Delle Monache, S. P. Xie, and F. M. Ralph, 2019: Improving atmospheric river forecasts with machine learning. *Geophys. Res. Lett.*, **46**, 10 627–10 635, https://doi.org/10.1029/2019GL083662.

Chen, L.-C., and A. A. Bradley, 2006: Adequacy of using surface humidity to estimate atmospheric moisture availability for probable maximum precipitation. *Water Resour. Res.*, **42**, W09410, https://doi.org/10.1029/2005WR004469.

Chen, X., and F. Hossain, 2016: Revisiting extreme storms of the past 100 years for future safety of large water management infrastructures. *Earth's Future*, **4**, 306–322, https://doi.org/10.1002/2016EF000368.

Cobb, A., L. Delle Monache, F. Cannon, and F. M. Ralph, 2021: Representation of dropsonde-observed atmospheric river conditions in reanalyses. *Geophys. Res. Lett.*, **48**, e2021GL093357, https://doi.org/10.1029/2021GL093357.

Corrigan, P., D. D. Fenn, D. R. Kluck, and J. L. Vogel, 1999: Probable maximum precipitation for California. NOAA Hydrometeorological Rep. 59, 419 pp., https://www.weather.gov/media/owp/hdsc_documents/PMP/HMR59.pdf .

English, J. M., D. D. Turner, T. I. Alcott, W. R. Moninger, J. L. Bytheway, R. Cifelli, and M. Marquis, 2021: Evaluating operational and experimental HRRR model forecasts of atmospheric river events in California. *Wea. Forecasting*, **36**, 1925–1944, https://doi.org/10.1175/WAF-D-21-0081.1.

Enzel, Y., L. L. Ely, P. K. House, V. R. Baker, and R. H. Webb, 1993: Paleoflood evidence for a natural upper bound to flood magnitudes in the Colorado River basin. *Water Resour. Res.*, **29**, 2287–2297, https://doi.org/10.1029/93WR00411.

Gangrade, S., S.-C. Kao, T. T. Dullo, A. J. Kalyanapu, and B. L. Preston, 2019: Ensemble-based flood vulnerability assessment for probable maximum flood in a changing environment. *J. Hydrol.*, **576**, 342–355, https://doi.org/10.1016/j.jhydrol.2019.06.027.

Grell, G. A., and D. Dévényi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.*, **29**, 1693, https://doi.org/10.1029/2002GL015311.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, https://doi.org/10.1175/MWR3199.1.

Ishida, K., M. L. Kavvas, F. Asce, S. Jang, Z. Q. Chen, A. M. Asce, N. Ohara, and M. L. Anderson, 2015a: Physically based estimation of maximum precipitation over three watersheds in Northern California: Atmospheric boundary condition shifting. *J. Hydrol. Eng.*, **20**, 04014052, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001026.

——, ——, S. Jang, Z. Q. Chen, N. Ohara, and M. L. Anderson, 2015b: Physically based estimation of maximum precipitation over three watersheds in Northern California: Relative humidity maximization method. *J. Hydrol. Eng.*, **20**, 04015014, https://doi.org/10.1061/(ASCE)HE.1943-5584.0001175.

Jankov, I., J.-W. Bao, P. J. Neiman, P. J. Schultz, H. Yuan, and A. B. White, 2009: Evaluation and comparison of microphysical algorithms in ARW-WRF Model simulations of atmospheric river events affecting the California coast. *J. Hydrometeor.*, **10**, 847–870, https://doi.org/10.1175/2009JHM1059.1.

Jiménez, P. A., J. Dudhia, J. F. González-Rouco, J. Navarro, J. P. Montávez, and E. García-Bustamante, 2012: A revised scheme for the WRF surface layer formulation. *Mon. Wea. Rev.*, **140**, 898–918, https://doi.org/10.1175/MWR-D-11-00056.1.

Leung, L. R., and Y. Qian, 2009: Atmospheric rivers induced heavy precipitation and flooding in the western U.S. simulated by the WRF regional climate model. *Geophys. Res. Lett.*, **36**, L03820, https://doi.org/10.1029/2008GL036445.

Lundquist, J., M. Hughes, E. Gutmann, and S. Kapnick, 2019: Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks. *Bull. Amer. Meteor. Soc.*, **100**, 2473–2490, https://doi.org/10.1175/BAMS-D-19-0001.1.

Mahoney, K., C. McColl, D. M. Hultstrand, W. D. Kappel, B. McCormick, and G. P. Compo, 2021: Blasts from the past: Reimagining historical storms with model simulations to modernize dam safety and flood risk assessment. *Bull. Amer. Meteor. Soc.*, **103**, E266–E280, https://doi.org/10.1175/BAMS-D-21-0133.1.

Martin, A., F. M. Ralph, R. Demirdjian, L. DeHaan, R. Weihs, J. Helly, D. Reynolds, and S. Iacobellis, 2018: Evaluation of atmospheric river predictions by the WRF Model using aircraft and regional Mesonet observations of orographic precipitation and its forcing. *J. Hydrometeor.*, **19**, 1097–1113, https://doi.org/10.1175/JHM-D-17-0098.1.

MGS Engineering Consultants, Inc., 2005: Stochastic modeling of extreme floods on the American river at Folsom dam: Flood-frequency curve extension. USACE Research Doc. 48, 54 pp., https://www.hec.usace.army.mil/publications/ResearchDocuments/RD-48.pdf.

Michaelis, A. C., A. C. Martin, M. A. Fish, C. W. Hecht, and F. M. Ralph, 2021: Modulation of atmospheric rivers by mesoscale frontal waves and latent heating: Comparison of two U.S. West Coast events. *Mon. Wea. Rev.*, **149**, 2755–2776, https://doi.org/10.1175/MWR-D-20-0364.1.

Micovic, Z., M. G. Schaefer, and G. H. Taylor, 2015: Uncertainty analysis for probable maximum precipitation estimates. *J. Hydrol.*, **521**, 360–373, https://doi.org/10.1016/j.jhydrol.2014.12.033.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, https://doi.org/10.1029/97JD00237.

Morrison, H., G. Thompson, and V. Tatarskii, 2009: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Mon. Wea. Rev.*, **137**, 991–1007, https://doi.org/10.1175/2008MWR2556.1.

Newman, A. J., and Coauthors, 2015: Gridded ensemble precipitation and temperature estimates for the contiguous United States. *J. Hydrometeor.*, **16**, 2481–2500, https://doi.org/10.1175/JHM-D-15-0026.1.

Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116**, D12109, https://doi.org/10.1029/2010JD015139.

O'Connor, J. E., G. E. Grant, and J. E. Costa, 2002: The geology and geography of floods. *Ancient Floods, Modern Hazards: Principles and Applications of Paleoflood Hydrology*, P. K. House et al., Eds., Vol. 5, Amer. Geophys. Union, 359–385.

Ohara, N., and Coauthors, 2011: Physically based estimation of maximum precipitation over American river watershed, California. *J. Hydrol. Eng.*, **16**, 351–361, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000324.

——, M. L. Kavvas, M. L. Anderson, Z. Q. Chen, and K. Ishida, 2017: Characterization of extreme storm events using a numerical model-based precipitation maximization procedure in the feather, Yuba, and American river watersheds in California. *J. Hydrometeor.*, **18**, 1413–1423, https://doi.org/10.1175/JHM-D-15-0232.1.

Palmer, T. N., R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer, 2009: Stochastic parametrization and model uncertainty. ECMWF Tech. Memo. 598, 44 pp., http://www.ecmwf.int/sites/default/files/elibrary/2009/11577-stochastic-parametrization-and-model-uncertainty.pdf.

Pleim, J. E., 2007: A combined local and nonlocal closure model for the atmospheric boundary layer. Part I: Model description and testing. *J. Appl. Meteor. Climatol.*, **46**, 1383–1395, https://doi.org/10.1175/JAM2539.1.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102, https://doi.org/10.1256/qj.04.106.

——, M. Leutbecher, A. Weisheimer, T. Stockdale, L. Isaksen, and M. Bonavita, 2011: Representing model uncertainty: Stochastic parametrizations at ECMWF. *ECMWF Newsletter*, No. 129, ECMWF, Reading, United Kingdom, 19–24, https://doi.org/10.21957/fbqmkhv7.

Skamarock, W. C., and J. B. Klemp, 2008: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.*, **227**, 3465–3485, https://doi.org/10.1016/j.jcp.2007.01.037.

Smith, J. A., and M. L. Baeck, 2015: "Prophetic vision, vivid imagination": The 1927 Mississippi River flood. *Water Resour. Res.*, **51**, 9964–9994, https://doi.org/10.1002/2015WR017927.

Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, https://doi.org/10.1175/2008MWR2387.1.

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2.

Toride, K., Y. Iseri, M. D. Warner, C. D. Frans, A. M. Duren, J. F. England, and M. L. Kavvas, 2019: Model-based probable maximum precipitation estimation: How to estimate the worst-case scenario induced by atmospheric rivers? *J. Hydrometeor.*, **20**, 2383–2400, https://doi.org/10.1175/JHM-D-19-0039.1.

USWB, 1961: Interim report: Probable maximum precipitation in California. USWB Hydrometeorological Rep. 36, 226 pp., https://www.weather.gov/media/owp/oh/hdsc/docs/HMR36.pdf.

WMO, 2009: Manual on estimation of probable maximum precipitation (PMP). 3rd ed. Tech. Rep. WMO-1045, 291 pp., https://damfailures.org/wp-content/uploads/2020/10/WMO-1045-en.pdf.