Title

The “Perfect Storm”: Leveraging Atmospheric Models to Improve Probable Maximum Precipitation (PMP) Estimates

Author

Tarouilly, Emilie G

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The “Perfect Storm”:

Leveraging Atmospheric Models to Improve Probable Maximum Precipitation (PMP) Estimates

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in Civil Engineering

by

Emilie Gaelle Margot Tarouilly

2023

ABSTRACT OF THE DISSERTATION

The “Perfect Storm”:

Leveraging Atmospheric Models to Improve Probable

Maximum Precipitation (PMP) Estimates

by

Emilie Gaelle Margot Tarouilly

Doctor of Philosophy in Civil Engineering

University of California, Los Angeles, 2023

Professor Dennis P. Lettenmaier, Chair

The flood that would result from the greatest depth of precipitation “meteorologically possible”, or Probable Maximum Precipitation (PMP) is used in the design of dam spillways and other high-risk structures. Historically, PMP has been estimated by scaling precipitation depth-area-duration relationships obtained from severe historical storms. Over the last decade, numerical weather prediction models have been leveraged to instead predict precipitation resulting from the addition of moisture at the model boundaries (called relative humidity maximization, or RHM). Despite the major improvement this represents, model-based methods have not yet been applied to produce operational PMP estimates. Several sources of uncertainty still limit their applicability: (i) model uncertainty, (ii) uncertainty regarding whether maximum storm efficiency (moisture conversion to precipitation) was achieved by observed storms and (iii) uncertainty caused by a lack of physically-constrained

guidance on moisture amplification. These uncertainties result in differences in model-based PMP estimates among studies that used different model setups, datasets and moisture amplification approaches, which limits confidence in the estimates: could PMP be much larger than a given study indicates? Or are we over-designing for unrealistically large events? Focusing on mountainous watersheds in the coastal western U.S. affected by atmospheric river (AR) storms (the Feather and Willamette River basins among others), I seek to identify the largest sources of uncertainty affecting model-based PMP estimates and ways to reduce these uncertainties. Using the Weather Research and Forecasting (WRF) regional atmospheric model, I performed experiments in which I reconstructed and amplified several severe storms. First (Chapter 2), I produced an ensemble of simulations to characterize model uncertainty (parametrization and propagation of model error). I found that the spread in PMP totals among ensemble members was modest, ranging from +/- 7% of the ensemble mean. Next (Chapter 3), I simulated artificial storms selected from a large sample (~1500 years) produced by a global climate model (the CESM2 large ensemble, or CESM2-LE). I found that the largest PMP estimates they produced was only 8% larger than estimates obtained from observed storms. Finally (Chapter 4), I found that the way moisture maximization is implemented has the largest impact on PMP estimates (with saturation of the atmosphere resulting in up to 30% larger precipitation totals than other approaches). I as a result developed a method that uses climatology to objectively determine the magnitude of moisture amplification. The understanding of the uncertainties in model-based PMP estimates and the tools I developed to characterize and reduce these uncertainties provide key information for the development of more reliable model-based PMP estimation guidance.

The dissertation of Emilie Gaelle Margot Tarouilly is approved.

Alexander Hall

Steven Margulis

Timu Gallien

Dennis P. Lettenmaier, Committee Chair

University of California, Los Angeles

2023

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am grateful to my advisor Dennis Lettenmaier for giving me the opportunity to come to UCLA and undertake this work. I thank him and my committee members Steve Margulis, Timu Gallien and Alex Hall for their guidance. I especially want to thank Stefan Rahimi, Daniel Steinhoff and Forest Cannon who taught me how to run the WRF model.

I gratefully acknowledge all those who informally helped me and gave me advice, including Michael Anderson, Xiaodong Chen, Rory Nathan, Conrad Wasko, Ed Tomlinson, Kevin Griebenow, Chandra Pathak, Sasha Gershunov, Zoran Micovic, Mel Schaefer, Kelly Mahoney, Katie Holman and Jay Cordeira.

I want to thank my labmates (Zhaoxin Ban, Solomon Vimal, Qian Cao and Mu Xiao in particular) and other students and postdocs at UCLA (Jessica Fayne, Sarfaraz Alam, Akash Koppa, Kayden Haleakala, Yiwen Fang, Maria Winters, Dongyue Li) for being there for me when I needed them. Special thanks to Jessica Fayne for convincing me to apply for the FINESST fellowship.

While not involved in my PhD research, I want to thank my Masters advisors Ivan Stoianov and Danny O'Hare for making me realize that I enjoyed doing research.

A portion of my work was performed in collaboration with the Center for Western Weather and Water Extremes (CW3E) at UCSD. My PhD work was in part funded by a NASA Future Investigators in Earth and Space Science and Technology (FINESST) Grant 80NSSC20K1663.

Curriculum Vitae

EDUCATION

MSc Civil and Environmental Engineering | Imperial College London (UK), 2015

BSc Biology and Ecology | Imperial College London (UK), 2014

INDUSTRY EXPERIENCE

Thames Water Utilities, Water Resources Modeler (London, UK) (2015-2017)

CONFERENCE PRESENTATIONS

Am. Meteorological Society (AMS) Annual Meeting (2023)

Am. Geophys. Union (AGU) Fall Meeting (2019, 2020, 2021, 2022)

European Geophys. Union (EGU) General Assembly (2021, 2022)

TEACHING EXPERIENCE

Teaching Assistant, Climate Dynamics (Atmospheric Sciences Department) (2019-2021)

AWARDS

American Water Works Association (AWWA) scholarship, $10,000 (2019)

Future Investigators in NASA Earth Sciences and Technology (FINESST) Fellowship, $135,000 (2019-2022)

Dimitris N. Chorafas Foundation award, $5,000 (2023)

PUBLICATIONS

Physically-Constrained Model-Based Moisture Amplification for Probable Maximum Precipitation (PMP) Estimation. Tarouilly E., Holman K., Lettenmaier D. P., *Journal of Hydrometeorology*. In review.

Did Historical Storms Used in Probable Maximum Precipitation (PMP) Estimation Reach Maximum Efficiency? A Large Model Ensemble Approach. Tarouilly E., Rahimi S., Cordeira J., Lettenmaier, D. P., *Journal of Geophysical Research*, In review.

Improving Confidence in Model-Based Probable Maximum Precipitation: How Important is Model Uncertainty in Storm Reconstruction and Maximization? 2022, Tarouilly, E., Cannon, F. and Lettenmaier, D. P., *Journal of Hydrometeorology*. https://doi.org/10.1175/JHM-D-22-0044.1

Western U.S. Superfloods in the Recent Instrumental Record. 2021, Tarouilly E., Li D., Lettenmaier D. P., *Water Resources Research.* https://doi.org/10.1029/2020wr029287.

Chapter 1: Introduction

How severe can a flood be in a given river basin? Hydrologist Robert Horton stated nearly a century ago that "it is not difficult to show that […] there is a natural limitation to rain intensity" (letter to F. Schmidt, 18 Nov. 1927), yet upper bounds to either rain or floods, if any, have not been identified. The question, however, is of both fundamental interest and practical importance: any critical infrastructure the failure of which would be catastrophic (including large dams but also nuclear power plants) must be able to pass the most severe flood possible (or our best estimate of it). In the U.S., this is the purpose of the concept of probable maximum flood (PMF), "a flood that can be expected from the most severe combination of critical meteorologic and hydrologic conditions that are reasonably possible in a region" (FERC, 2001).

The PMF is obtained by combining the probable maximum precipitation (PMP) with wet antecedent hydrologic conditions, typically the 100-year return period soil moisture or snowpack. PMP is defined by the WMO (2009) as the "maximum depth of precipitation that can occur over a specified area". As the upper bound of precipitation is unknown, the PMP is only an approximation of it for engineering purposes (e.g., the sizing of dam spillways). PMP estimates for the U.S. and their estimation procedures are described in NOAA's hydrometeorological reports (HMRs, e.g., HMR59 for California; Corrigan et al., 1999).

Following HMR guidelines (e.g., HMR 57 for the Pacific Northwest; Hansen et al., 1994, and HMR59 for California; Corrigan et al., 1999), PMP estimation starts with the most severe storms that were observed at a given location. Storms from nearby areas within a "homogenous region" may be transposed over the river basin of interest to identify as many relevant severe storms as possible. The storms are then amplified, with the implied

assumption that historical (observed) storms achieved maximum storm efficiency (moisture conversion to precipitation) but may have been moisture-limited. Therefore, the storms' depth-area-duration (D-A-D) relationships are scaled according to how much moisture could have been available, typically the ratio of the climatological maximum dew point to the dew point temperatures during the storm. HMR procedures often additionally include steps for enveloping precipitation from the scaled D-A-D relationships. The end product is an index map of 10-mi^2, 24-hr PMP values for a region (see HMR 59 for an example of the index map for California; Corrigan et al., 1999), which can then be apportioned to the relevant area and duration for a basin of interest by the user.

The HMR procedures date to the beginning of the era of major dam construction in the U.S. (see e.g. HMR 41 for the Tennessee River basin; Schwarz et al., 1965) and were developed assuming limited knowledge of precipitation processes and available data. As a result, they are based on many assumptions, including the use of surface dew point temperatures with the pseudo-adiabatic assumption to determine dew points in the atmospheric column, and the assumption that precipitation scales linearly with moisture increases. These assumptions have since been questioned by a number of numerical weather prediction model studies (e.g., Zhao et al., 1997; Yang & Smith, 2018; Chen & Bradley, 2006; Kim et al., 2020). Additionally, over 20 years of new data, possibly containing more severe storms, have become available since the most recent updates to PMP estimates in 1999 (HMR 59; Corrigan et al., 1999). As a result, there is an urgent need to re-evaluate and modernize PMP guidelines to better reflect the latest data and current understanding of precipitation processes.

Numerical weather prediction models provide a new tool to test hypotheses about the controls on extreme precipitation and to design more physically-based PMP estimation procedures accordingly. The Weather Research and Forecasting (WRF) model (Skamarock & Klemp,

2008) is the current standard in model-based PMP studies. WRF (and its predecessor the MM5 model; Grell et al., 1994) were developed by the National Center for Atmospheric Research (NCAR) to become an operational forecast model as well as a research tool for the weather and climate communities. WRF is a regional model for which boundary conditions must be specified by a global dataset, such as a reanalysis dataset when the goal is to reconstruct past conditions (as is the case here) or a global climate model (for future climate projections). Here, I use the ERA5 reanalysis (1950-2020) produced by the ECMWF global atmospheric model with observation data (including satellite data) assimilated into the model runs (Hersbach et al., 2020).

In the context of climate change and population increases, dams are essential to both water supply and flood control, particularly in the U.S. West where precipitation patterns are highly variable. Numerical weather prediction models such as WRF perform particularly well in coastal parts of the Western U.S. affected by atmospheric river (AR) storms, i.e., synoptic scale weather patterns interacting with topography (in comparison with e.g., convective storms in the U.S. interior). I therefore focus on four river basins: the Feather River basin (Oroville dam) in California (Chapters 2 and 3), the Coos River (Pony Creek dam), the Illinois River (McMullen dam) and the Willamette River upstream of Salem (multiple large dams), all in Oregon (Chapter 4). The need for new PMP estimates in that area is motivated by increasing understanding of precipitation processes in AR storms (Ralph et al., 2019) and concerns about the safety of several high-hazard dams. Another motivating factor for updating the methodology underlying PMP estimates is the impact of global warming on moisture availability via the Clausius-Clapeyron relationship, which is expected to cause an increase in PMP estimates and makes relying solely on past observations increasingly untenable. While this thesis focuses on current climate PMP, the methods developed here can

be used directly to estimate future climate PMP using storms from future climate projections instead of historical reanalyzes.

As currently implemented, model-based PMP follows the logic of HMR guidelines in transposing and amplifying the moisture of a severe historical storm. However, model-based PMP differs from HMR guidelines in that the model formulation is used to predict precipitation that results from moving a severe historical storm to a new location and adding moisture at the WRF model boundaries. While model-based approaches represent a significant improvement over HMR approaches, they have not yet been used to produce operational estimates, in part due to the uncertainties that are suspected to affect them: characterizing and reducing these uncertainties is the goal of this thesis.

A key source of uncertainty in model-based PMP estimates is model uncertainty: 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 among reconstructed storm totals and hence in PMP estimates. The fact that the PMP studies cited above (Ohara et al. 2011; Ishida et al. 2015a,b) have used different models (including MM5, a predecessor of the WRF Model) and setups (domains, resolution, parametrizations) may explain differences in the PMP estimates they produce, but this to my knowledge has not been assessed. My approach (Chapter 2) is to design an ensemble of simulations that captures those important sources of uncertainty (different ensemble members may have e.g., different parameter sets and/or perturbations introduced at the beginning of the simulation) such that their impact on model-based PMP estimates can be quantified.

Another source of uncertainty is that storm efficiency (moisture conversation to precipitation) is assumed to be already maximized in the severe historical storms used as the basis for PMP

estimation. Given the relatively short duration of the observed historical record (typically less than 100 years), it is essentially certain that storms with higher efficiencies than those observed could occur. I propose that a PMP storm – or a reasonably similar storm – can be found, in which case there is no need for amplification of either moisture or storm efficiency, so long as a large enough storm sample is available. In Chapter 3, I examine storms from the CESM2 large ensemble (CESM2-LE) of global model runs for the historical (1900-2014) period, effectively pooling a historical climatology that is ~16.5 times the length of ERA5. The larger sample size from the CESM2-LE allows for an approximately 1-in-1000-year storm to be examined within a PMP context to inform more robust PMP estimates.

Finally, another important barrier in model-based PMP estimation is that there currently exist multiple different implementations of moisture maximization, which involve adding different amounts of moisture at different locations and which produce different amplified precipitation estimates. Some authors increase moisture all the way up to saturation (relative humidity maximization, or RHM; Ishida et al., 2015), while others use a multiplicative factor to increase moisture where it was already present (relative humidity perturbation, or RHP) which they suggest more realistically preserves the storm characteristics than RHM but do not provide guidance on how to select the multiplier (Toride et al., 2019). The reason for these different possible approaches is that while the model provides a quantitative, reproducible way to convert moisture to precipitation, how moisture is added remains a somewhat subjective decision. Saturation is arguably a physical upper bound but may not be realistic. To address this issue, I propose (Chapter 4) an approach which I call RHP-ratio that builds on RHP but uses historical maximum moisture as a way to consistently determine the magnitude of the moisture increases.

The overarching goal of my dissertation is to characterize sources of uncertainty affecting model-based PMP estimates and to develop estimation guidelines that are constrained by physical reasoning to improve consistency, defensibility and reproducibility. I hope that these improved model-based guidelines will ultimately provide a robust alternative to HMR methods that are better in-line with our evolving knowledge of precipitation processes and applicable in a changing climate.

Chapter 2: Improving Confidence in Model-Based Probable Maximum Precipitation: How Important is Model Uncertainty in Storm Reconstruction and Maximization?

A version of this chapter has been published in the *Journal of Hydrometeorology*: Tarouilly, E., Cannon, F. and Lettenmaier, D. P. (2022). "Improving Confidence in Model-Based Probable Maximum Precipitation: How Important is Model Uncertainty in Storm Reconstruction and Maximization?", 24, 257-267, DOI: 10.1175/JHM-D-22-0044.1

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 U.S., 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 & 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 & Hossain, 2016). These developments are prompting a re-evaluation of PMP estimates, starting with the use of model-derived precipitation to fill-in data gaps and apply HMR PMP methods when observations are limited (e.g., Mahoney et al., 2021 for Colorado and New Mexico).

New methods to estimate PMP directly using the Weather and Forecasting (WRF) model (Skamarock et al. 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., 2014; 2015), consists of 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 hours) 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 parametrizations, 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, and 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., 2014; 2015) 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, 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 & 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, then produce an ensemble of PMP values (rather than a single estimate) in order to adequately reflect uncertainty. The science questions we address are:

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 parametrizations, and model errors)?

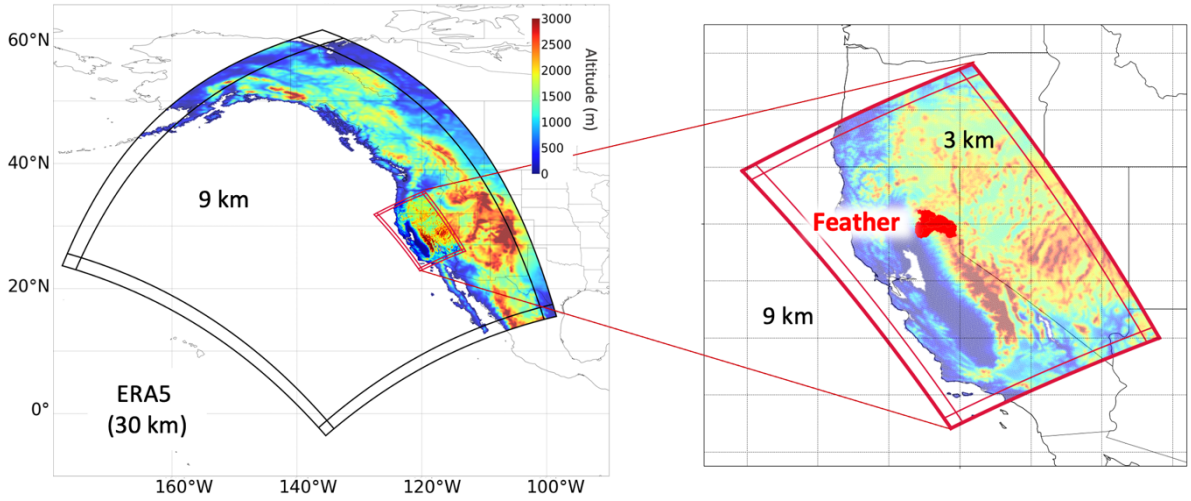
3. How important is this uncertainty relative to the size of the maximization signal?

2. Methods

2.a. Study area and storms of interest

Our study area is the Feather River watershed (3600 sq mi) upstream of Oroville dam, California (Figure 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., 2014; 2015; Ohara et al., 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 & 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-hour precipitation totals, respectively.



Chap. 2. 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.

2.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 et al., 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 U.S. (Figure 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 in this region (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 used in the maximization.

Chap. 2. Table 1. West-WRF physics parametrizations 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 Devenyi, 2002)
Boundary layer scheme	Yonsei University (Hong et al. 2006)
Long-wave physics	RRTMG (Mlawer et al. 1997)
Short-wave physics	RRTMG (Mlawer et al. 1997)
Surface layer physics	Revised MM5 (Jimenez et al. 2012)
Land surface physics	Unified Noah Land Surface Model (Niu et al., 2011)

2.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 degree) 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). 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.

2.d. Baseline storm maximization for PMP estimation

The maximized storm simulations were produced using a technique developed by Ishida et al. (2015) called Relative Humidity Maximization (RHM). It is currently the most widely used model-based PMP technique (see e.g., Gangrade et al., 2017). 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 the 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-hour total precipitation (for the same window that produced the highest observed totals) produced by 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 continued to work with both storms throughout this paper and refer to "maximized" rather than "PMP" totals.

2.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 model precipitation: boundary and initial condition errors, model parametrization and upscale propagating errors

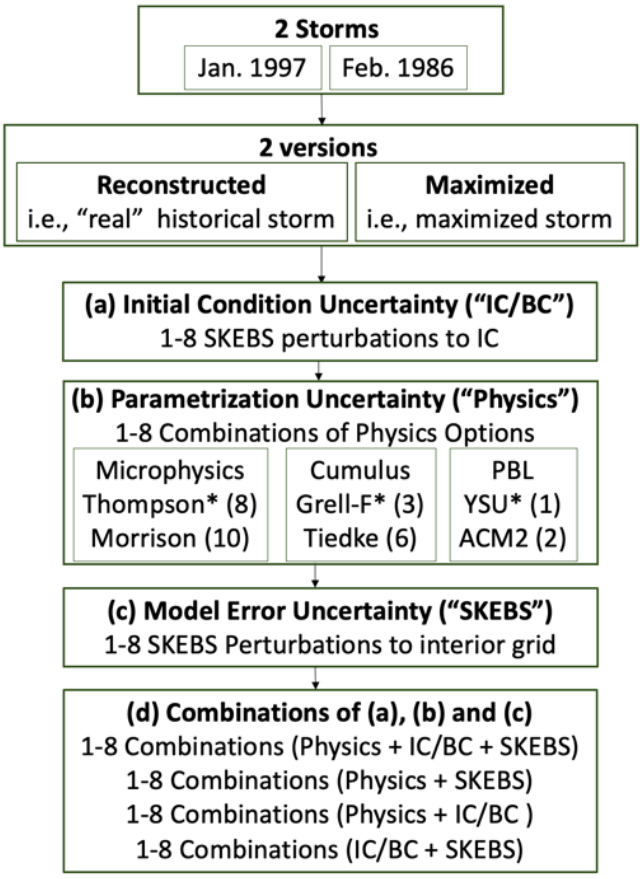
(Berner et al., 2015; Cannon et al., 2022). The remainder of this section describes the design of the ensembles, which aims to balance computational cost with the generation of realistic spread. Figure 2 provides an overview of all experiments, which resulted in 56 ensemble members for each version (reconstructed and maximized) of each storm.

Uncertainty due to the choice of physics parametrization was assessed by testing different combinations of possible microphysics, cumulus and boundary layer schemes. These are the parametrizations 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 Devenyi, 2002)- 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 (Figure 2, box b). This choice of schemes is guided by Cannon et al. (2022), who identified scheme 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 parametrization, 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 stream function 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) (Figures 2, boxes a and c, respectively).

In addition to ensemble members that consist of either different physics parametrization, or initial/boundary condition perturbations, or interior grid perturbations alone, we produced runs that consist of various combinations of the experiments above (Figure 2, box d). 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 parametrization, upscale propagating errors or a combination) but rather to assess the magnitude of the uncertainty produced by all of those different sources together.



Chap. 2. Fig. 2. 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 box (b), the asterisks denote the physics options used in the West-WRF setup and numbers in parentheses refer to the WRF parametrization code.

3. Results

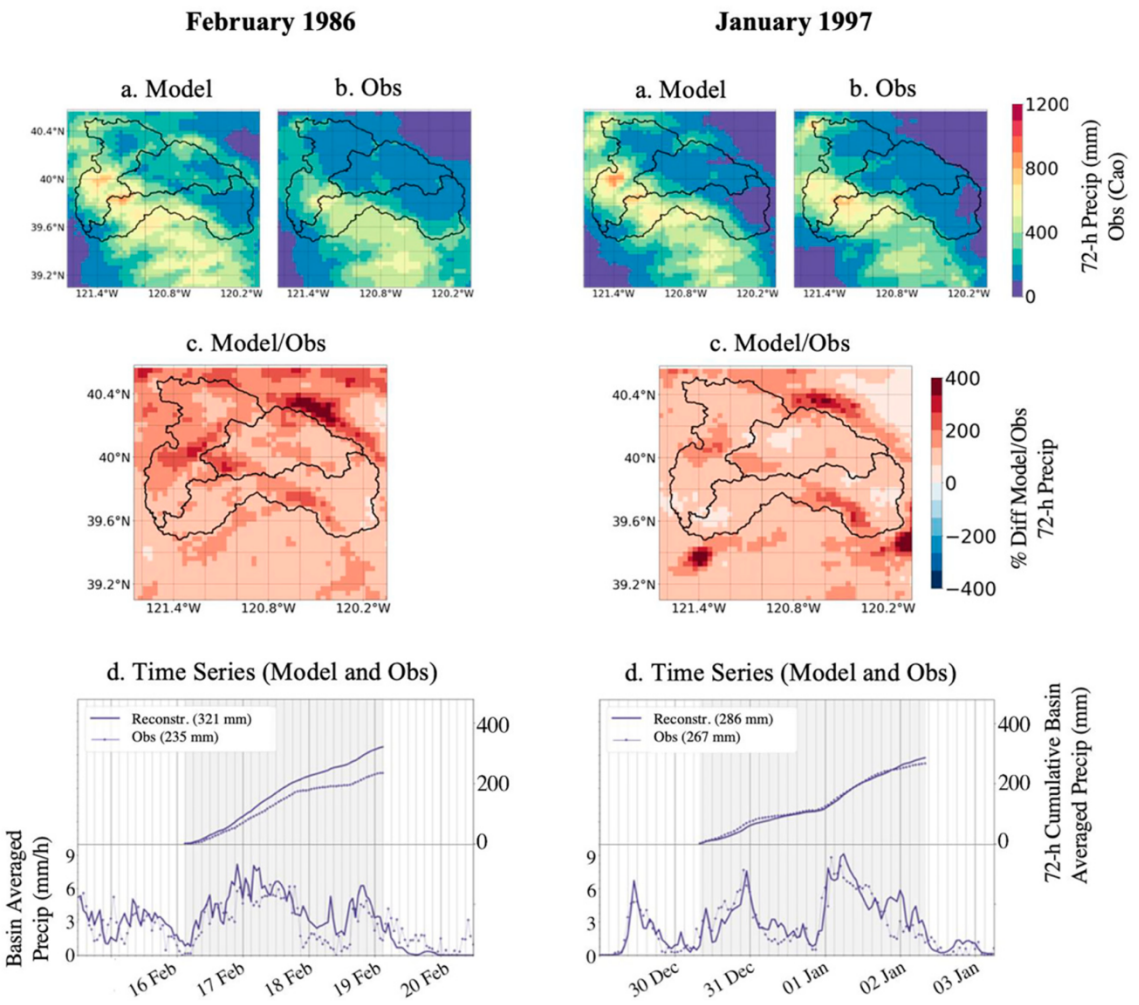
3.a. Baseline storm reconstructions

3.a.1. Model evaluation

We first assessed performance of the baseline storm reconstructions (West-WRF, no perturbations) against which we subsequently (Section 3c) 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

(Figure 3d, bottom row). The location of precipitation centers (Figure 3a and b, top row) is also well represented, but the model has a wet bias at mid-elevations (Figure 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 all 72 hours during the storm) exceed 0.7 for February 1986 and 0.8 for January 1997, as has been pointed out before (e.g., Ishida et al. (2014) 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 biases in 72-hour precipitation totals, which are arguably an important characteristic to reconstruct for PMP estimation. The reconstructed basin-averaged 72-hour 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 modelled precipitation during the first half of February 18. On the other hand, the reconstructed basin-averaged 72-hour 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. (2014) 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.



Chap. 2. 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-hour total precipitation. a-b: Maps of modeled and observed 72-hour total precipitation, respectively. c: Maps of percentage difference (modeled/observed). d: Precipitation time series (modeled and observed) with 72-hour precipitation totals shown in the legend.

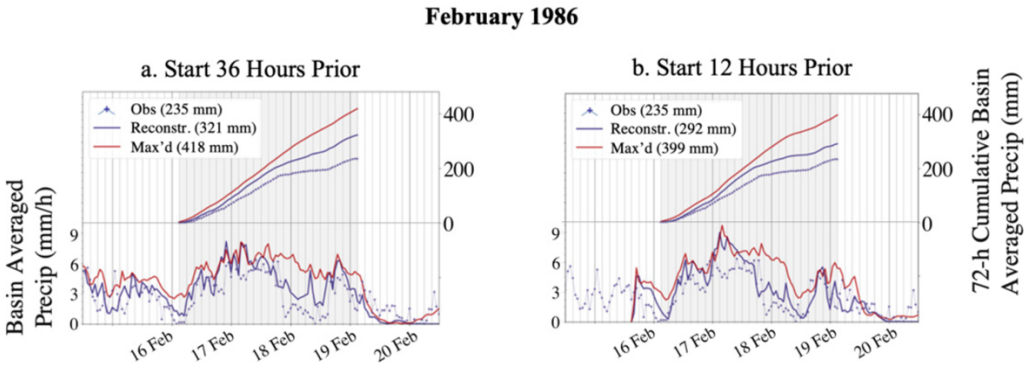
The performance of our reconstructions reflects the current capabilities and limitations of mesoscale precipitation modeling. The wet bias we observe in the Sierra Nevada is acknowledged by a large body of work on WRF performance (e.g., Cannon. et al., 2022; English et al., 2021; Caldwell et al., 2009; Jankov et al., 2009). Further 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 timescale. 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., 2010) 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.

3.a.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-hour storm period would improve modeled precipitation, and whether this would have an impact on the maximized total. While the model does require some spin-up, 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-hour basin-averaged total decreases from 37% of observed totals (starting 36 hours prior as in the rest of this study) to 24% (starting 12 hours 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 (Figure 4). The maximized 72-hour totals as a result decrease from 418 mm 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.

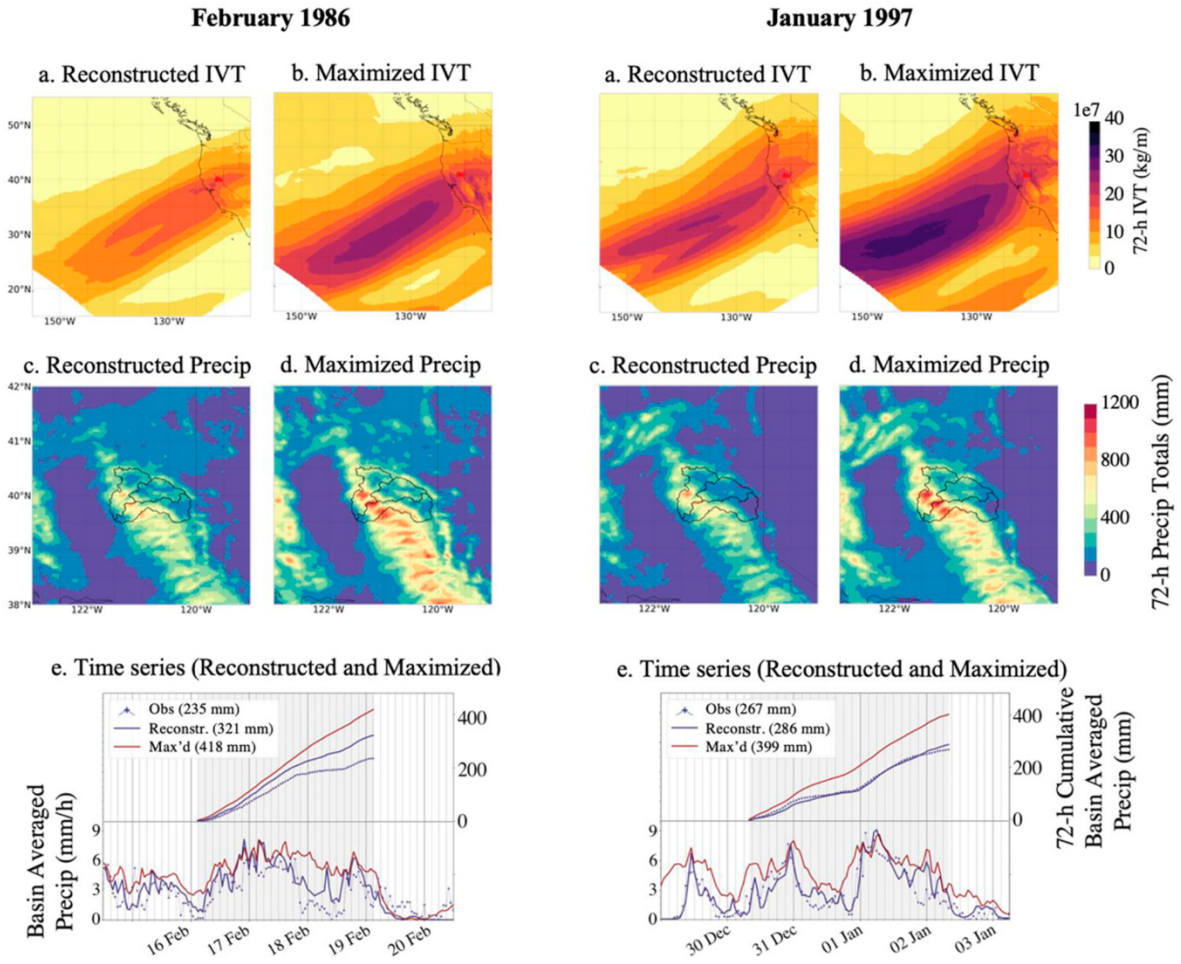


Chap. 2. Fig. 4. Precipitation time series for reconstructed and maximized runs precipitation (with 72-hour precipitation totals shown in the legend) started a: 36 hours prior (as in the rest of this study, same as Figure 3d, shown again here for comparison) and b: 12 hours prior to the onset of the 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.

3.b. Baseline storm maximization

Next, we performed the maximized simulations using the RHM technique of Ishida et al. (2015) 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 (Figure 5a and b, top row). The precipitation increases produce basin averaged maximized totals (418 mm 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 (Figure 5c, d and 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.



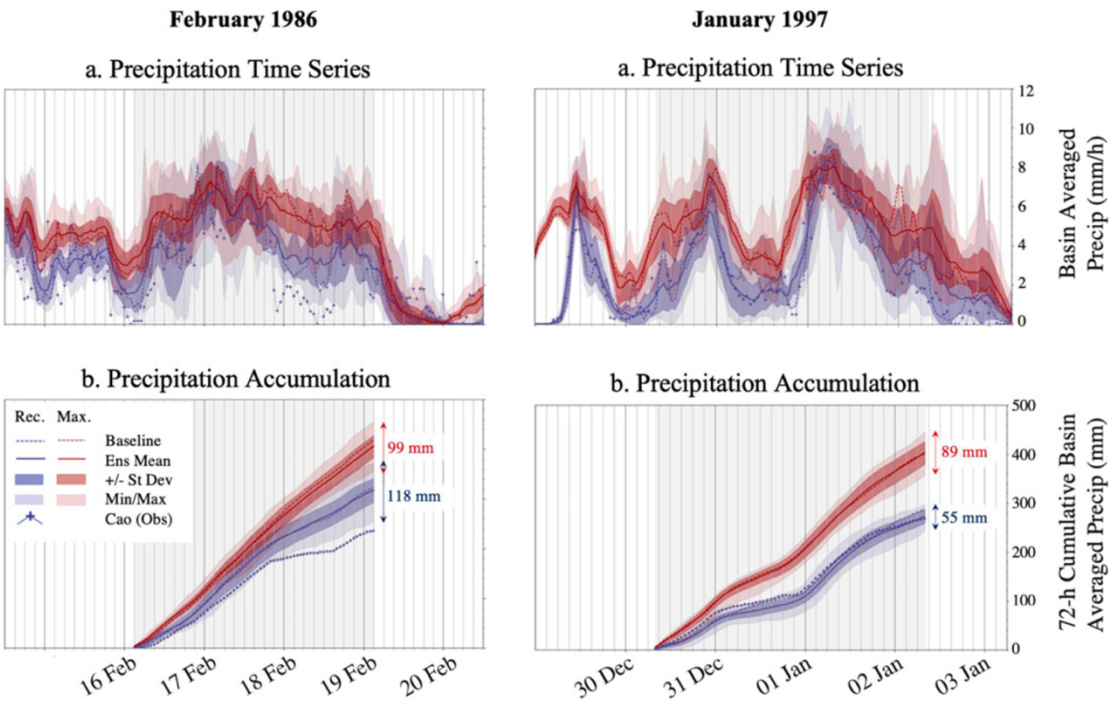
Chap. 2. Fig. 5. Maximized simulations for the February 1986 and January 1997 storms show increased precipitation totals compared to reconstructed simulations. a-b: Maps of reconstructed and maximized 72-hour total integrated vapor transport (IVT), respectively. c-d: Maps of reconstructed and maximized 72-hour total precipitation, respectively. e: Precipitation time series (observed, reconstructed, and maximized) with 72-hour precipitation totals shown in the legend.

3.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 Figure 6), keeping in mind the issues with the baseline reconstruction of February 1986 (Section 3a). We find that the February 1986 reconstructed ensemble encounters the same issues as the baseline run (dashed blue lines): the majority of the observations (blue dots) are missed (Figure 6). Even the lowest ensemble members overestimate precipitation during the trough on the first half of February 18th, 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 parametrization 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 encapsules 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.

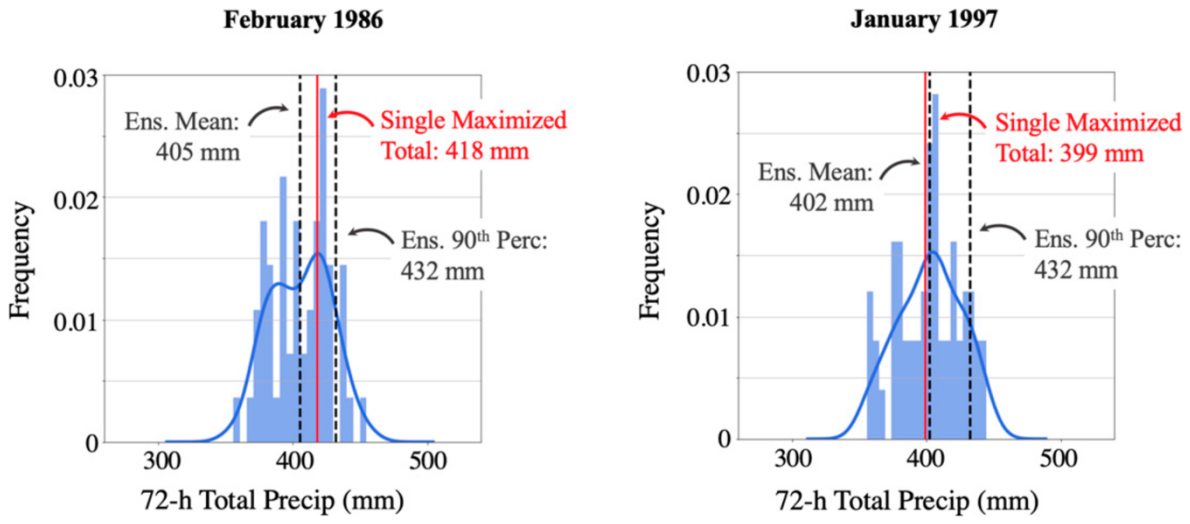
We next examine the two maximized ensembles (shaded red areas on Figure 6) and what they tell us about the robustness of the PMP estimates. The maximized ensembles are very similar between the two storms (unlike the reconstruction ensembles): there is almost the same spread among maximized runs for February 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 ensembles 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 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 February1986 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.



Chap. 2. 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-hour peak storm period. The arrows and numbers indicate the spread between the lowest and highest ensemble members.

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 Figure 7, we show the ensemble mean and 90th percentile on histograms of 72-hour precipitation resulting from maximization of the two storms, together with the single-value (West-WRF configuration, no perturbations) estimate. Our ensemble suggests 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.



Chap. 2. 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 ensembles means and 90th percentiles values in black.

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 spatial structure of observations well, but a wet bias exists that can be large (5% of maximized 72-hour 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 U.S. 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., 2014; 2015; 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 U.S. (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 towards 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.

Chapter 3: Did Historical Storms Used in Probable Maximum Precipitation (PMP) Estimation Reach Maximum Efficiency? A Large Model Ensemble Approach

A version of this chapter is currently under review at the *Journal of Geophysical Research*: Tarouilly E., Rahimi S., Cordeira J., Lettenmaier, D. P. "Did Historical Storms Used in Probable Maximum Precipitation (PMP) Estimation Reach Maximum Efficiency? A Large Model Ensemble Approach".

1. Introduction

There is a growing need to adapt flood safety guidelines and infrastructure as the climate changes. This need is particularly acute in California, where high population density and rainfall variability cause reliance on a large number of dams for water supply and flood control. Probable maximum precipitation (PMP), defined as the largest amount of precipitation that could occur over a given river basin, is a central aspect of flood safety and a key design criterion for dams and other high-risk infrastructure. PMP is an engineering concept (an approximation of the upper bound of accumulated precipitation), which in turn is used to estimate the resulting flood, (probable maximum flood or PMF). PMP and PMF were defined in response to the need for a single value that can be used to size dam spillways. The difficulty lies in that there are no known upper bounds on precipitation. Hydrologist Robert Horton stated that "it is not difficult to show that […] there is a natural limitation to rain

intensity" nearly a century ago (letter to F. Schmidt, 18 Nov. 1927), yet establishing the magnitude of a possible upper bound has been a challenge. This is due to the uncertainties regarding the combinations of different drivers that could modulate the upper bounds of precipitation, which PMP guidelines attempt to estimate.

Several issues have been identified with existing PMP estimation guidelines. Detailed procedures are outlined in NOAA's Hydrometeorological Reports (HMRs), such as HMR 59 for California (Corrigan et al., 1999). For instance, HMR guidelines recommend scaling the most severe historical storm's precipitation totals according to how much more moisture could have been available (via the Clausius-Clapeyron relationship) in order to obtain a PMP estimate. This moisture maximization approach, developed at a time when understanding of precipitation processes was limited, requires multiple assumptions. One is that severe historical storms achieved maximum storm efficiency (moisture conversion to precipitation), implying that moisture is the only storm characteristic to be maximized in determining PMP. Given the relatively short duration of the observed historical record at the time the original methods were developed, however, there is no certainty that even the most severe historical storms achieved maximum efficiency. Furthermore, additional assumptions regarding precipitation scaling with moisture are required, which have since been questioned (Zhao et al., 1997; Yang & Smith, 2018; Chen & Bradley, 2006; Ohara et al., 2017). As a result, there is a growing need to re-evaluate underlying assumptions in PMP estimation procedures.

Advances in numerical weather prediction modeling provide an opportunity to test assumptions and develop PMP estimation guidelines that better incorporate a current understanding of precipitation processes. The Weather Research and Forecasting (WRF) model (Skamarock & Klemp, 2008) is the current standard in most model-based PMP studies. As currently implemented, model-based PMP follows the logic of HMR guidelines

(Corrigan et al., 1999) in reconstructing and maximizing the moisture within a severe historical storm. Model-based PMP differs from HMR guidelines in that moisture is added at the model boundaries to generate more precipitation according to the WRF model formulation, rather than assuming that precipitation scales linearly with moisture availability. A number of studies in the last decade have contributed to a more consistent and defensible framework for model-based PMP estimation. These include the first modeling study to vary moisture in a dynamically consistent manner (Zhao et al., 1997), a recent body of work establishing that moisture flux rather than moisture is the key control on PMP (Smith & Baeck, 2015; Yang & Smith, 2018; Su & Smith, 2021; 2023), and studies that have attempted to define step-by-step guidelines for model-based PMP estimation (Ohara et al., 2011; Ishida et al., 2015 a & b; Toride et al., 2019; Ohara et al., 2017; Toride et al, 2019; Trinh et al., 2021). Taken together, the above studies suggest that model-based PMP provides a more physically-based framework than HMR methods in which to test assumptions and develop more accurate PMP estimates.

Model-based methods now accurately reflect the way precipitation is controlled by moisture, but not by storm efficiency, as storm efficiency is assumed to be already maximized in the most extreme historical storms. Given the relatively short duration of the observed historical record (typically less than 100 years), it is essentially certain that storms with higher efficiencies than those observed could occur (and likely have occurred). If this is the case, simultaneously maximizing (or optimizing) a combination of a large number of storm characteristics, including but not limited to moisture, is a challenge. Chen & Hossain (2018) highlight that atmospheric instability and large-scale convergence, in addition to moisture availability, are closely related to extreme precipitation totals. Recent papers by Smith & Baeck (2015) and Su & Smith (2021, 2023) similarly highlight that moisture flux and convergence, rather than moisture content, control extreme precipitation. However, there is

currently no clear picture of how specific variables control precipitation, what their upper bounds may be, or in what combination they maximize precipitation. Furthermore, the relationship between some variables (e.g., wind) and precipitation can be highly non-linear, making it difficult to determine how they should be modified so as to optimize precipitation. As a result, a "trial-and-error" approach to identify the optimal combination of multiple controls (moisture and efficiency) that would produce a PMP storm is not feasible, and there currently isn't a path for incorporating storm efficiency into PMP estimation.

To circumvent this issue, we propose here that a PMP storm – or a reasonably similar storm – can be found, in which case there is no need for amplification of either moisture or storm efficiency, so long as a large enough storm sample is available. Focusing on the Feather River basin in California, we present here a first attempt at this strategy using dynamically downscaled storms from the second version of the Community Earth System Model Large Ensemble (CESM2-LE; Danabasoglu et al., 2020; Rodgers et al., 2021) and the ECMWF's Fifth reanalysis (ERA5; Hersbach et al., 2020). The CESM2-LE consists of 10 separate downscalable GCM simulations of ~115 years in length (1900-2014), effectively pooling a historical climatology that is ~16.5 times the length of ERA5 (1950-2020). The larger sample size from the CESM2-LE allows for an approximately 1-in-1000-year storm to be examined within a PMP context to inform more robust PMP estimates. By comparing these storms to the most extreme storms in ERA5, we provide a first assessment of whether more severe storms exist in a large ensemble and whether they could produce (with or without moisture maximization) a larger PMP than current estimates.

The aim of this study is to better understand storm characteristics that have the greatest potential to influence PMP estimates. A key assumption of PMP methods (whether HMR or model-based), which we refer to as assumption 1, is that historical storms achieved maximum

efficiency. To address this issue, we search the CESM2-LE dataset for any storms with higher efficiencies than historical ones. Another assumption, which we refer to as assumption 2, is that storm efficiency does not change as a result of moisture maximization. We examine this assumption as well by comparing efficiency in reconstructed and amplified storms. Ultimately, our goal is to improve confidence in PMP estimates by using newly available modeling tools to ascertain whether or not the PMP could be larger than previously anticipated based on the current HMR approach and historical storms. While we do not expect to provide a final answer as to whether an upper bound on precipitation exists, we do expect that our work will result in improved understanding of the controls (storm efficiency in particular) on storms more severe than those examined in the context of PMP estimation. In particular, we address the following questions:

1. How can the upper bound of storm efficiency be approximated?

2. How close to maximum storm efficiency are the severe historical storms used in PMP estimation?

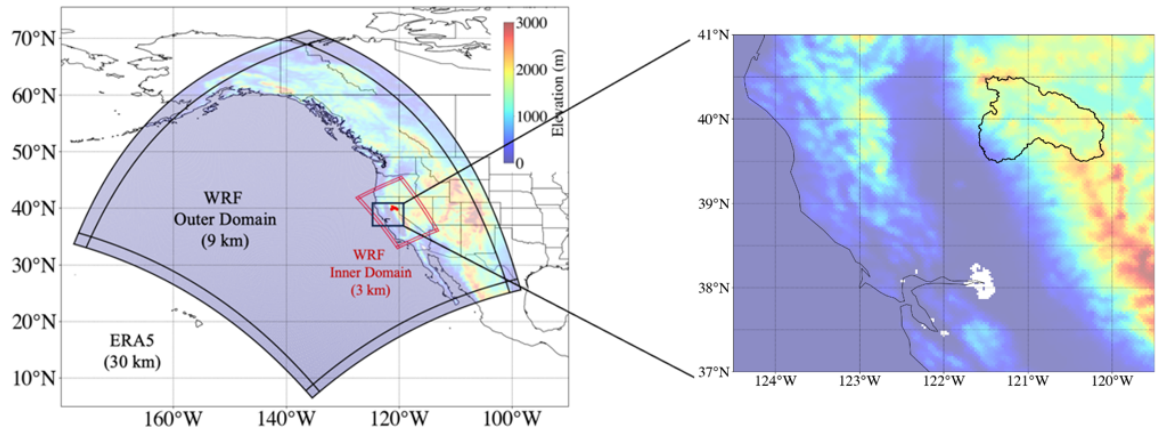
3. Does moisture maximization for PMP estimation affect storm efficiency?

2. Methods

2.a. Study area

Our study area is the Feather River watershed (~3600 mi^2, or 9300 km^2) upstream of Oroville Dam, California (Fig. 1). Our choice of this location is guided by the existence of earlier PMP estimation work in this region (e.g., Ishida et al., 2015a;b; Ohara et al., 2011) as well as work investigating local precipitation maxima in the Feather River watershed (Reeves et al., 2017). River basins located on the western slopes of the Sierra Nevada are an ideal

environment for investigating PMP-magnitude storms because the dominant precipitation mechanism is orographic forcing driven by onshore lower tropospheric moisture flux within landfalling atmospheric rivers (ARs), which is well represented by numerical weather models (Martin et al., 2018).



Chap. 3. Fig. 1. Location of the two nested 9 km (black) and 3 km (red) modeling domains and of the Feather River watershed.

2.b. Modeling approach

We perform WRF reconstruction and amplification experiments of severe storms selected based on the CESM2-LE's and ERA5's native resolution using 72-h basin average precipitation totals over the Feather River basin. We use the 72-h duration as it is typically the critical duration in large basins and has been used for this reason in earlier PMP studies for Oroville dam (Ishida et al., 2015a;b; Ohara et al., 2011). We select storms using precipitation (as opposed to another indicator of severe AR-related storms, e.g., integrated vapor transport or IVT), because precipitation is the variable we are ultimately interested in and, despite the coarse spatial resolution of the CESM2-LE, storms selected that way have equally large amounts of precipitation before and after WRF downscaling (not shown). Storms can be selected from any season, although almost all the storms we selected happen to be in fall or winter.

2.c. Forcing datasets

2.c.1. ERA5

ERA5 is generally acknowledged as the best performing reanalysis dataset for representing AR conditions due to its advanced data assimilation scheme and high (30-km) spatial resolution (Cobb et. al., 2021). As a reanalysis dataset, ERA5 provides both a reconstruction of the historical period for severe storm identification and a forcing dataset for high-resolution WRF simulations of said storms. We use both ERA5 (1979-2020) and the ERA5 back-extension (1950-1978).

2.c.2. CESM2-LE

The CESM2-LE is a 100-member ensemble of global climate model simulations spanning 1850-2100, 10 of which are archived with sufficient variables and time frequency to provide lateral and initial conditions for WRF simulations. CESM2 has contributed to the Coupled Model Intercomparison Project (CMIP) and has been identified as among the most skillful climate models in simulating western North American weather and climate (Knutti et al., 2013, Simpson et al., 2020, Krantz et al., 2021). CESM2 offers substantial improvements over its predecessor, CESM1 (Danabasolu et al., 2020). The different CESM2-LE ensemble members are produced by perturbations introduced at the beginning of each simulation. The CESM2-LE is designed to produce an ensemble of several realizations, with each realization characterized by a different phasing of internal climate variability.

Here, we use the historical (1900-2014) segments of the 10 ensembles that can be used to force WRF simulations. The CESM2-LE ensemble provides us, in addition to ERA5 reanalysis, with another ~1150 years of simulations containing different storms that we

consider equally plausible as historical ones. The use of the CESM2-LE dataset to identify extreme storm conditions that are relevant to PMP, but may not have occurred in the short historical record from ERA5, is an important contribution of this study. WRF ingests the CESM2-LE outputs, defined at the land surface, on 23 isobaric atmospheric levels, and in four soil layers, at different time intervals depending on data availability. To provide the CESM2-LE outputs to WRF, we created new software to replace the *ungrib.exe* program within WRF. Three-dimensional horizontal winds, temperature, and specific humidity, as well as surface pressure are updated at the lateral boundaries every six hours. Sea surface temperatures are assumed to be static fields in these short meteorological runs, although they are allowed to vary in preprocessing periods. WRF simulations are initialized to soil temperature and moisture fields directly from the CESM2-LE and evolve unconstrained following model initialization. No bias correction of the CESM2-LE outputs was performed prior to downscaling, the impacts of which have been shown to be important in modulating the statistics of dynamically downscaled precipitation (Bruyere et al., 2014; Holland et al., 2010). However, we believe the use of the CESM2-LE without bias correction is appropriate for the purpose of PMP estimation as the CESM2 model is known for its wet bias (Gershunov et al., 2019), and therefore provides a "worst-case" PMP estimate.

2.d. WRF model setup for model reconstructions

We follow the "West-WRF" model configuration (version 3.7.1) (Martin et al., 2018) developed and used by the Center for Western Weather and Water Extremes (CW3E) as its operational forecast model. We chose the West-WRF configuration as it is specifically tailored for representing extreme precipitation associated with ARs along the U.S. West Coast. We use two nested domains with resolutions of 9 and 3 km over the coastal Western U.S. (Fig. 1). The cumulus scheme is turned off in the inner domain. The wide area covered

by the outer domain provides coverage for the storm track regions and the Aleutian low, which both play an important role in steering ARs toward the U.S. West Coast. We evaluate the WRF reconstructions of historical storms through comparisons with the Cao et al. (2019) and Livneh et al. (2015) gridded precipitation datasets. The Cao hourly gridded precipitation (1/32 deg.) was produced using the Mountain Mapper method for 1985-2017 using hourly and daily data. The Livneh gridded precipitation (1/16 deg.) uses the SYMAP algorithm as implemented by Livneh et al. (2013) for the period 1950-2013 over the conterminous U.S.

2.e. Moisture maximization

The maximized storm simulations are produced using a widely-used technique called relative humidity maximization (RHM; Ishida et al., 2015). The logic of RHM is to amplify a severe historical storm's precipitation (as reconstructed by the WRF model) by providing it with additional moisture at the domain boundaries. This is accomplished by setting relative humidity (RH) to 100% at all grid cells in the forcing dataset which provide initial and boundary conditions to the WRF simulations, so that additional moisture flows into the model domain. The basin-averaged 72-h total precipitation produced by the moisture-maximized simulations is calculated for each storm, and the storm with the largest maximized 72-h total is retained as the PMP estimate. However, we emphasize that our goal here is not to obtain a PMP estimate but rather to examine how different storms that have the potential to yield the PMP react to the moisture maximization step of PMP methodology, therefore we refer to "amplified" rather than "PMP" storms throughout this paper.

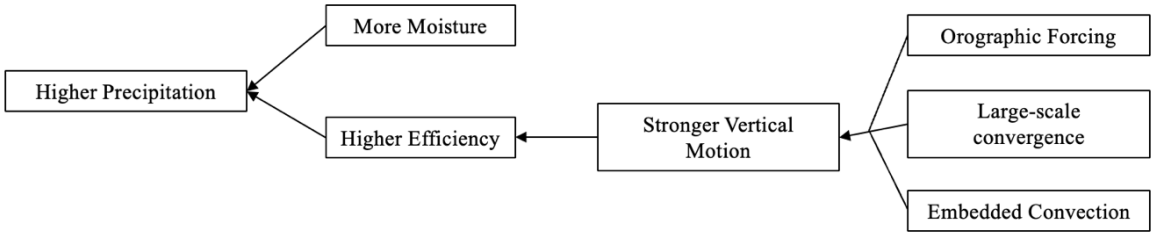
2.f. Analysis approach

We investigate differences in the behavior of storms before and after moisture maximization, as well as differences between ERA5 and CESM2-LE storms, the latter being selected from a

much larger sample. We investigate the differences in precipitation totals, which we explain from the perspective of storm efficiency (see Fig. 2 for an overview of the controls on precipitation, and Table 1 for definitions). The rationale for this approach is that, while moisture will be maximized as part of the moisture maximization procedure, storm efficiency is assumed to be near its maximum in the un-amplified storm (assumption 1). Whether or not this assumption is valid could have a large impact on the PMP estimate. Therefore, we assess whether storms with higher efficiencies than observed in the historical record can be found, and whether efficiency and its controls (horizontal wind speeds and in turn, vertical winds, and CAPE) respond to moisture maximization (assumption 2).

Chap. 3. Table 1. Definitions of variables used in the analysis of severe storms that have the potential to yield PMP estimates.

Variable	**Meaning**
IVT (kg/m/s)	$IVT = \frac{1}{g}\int_{p_{sfc}}^{100hPa} qVdp$ where q is the specific humidity (kg/kg) and V is the horizontal wind speed vector (m/s)
CAPE (J/kg)	Accumulated buoyant energy from the level of free convection (LFC) to the equilibrium level (EL) for the parcel with maximum equivalent potential temperature in the column
Precipitable water (mm)	$PW = \frac{1}{\rho g}\int_{p_{sfc}}^{100hPa} w\ dp$ where w is the mixing ratio (g/kg)
Storm efficiency	Precipitation (mm) / Precipitable water (mm)



Chap. 3. Fig. 2. Summary of the controls on precipitation: precipitation is driven by moisture and storm efficiency (moisture conversion to precipitation), which in turn is closely related to vertical motion (Mahoney et al., 2016).

3. Results

3.a. Storm precipitation

3.a.1. Precipitation of ERA5 storm reconstructions

We first examine the precipitation behavior of the 10 ERA5 storms using their WRF reconstructions. These storms are the 10 most severe by 72-h precipitation totals during the 1950-2020 ERA5 record. All are extreme storms: reconstructed basin averaged historical storm 72-h precipitation totals range from 170-286 mm (average 228 mm across the 10 ERA5 storms). Precipitation totals are shown both on Figs. 3a and 5a. All 10 storms occurred between October and March in association with landfalling ARs that either ranked as AR4 or AR5 according to the Ralph et al. (2019) AR Scale or were clusters (i.e., back-to-back storms, or AR families; Fish et al. 2021) of lower-category storms.

We evaluate the quality of the 10 historical storms' precipitation reconstructions relative to observations. Observations refer to either Livneh or Cao, or the average of the two datasets when both are available, as indicated on Fig. 3a. (dark blue line with dots). The reconstructions tend to produce slightly larger (between 67% and 145%) 72-h totals compared to observations. Given observational uncertainty in the gridded data sets

(evidenced by differences between the Cao and Livneh data in cases when both are available, which model uncertainty is in some cases approaching, see Lundquist et al., 2019) the reconstructed and observed storm precipitation and spatial patterns may be considered similar. As Tarouilly et al. (2022) previously noted, the temporal pattern of precipitation is typically well reproduced by reconstructions (Fig. 3a, light blue lines), except occasionally for the timing of some precipitation peaks. The spatial pattern of precipitation is closely reproduced, following orographically favored regions of the Feather River watershed. We highlight that the least well reconstructed storms are typically the earlier storms for which less satellite data was assimilated in the ERA5 reanalysis (forcing dataset). The appropriateness of the reconstructions is further discussed in Section 3.a.3. in the context of the magnitude of precipitation totals after moisture maximization.

3.a.2. Precipitation of the CESM2-LE storm reconstructions

We next examine the 10 most intense storms downscaled from the CESM2-LE (selected from a much larger sample than the ERA5 storms). The 10 CESM2-LE storms occurred during a more restricted part of the year than the ERA5 storms: all occurred either in December or January, except for one in November. Precipitation totals are shown both in Figs. 3b and 5b. Precipitation totals for all but two CESM2-LE storms are greater than any of the ERA5 storms: basin averaged 72-h reconstructed CESM2-LE storm totals range between 236-370 mm, with an average of 325 mm across the 10 storms. The 10 CESM2-LE storm totals are on average 43% larger than the 10 ERA5 storms. The largest CESM2-LE storm total (370 mm) is 29% greater than the largest ERA5 storm (286 mm). The difference between the averages of the 10 ERA5 and 10 CESM2-LE storms is more pronounced than between their most extreme storms.

The spatial patterns of CESM2-LE storms before amplification are similar to those of ERA5 storms (Fig. 4a and 4c), although precipitation is stronger in already high precipitation orographic areas. Regarding temporal patterns, while most CESM2-LE storms have broader periods with relatively high precipitation, the magnitude of peaks generally is not increased much compared to ERA5 storms. Similarly, Rahimi et al., (2023) found that the transmission of mean-state biases from GCMs to WRF were larger when considering 5-day events compared to 1-day (i.e., peak) precipitation totals. The small increases in precipitation (both 72-h totals and peaks) despite the known wet biases in the CESM2-LE as well as the larger sample size suggests that precipitation may be approaching an upper limit.

3.a.3. Precipitation of ERA5 and CESM2-LE storms amplified by RHM

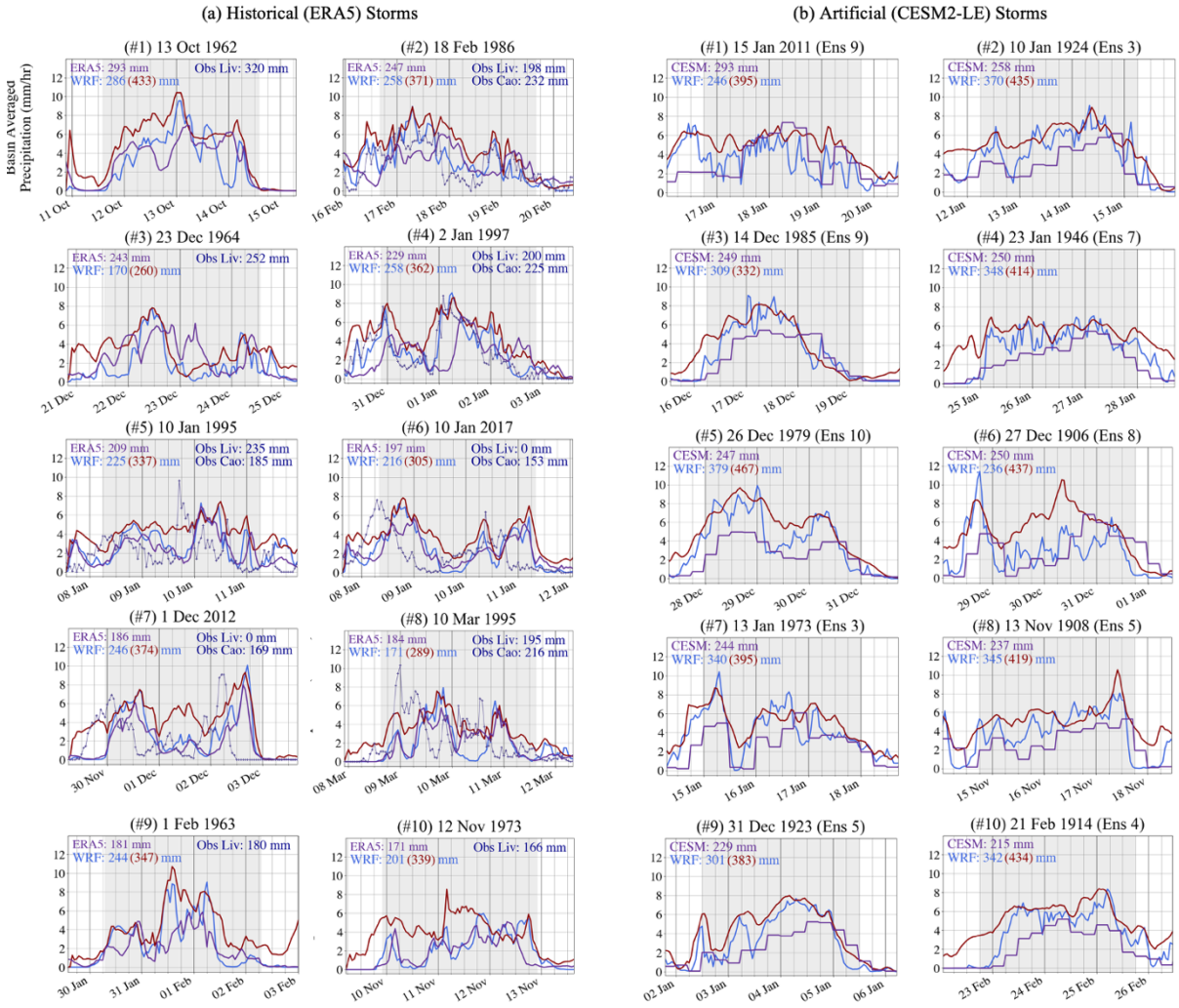
Moisture maximization of ERA5 storms results in increases in precipitation totals of +40-69%, with an average increase of +51% (Fig. 5a). The largest amplified 72-h total precipitation produced is 433 mm. For CESM2-LE storms on the other hand, moisture maximization tends to produce smaller average increases in precipitation totals of +31% and a wider range of 7-85% increases (Fig. 5b). We investigate reasons for those differences in Section 3.b. Both ERA5 and CESM2-LE precipitation is enhanced in orographic areas (Fig. 4b and 4d). As a result of the relatively smaller precipitation increases and larger reconstructed totals before moisture maximization, the largest amplified CESM2-LE 72-h storm total precipitation (467 mm) is only 7% larger than the largest amplified ERA5 storm (433 mm) (Fig. 5). These results suggests that the PMP estimated from amplified CESM2-LE storms would likely only be slightly larger than if obtained from ERA5 storms (or historical observations, as is currently the case for HMR estimates).

We highlight that the changes in precipitation, whether between reconstructed ERA5 and CESM2-LE storms (due to sample size), or between a given reconstructed and amplified storm (due to moisture maximization), are similar. The magnitudes of the increases are similar: reconstructed CESM2-LE storms are on average 43% larger than reconstructed ERA5 storms, while moisture maximization causes 45% increases in ERA5 storm precipitation and 31% increases in CESM2-LE storm precipitation (Fig. 5). In addition, amplified ERA5 storms and un-amplified CESM2-LE storms are similar: both display broadened periods of high precipitation without changes in peak magnitudes and intensified precipitation in orographic areas compared to un-amplified ERA5 storms. This is an important result that suggests the PMP storm can be approximated using either moisture maximization or a large sample size and is therefore quite robust. The value of using artificial storms lies in the fact that they can be combined with moisture maximization to be further amplified (even if slightly).

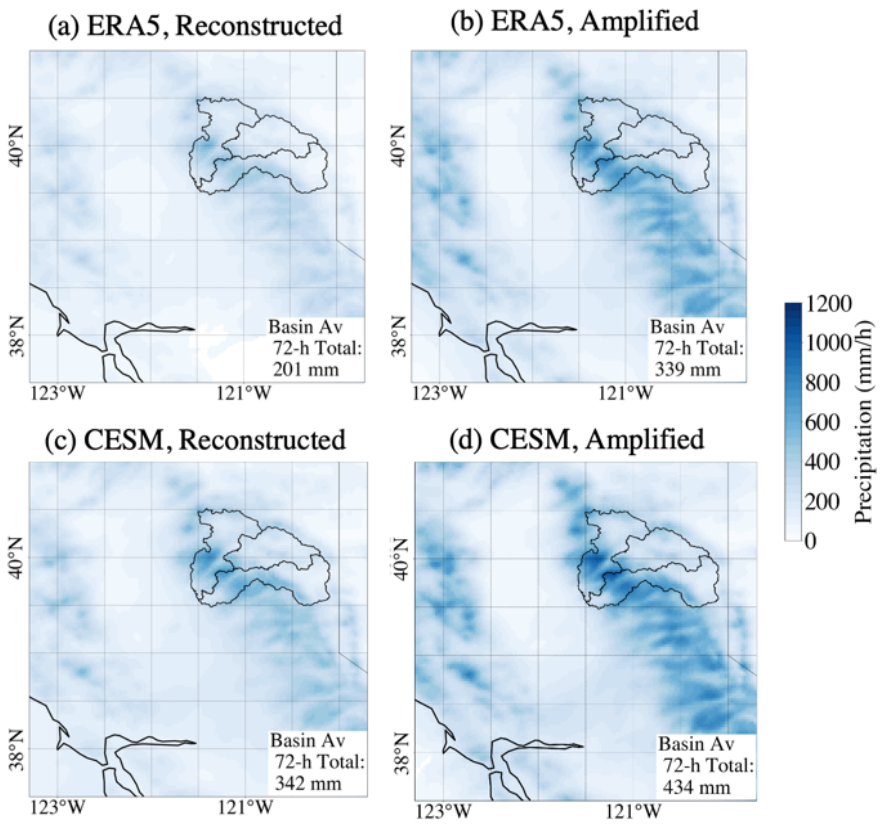
We also note that moisture maximization fills in precipitation gaps, such that pre-amplification differences in spatial or temporal pattern across storms (due either to the choice of storm or the quality of reconstructions) tend to be reduced (Fig. 3, red lines). We observe less storm-to-storm variation among amplified than reconstructed storms. This result also points to a degree of robustness in model-based PMP estimation, which produces similar values even if obtained from storms that are imperfectly simulated.

We hypothesize that a sufficiently large sample size would therefore contain a storm requiring very little amplification of moisture or any other aspects of the storm, if any (i.e., it would contain the PMP storm, or a storm of near-PMP magnitude). This would make it easier to apply the same approach to any watershed in the country as opposed to using amplification methods that require location-specific decisions. Questions remain as to what size this sample

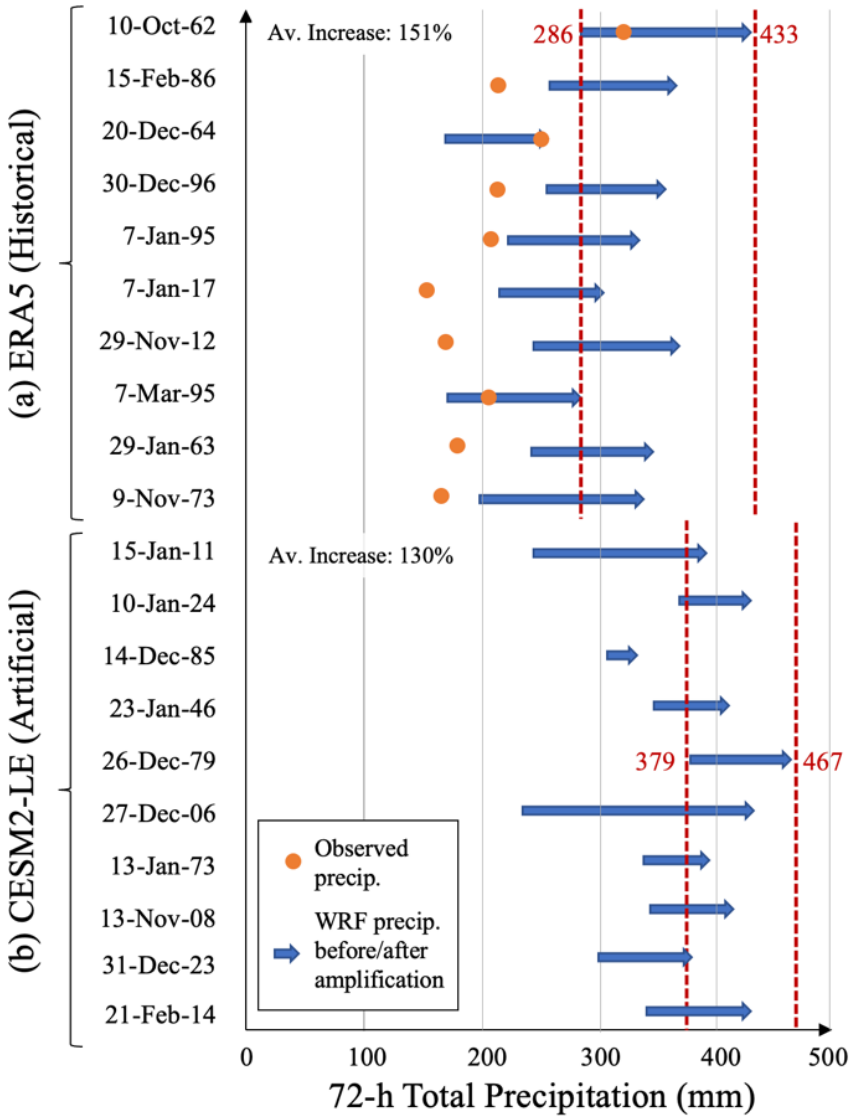
should be to ensure finding the PMP storm. The following sections investigate differences in storm efficiency to explain why (1) CESM2-LE storms have larger precipitation totals than ERA5 storms before amplification and (2) moisture maximization produces smaller precipitation increases in the CESM2-LE compared to ERA5 storms.



Chap. 3. Fig. 3. Basin-averaged precipitation time series for (a) the 10 ERA5 and (b) the 10 CESM2-LE storms, including precipitation in the forcing datasets (ERA5 and CESM2-LE, respectively) in purple and WRF-reconstructed (light blue) and amplified (red) precipitation. For historical storms, observed precipitation from Cao et al. (2019) is additionally shown in dark blue (dots); Livneh et al. (2015) precipitation is not plotted as it is daily. Basin-averaged storm total precipitation values are displayed at the top of each plot. Grey shading represents the 72-h storm duration. WRF-reconstructed precipitation totals of historical storms are similar to observations.



Chap. 3. Fig. 4. Composite 72-h total precipitation maps, i.e., precipitation averaged across each of the 10 reconstructed ERA5, amplified ERA5, reconstructed CESM2-LE and amplified CESM2-LE storms. Spatial patterns of precipitation follow orographically favored regions and are therefore similar in ERA5 vs. CESM2-LE storms and in reconstructed vs. amplified storms.



Chap. 3. Fig. 5. Magnitude of the amplification by moisture maximization of ERA5 and the CESM2-LE 72-h precipitation totals. The beginning of each arrow indicates the reconstructed total, while the end of the arrow indicates the amplified total. Orange dots show observed totals for historical storms. Vertical red dotted lines indicate the largest ERA5 (CESM2-LE, respectively) totals before and after amplification. CESM2-LE storms produce slightly larger precipitation totals than ERA5 storms before amplification, but do not get amplified as much by moisture maximization.

3.b. Efficiency of reconstructed and amplified ERA5 and CESM2-LE storms

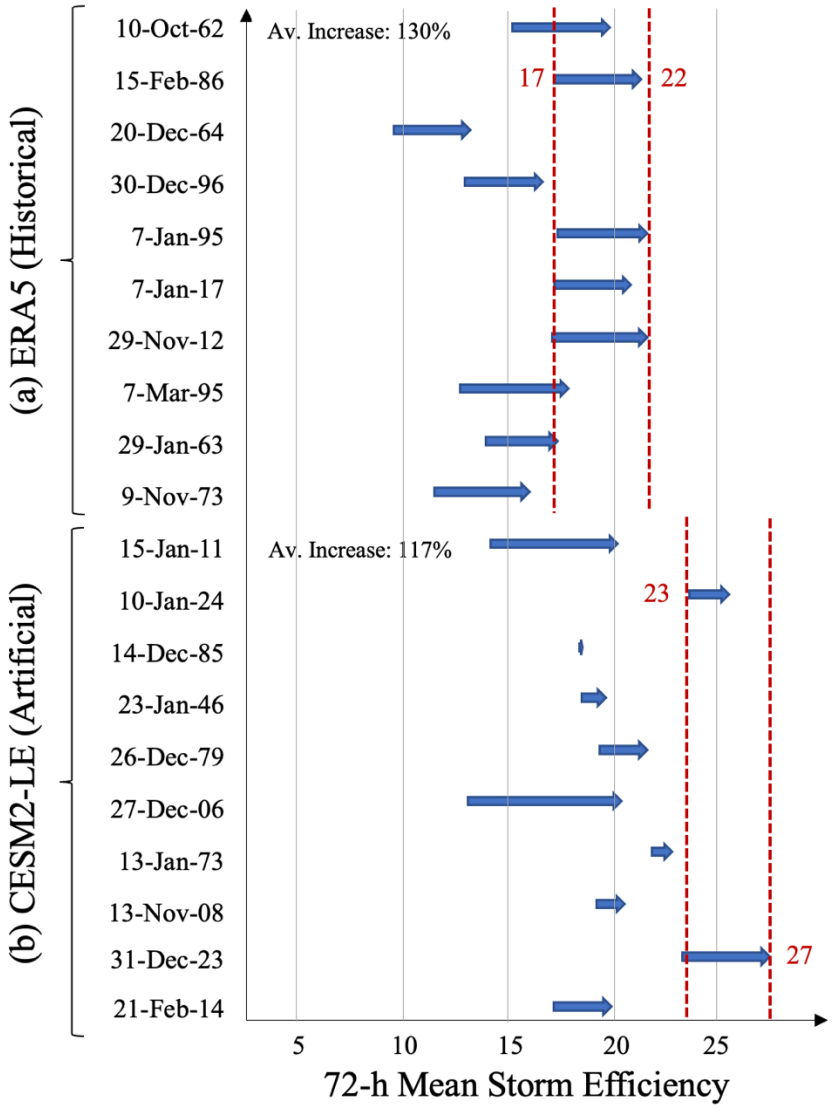
3.b.1. Efficiency of ERA5 vs. CESM2-LE storms (before amplification)

This section investigates the larger precipitation totals noted, before amplification, in CESM2-LE storms compared to ERA5 storms. While these differences can be caused either by differences in moisture amounts or storm efficiency (transforming moisture into precipitation), we focus on efficiency as the assumption (assumption 1) that storm efficiency is already maximized in historical storms is central to PMP methodology (both HMR and model-based). We therefore investigate whether higher-efficiency storms can be found in the CESM2-LE (given its large sample size). We find that storm efficiencies in most CESM2-LE storms are similar those in ERA5 storms, except for three that are higher (Fig. 6). Therefore, based on the CESM2-LE dataset, HMR assumption 1 may not be valid. The remainder of this section investigates why the CESM2-LE storms tend to have high efficiencies, while the following section (Section 3.b.2) evaluates implications of higher efficiencies for the magnitude of the PMP estimate.

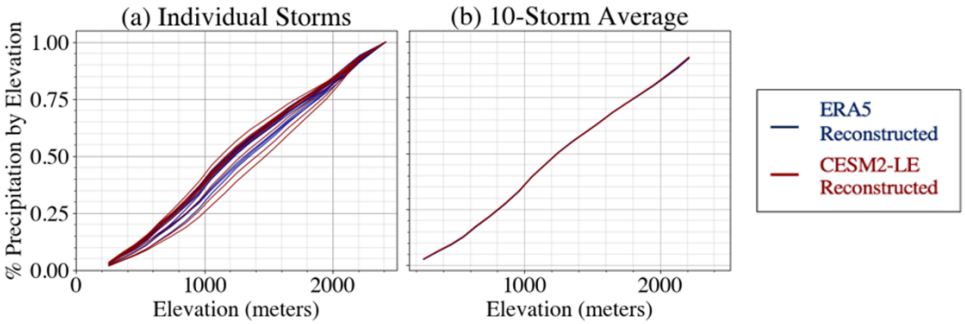
We observe little difference on average between ERA5 and CESM2-LE storms, which all exhibit weak orographic gradients (similar proportion of precipitation across elevations indicating weak dependence of uplift on topography), despite some event-to-event variability (Fig. 7). This is consistent with the fact that all storms are landfalling ARs with southwesterly water vapor flux and IVT, and all are similarly centered over the Feather River basin. We next examined equivalent potential temperature profiles to assess whether instability and convection could explain the differences in efficiency between ERA5 and CESM2-LE storms. We found that CESM2-LE storms on average have less-stable profiles than ERA5

ones (Fig. 8). Nevertheless, it does not appear that these differences are large enough for convection to explain the differences we observe.

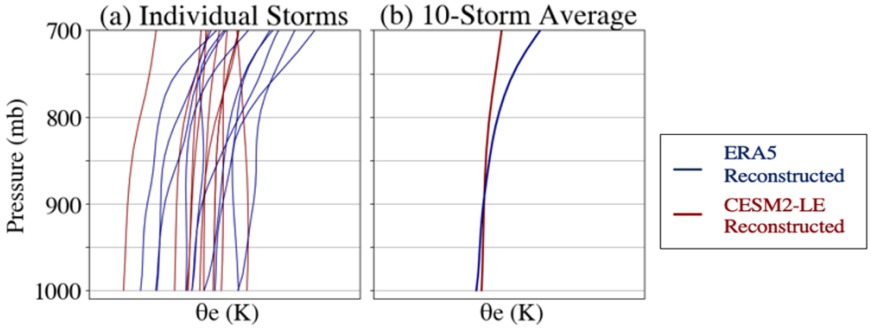
Finally, we investigated whether convergence could be driving the differences in efficiency. CESM2-LE storms on average have a deeper pressure trough (980 mb vs. 990 mb), stronger horizontal wind towards the basin, and stronger IVT than ERA5 storms (Fig. 9). These larger (i.e., mesoscale and synoptic-scale) processes associated with CESM2-LE storms may play an important role in producing higher storm efficiencies. This difference is important as storms with different characteristics may respond to moisture maximization more than others, therefore having a number of higher-efficiency storms to amplify may produce a large PMP despite none of them exceeding historical efficiency. The next section evaluates how storm efficiency responds to moisture maximization in ERA5 versus CESM2-LE storms.



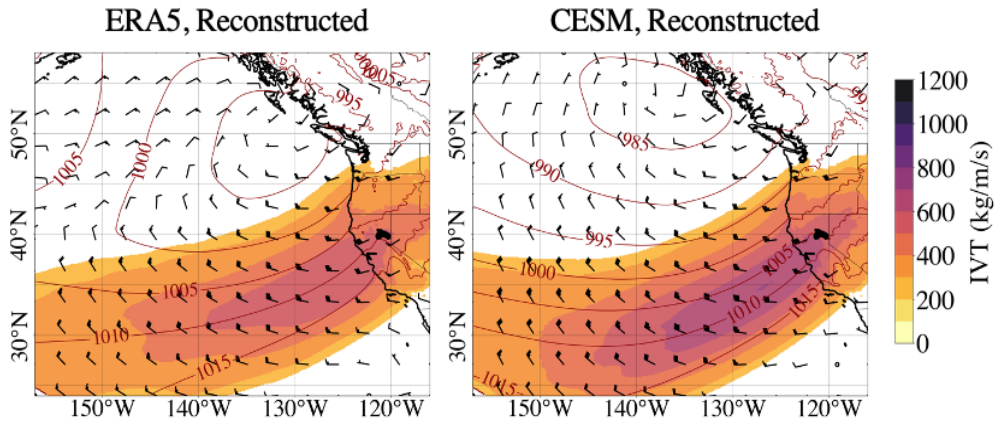
Chap. 3. Fig. 6. Magnitude of the changes in storm efficiency as a result of storm amplification by moisture maximization of ERA5 and the CESM2-LE storms. The beginning of each arrow indicates the reconstructed efficiency, while the end of the arrow indicates the efficiency after amplification. Vertical red dotted lines indicate the largest ERA5 (CESM2-LE, respectively) efficiencies before and after storm amplification. Storm efficiency tends to increase during moisture maximization, especially for ERA5 storms. CESM2-LE storms have slightly higher efficiencies than ERA5 storms both before and after amplification.



Chap. 3. Fig. 7. Cumulative precipitation percentage by elevation showing, (a) shows each storm individually, while (b) shows the average of the 10 ERA5 and 10 CESM2-LE storms, respectively. While there is storm-to-storm variability, ERA5 (blue) and CESM2-LE (red) storms do not, on average, have different orographic forcing.



Chap. 3. Fig. 8. Equivalent potential temperature profiles upstream of the study basin at locations within the box shown on Fig. 10b where IVT>250 kg/m/s. (a) shows each storm individually, while (b) shows the average of the 10 ERA5 and 10 CESM2-LE storms, respectively. ERA5 (blue) and CESM2-LE (red) storms on average have only slightly different stability profiles.



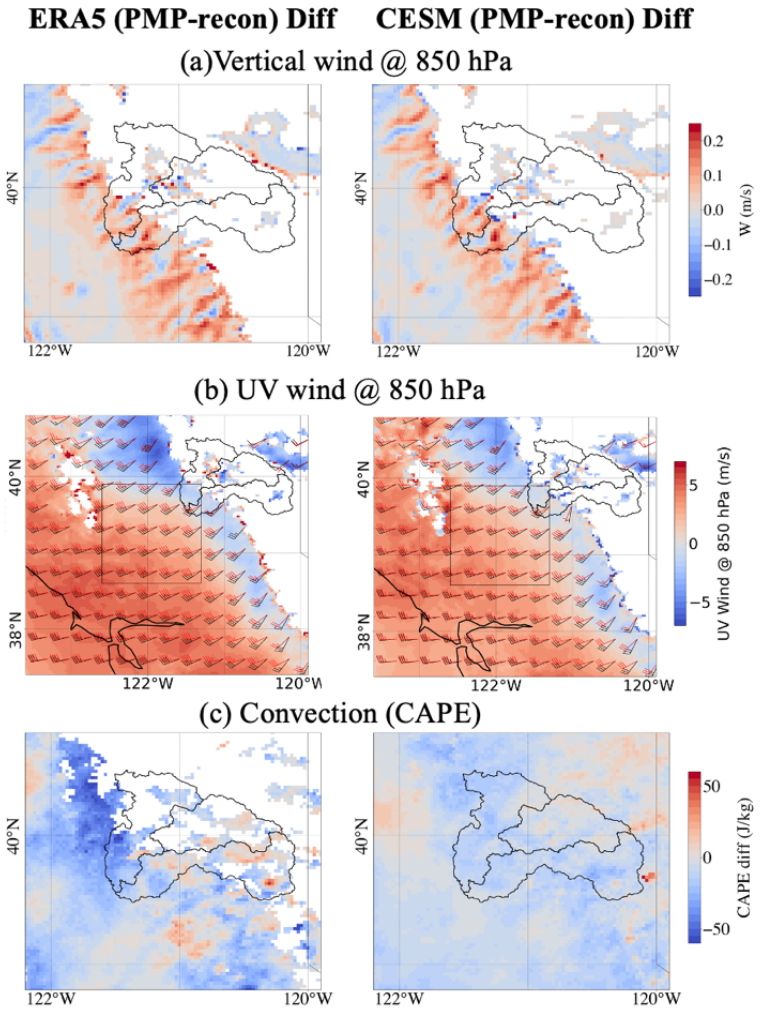
Chap. 3. Fig. 9. Maps of IVT (kg/m/s; shaded), sea level pressure (hPa; contours) and 850-hPa winds (m/s; vector) averaged over each 72-h storm period, then averaged over the 10 ERA5 storms (left) and the 10 CESM2-LE storms (right). CESM2-LE storms on average have stronger convergence and deeper pressure troughs than ERA5 storms.

3.b.2. Efficiency of reconstructed vs. amplified storms

We note a number of changes to moisture and storm efficiency during the moisture maximization of both ERA5 and CESM2-LE storms. Before maximization, the CESM2-LE storms contain on average more moisture and are closer to saturation (not shown), such that the relative increases in precipitation are not as large as in ERA5 storms. Unexpectedly however, we find sometimes large increases in precipitation efficiency upon moisture maximization in both ERA5 and CESM2-LE storms – with up to 50% increases in some ERA5 storms (Fig. 6). The increase in efficiency appears to be influenced by the mesoscale and synoptic-scale characteristics of the landfalling ARs responsible for sustained non-moisture factors responsible for precipitation (i.e., convergence) which are modified upon moisture maximization. Note that PMP guidelines (assumption 2) assume these mesoscale and synoptic-scale characteristics are not modified.

Additional meteorological differences among the reconstructed and amplified storms explain the changes in storm efficiency during moisture maximization (Fig. 10). These include widespread increases in 850mb vertical velocities along the Central Valley and Sierra Foothills, suggesting overall more kinematic convergence, as well as increases in horizontal wind speeds. The increases in both vertical and horizontal velocities after moisture maximization appear stronger for ERA5 than for CESM2-LE storms. The location of this implied quasi-zonal convergence, which is enhanced by moisture maximization, coincides with where the Sierra Barrier Jet – a south-southeast to north-northwest terrain-driven lower-tropospheric jet stream – which forms in response to strong landfalling ARs (Neiman et al., 2013; Kingsmill et al., 2013). This further suggests that the selected CESM2-LE storms likely already contained quasi-optimal mesoscale and synoptic-scale processes conducive to heavy precipitation during landfalling ARs that were sub-optimal in ERA5 storms before

amplification. Those processes are enhanced as a side-effect of moisture maximization in both sets of storms, but more so in ERA5 storms. We additionally observe an overall reduction in CAPE with moisture maximization in both sets of storms, implying that increases in precipitation were likely not convective. These results affirm the utility of using modeling as a tool for estimating PMP and assessing the direct and indirect impacts of both storm selection and moisture maximization on the meteorological processes responsible for extreme precipitation in landfalling ARs.



Chap. 3. Fig. 10. Difference between amplified and reconstructed (a) vertical wind velocities, (b) horizontal wind velocities and (c) CAPE, for the 10 ERA5 storms (left) and 10 CESM2-LE storms (right). The differences are calculated as amplified minus reconstructed values for each storm, then the average of the differences is calculated across each set of storms. The large increases in horizontal, and in turn, vertical wind velocities appear to be an important contributor to increases in storm efficiency during moisture maximization.

4. Conclusions

This study is a first step towards addressing a key question that limits our confidence in PMP estimates, both HMR and model-based: how do we know we correctly designed a hypothetical event if it has not occurred? We leveraged the large sample size of the CESM2-LE (~1150 years) to produce PMP estimates that are arguably more robust than those based on the observed historical record (~70 years). We additionally shift the focus from moisture (which has a known upper bound) to storm efficiency, and leverage the CESM2-LE to understand the controls on storm efficiency and the implications of possible higher storms efficiencies for PMP.

We highlight the following key findings and conclusions they lead to:

- Unamplified CESM2-LE storms have slightly higher efficiency, are more saturated with moisture and produce slightly larger precipitation totals than unamplified ERA5 storms, but do not exceed current PMP estimates. It does appear based on the CESM2-LE that higher storm efficiencies that what has been observed are possible, such that HMR assumption 1 may not be valid.
- Due to near-saturation and despite slight increases in storm efficiency, moisture maximization amplifies CESM2-LE storms only marginally, leading to only an 8% increase in the PMP estimate obtained from the CESM2-LE instead of historical storms. Part of that increase is due to an unintended increase in storm efficiency during moisture maximization, contrary to HMR assumption 2 that precipitation increases linearly with moisture availability. The ability of model-based PMP to account for changes in storm efficiency represents a major advance.

While the insights gained in this study motivate further use of ensemble datasets such as the CESM2-LE to better understand the controls on storm efficiency and "engineer" possibly more severe (even if marginally) PMP storms, our results also suggest that no storm modifications may be necessary if a large enough sample (the size of which remains to be determined) of storms was available. We suggest further testing whether this may be the case using other GCM ensembles, in addition to the CESM2-LE. No longer relying on the subjective and location-specific considerations of topography and moisture sources that are inherent to storm amplification would be a major step forward for the development of PMP guidelines across the U.S.

Robust model-based PMP methodology is important as the basis for producing model-based PMF both in the current and future climate. The WRF simulations that produced the PMP estimates can be used directly to force distributed hydrologic models, and other simulations could be designed to additionally simulate snowmelt and rain-on-snow events. The coupled PMP and PMF simulations would also lend themselves to investigating potential climate impacts, whether using a pseudo-global warming approach or by selecting storms from climate projections (hence the value of the work presented here using the same GCMs for the historical period that could also supply the future projections). Our results so far suggest that precipitation amounts in the current climate are likely bounded, and the above-mentioned coupled atmospheric-hydrologic simulations would confirm whether this translates into floods also being bounded, and whether this upper bound could change in a warmer climate.

Chapter 4: Physically-Constrained Model-Based Moisture Amplfiication for Probable Maximum Precipitation (PMP) Estimation

A version of this chapter is currently under review at the *Journal of Hydrometeorology*: Tarouilly E., Holman, K., Lettenmaier, D. P. "Physically-Constrained Model-Based Moisture Amplification for Probable Maximum Precipitation (PMP) Estimation".

1. Introduction

A warming climate, as well as population growth, are leading to growing infrastructure vulnerabilities, and hence the need to improve flood preparedness and the reliability of water supplies. Dams are a key aspect of societal resilience, provided that they can safely pass the most severe floods. The safety of large dams currently relies on a concept called probable maximum precipitation (PMP). PMP is defined as the largest amount of precipitation that could occur at a given location (WMO, 2009). PMP is in turn used to estimate the resulting flood (probable maximum flood or PMF) and determining requirements for safely passing the associated reservoir releases.

Existing PMP estimation procedures are outlined in NOAA's Hydrometeorological Reports (HMRs), such as HMR 57 for the U.S. Pacific Northwest (Hansen et al., 1994). In HMR guidelines, the PMP estimate is obtained by transposing the most severe historical storm's precipitation over a target river basin and scaling precipitation totals according to how much

moisture could have been available during the event (typically a historical maximum). This approach was developed at a time when understanding of precipitation processes and available data were limited. In particular, the method requires assumptions of a pseudoadiabatic lapse rate such that precipitable water can be estimated from surface dew points. It further requires an assumption that precipitation scales linearly according to moisture availability (Abbs, 1999). These assumptions have since been questioned (Zhao et al., 1997; Yang & Smith, 2018; Chen & Bradley, 2006; Kim et al., 2020), motivating the re-evaluation and updating of PMP estimation procedures.

Advances in numerical weather prediction modeling provide an opportunity to develop PMP estimation procedures that better reflect the complex physical processes that control precipitation than the approximations used in current HMR methods. The Weather Research and Forecasting (WRF) model (Skamarock & Klemp, 2008) is the current standard in model-based PMP studies (e.g., Kunkel et al., 2013 and Gangrade et al., 2018, in addition to the studies cited below). As currently implemented, model-based PMP has followed the logic of HMR guidelines in transposing and amplifying the moisture of severe historical storms. However, model-based PMP differs from HMR guidelines in that the model formulation is used to predict precipitation that results from moving a severe storm's moisture to a new location and further adding moisture at the model boundaries (rather than directly scaling precipitation). Using a model that reflects current state-of-the-art understanding of precipitation processes to do this is a key improvement over existing HMR techniques.

Despite the major improvement in the theoretical basis for model-based PMP approaches, these methods have so far not been used to produce operational estimates. One barrier is that there currently exist multiple different implementations of moisture amplification (summarized in Table 1), which involve adding different amounts of moisture at different

locations and which produce different amplified precipitation estimates. Some authors increase moisture all the way up to saturation (relative humidity maximization, or RHM; Ishida et al., 2015). Others use a multiplier to increase moisture according to the amounts of moisture already present (relative humidity perturbation, or RHP; Toride et al., 2019). They suggest RHP more realistically preserves the storm characteristics than RHM does, but do not provide guidance as to how to select the multiplier (Toride et al., 2019). For both RHM and RHP, the moisture increases can be performed either everywhere in the forcing dataset, or only in the path of the storm (called RHM-IVT or RHP-IVT where IVT refers to integrated vapor transport in kg/m/s as the method uses an IVT threshold to identify the path of the storm; Toride et al., 2019). The reason for these different possible approaches is that while the model (e.g., WRF) provides a quantitative and reproducible way of converting moisture to precipitation, how moisture is added remains a somewhat subjective decision. Saturation is arguably a physical upper bound but may not be realistic. In some cases, more moisture does not translate into more precipitation (Ohara et al., 2017). Here, we propose an approach which we call RHP-ratio, that builds on RHP but uses historical maximum moisture as a way to consistently select the magnitude of the moisture increases.

Another barrier to the implementation of model-based PMP approaches is that they have so far been developed for individual river basins, when in fact regional PMP estimates (e.g., statewide PMP or PMP for the same homogeneous regions as those covered by the HMRs) are needed. Producing PMP estimates (i.e., simulating multiple storms and testing different transposition and moisture amounts) for a large number of river basins one-by-one would require significant computational resources. It may be possible to optimally (or nearly optimally, given other uncertainties) center and amplify storms over multiple river basins within a homogeneous region, greatly reducing the number of simulations required to

produce regional PMP estimates. Another goal of this paper to evaluate whether this is possible.

This study aims to (1) characterize how sensitive PMP estimates are to decisions related to moisture and transposition amounts, (2) identify physically-based reasoning that can be used to define more defensible and applicable model-based implementations, and (3) demonstrate the proposed moisture amplification framework for the Willamette River system as well as smaller but high-hazard dams in coastal Oregon. We emphasize that our goal is not to produce a PMP estimate but rather to characterize and address important sources of uncertainty. In particular, we ask the following questions:

1. Can a storm be positioned over multiple river basins within a homogeneous region, i.e., does optimizing storm position for one river basin significantly reduce maximized precipitation totals for another?
2. How should atmospheric moisture be modified to produce worst-case precipitation totals while retaining realistic moisture perturbations? In particular:
    a. Should moisture be maximized (i.e., saturated) or only "amplified" using a multiplier? If so, how?
    b. Should moisture addition be limited to the path of the storm or applied to the entire domain?
    c. At what distance from the target river basins should the moisture perturbations be applied?

2. Methods

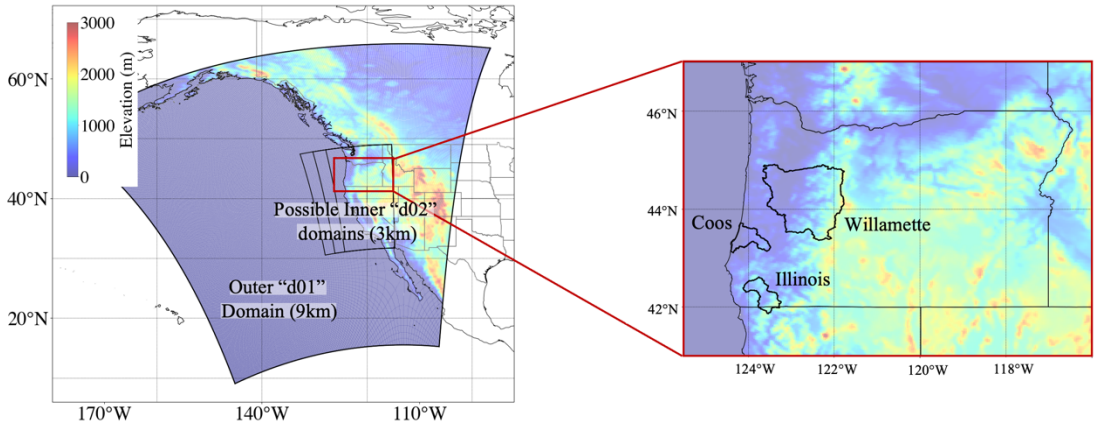
2.a. Study domain

We focus on mountainous coastal drainages in Oregon (U.S. Pacific Northwest). These river basins provide ideal cases for developing model-based PMP estimates because the atmospheric river (AR) storms and orographic effects that control precipitation in those areas are well represented by numerical models (Tarouilly et al, 2022). This study focuses on three river basins: the Coos including Pony Creek Dam (610 mi^2 or 1580 km^2), the Illinois including McMullen Dam (990 mi^2 or 2564 km^2) and the Willamette upstream of Salem (48,692 mi^2 or 126,112 km^2) (Fig. 1). These river basins were chosen for diversity in size, because they have dams with spillway sizing concerns, and in the case of the Willamette because of the existence of a previous model-based PMP study (Toride et al., 2019).

2.b. Model set-up

We used the Weather Research and Forecast (WRF) model (Skamarock & Klemp, 2008) which we parametrize using a configuration called "West-WRF" that was developed specifically for modeling weather extremes including AR storms in the western U.S. (Martin et al., 2018). We forced the WRF model runs every hour with ERA5 reanalysis (Hersbach et al., 2020) at 30 km spatial resolution. The model set-up uses two nested domains (outer domain "d01" and inner domain "d02" with 9 and 3 km spatial resolution, respectively) centered over the state of Oregon (Fig. 1). Sensitivity tests described in Section 3.e. additionally involve varying the location of the inner domain's western boundary, for which the three different possible inner domains are also shown on Fig. 1. Our simulations cover periods of roughly six days which allows to capture three days of storm conditions from

which we calculated maximized precipitation totals as well as at least one day beforehand for spin-up.



Chap. 4. Fig. 1. Study region showing nested 9 km (outer, "d01") and 3 km (inner, "d02") WRF modeling domains and three study river basins in Oregon. For the inner domain, three possible domains are shown, the implications of which are examined in Section 3.e.

2.c. Storms

We examine two extreme AR events: the storms of December 1964 and November 1996. The tropical-origin storm of December 1964 caused the greatest recorded streamflows (greater than the storm of 1955) in Oregon and other Western states; it was caused by warm unstable air masses and strong winds oriented orthogonally to Oregon's coastal mountains (Waananen et al., 1971). The storm of November 1996 was similarly caused by moist subtropical air masses which were steered toward Oregon by colder air masses further to the North (Risley, 2004). We chose those storms based on their large precipitation totals (within the largest 30 in ERA5 reanalysis) over all three river basins for a wide range of durations (1-, 6-, 24-, 48- and 72-hrs). We emphasize that our goal here is not to obtain a PMP estimate, which would require amplifying more storms, but rather to examine how severe storms that have the potential to yield the PMP respond to moisture amplification. The numbers we present in the

rest of this study are for 72-hr durations (which have been shown to produce the highest correlations with discharges for U.S. West Coast basins; Warner et al. 2012).

2.d. Validation datasets

Two gridded observations datasets were used to validate model precipitation. The Analysis of Record for Calibration (AORC; Kitzmiller et al., 2018) dataset containing hourly precipitation on a 0.008° grid was used for storms that occurred between 1979-present (i.e., the December 1964 storm). The Livneh (Livneh et al., 2015) 1/16° daily precipitation (1950-2013) was used on its own (prior to 1979) or in addition to AORC for more recent storms, i.e., the November 1996 storm.

2.e. Model-based PMP estimation

Our overarching goal was to develop an implementation of model-based PMP estimation that reflects current understanding of extreme ARs and is applicable to a region, rather than a specific river basin. Below, we describe different model-based PMP implementations used in previous studies as well as the improvements we develop here.

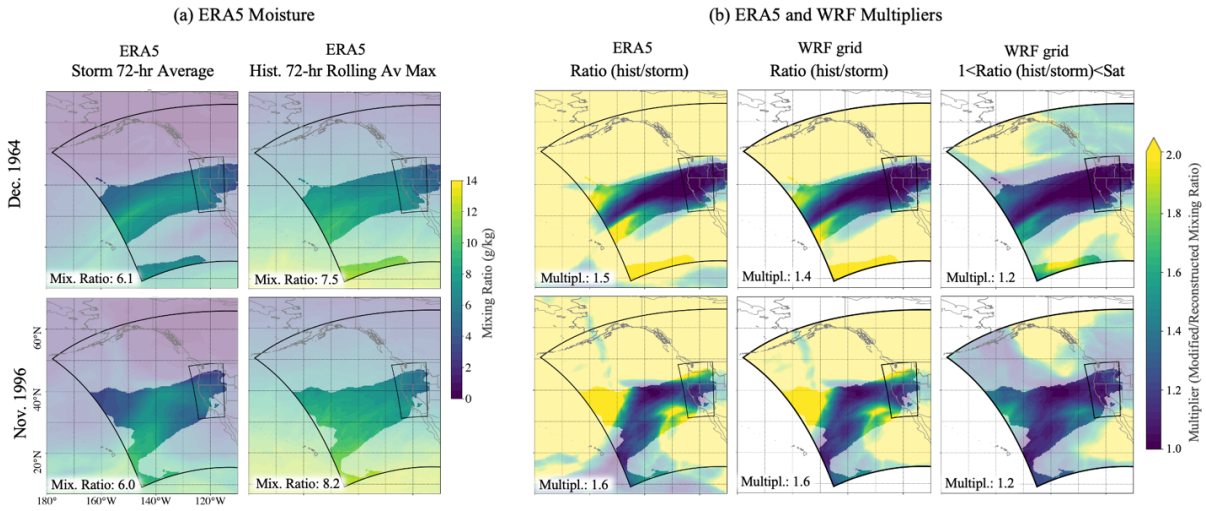
After storm selection and model reconstruction (with the goal of reproducing precipitation totals as closely as possible), the next step in obtaining PMP estimates is transposition. Transposition is the process of moving a storm in order to center its precipitation optimally over one or more river basins. Given the north-south orientation of topography along the Oregon coast, this is achieved by shifting storms to the north or south from their observed position. We performed the latitudinal shift by changing the coordinates of the atmospheric variables in the ERA5 forcing data and rerunning the WRF simulations using the shifted forcing data. We made an initial guess of the optimal storm location using a map of

reconstructed IVT averaged over the 72-hr storm period. We then quantified the forcing field displacement required to place the maximum IVT grid cell over the centroid of the river basin of interest. We performed shifts in 0.5-degree increments from the presumed optimal location until precipitation started to decrease over the river basins of interest. The moisture amplification steps described in the next paragraph are applied to each of the possible storm location we considered.

Next, the storms' moisture is amplified. Moisture refers to mixing ratio (g/kg) in all experiments. Several numerical approaches for amplifying moisture already exist; these are summarized in Table 1. All involve increasing moisture at the model boundaries. Moisture can be added at the boundary of the outer domain i.e., in the forcing dataset, which then forces the inner domain, or it can be added to the outer domain that then forces the inner domain. Because we used a wide outer domain to capture features that contribute to steering ARs towards the study basins, we took the latter approach. We implemented two existing approaches (RHM and RHP) and additionally propose an alternative method, which we call RHP-ratio. The rationale behind RHP-ratio is as follows: RHP has been shown to produce the most realistic amplified storms because the method amplifies moisture where it was already present in the observed storm (Toride et al., 2019). This approach is different from RHM, in which the atmosphere is saturated everywhere. However, there is no basis in RHP for selecting the amount by which the moisture should be amplified (e.g., multiplying by 1.2 vs. 1.6).

Here, we propose to increase moisture by the ratio of the historical maximum moisture, which we define as the maximum 72-hr rolling average within a 15-day window around the storm, to moisture during that storm (72-hr storm average). This ratio is calculated at every pixel location and level using hourly ERA5 reanalysis (Hersbach et al., 2020) between 1950-

2020, is re-gridded so that it can be applied to the WRF outer "d01" grid as a multiplier and is restricted to values between one and values that would cause saturation. Fig. 2 shows the ERA5 moisture amounts (panel a) and the multipliers at different steps (panel b). We suggest that RHP-ratio has several benefits over existing methods. It has the same degree of realism as RHP to the extent that it enhances existing moisture amounts, while having the added benefit of using historical information from a reanalysis dataset to produce physically constrained and reproducible multipliers. Further, RHP-ratio is more consistent with the HMR approach than RHM, allowing for comparisons and exploration of differences.



Chap. 4. Fig. 2. (a) Moisture amounts (mixing ratio, g/kg) and (b) multipliers used in RHP-ratio. Panel (a) shows ERA5 moisture amounts averaged over the 72-hr storm period ("Storm", left) and the historical maximum 72-hr average during the 15-day period centered around the time of the storm ("Hist.", right). Panel (b) shows from left to right the ratios obtained by dividing the latter by the former, the same ratios after re-gridding as applied to the WRF outer "d01" domain and the final multipliers after limiting the ratios to values between one and moisture increases that would cause saturation. The darker (not shaded) parts of the maps highlight locations where IVT > 250 kg/m/s in the original storm. All maps are for the average values between 500-1000 mb. The numbers on the maps are averages over the IVT > 250 kg/m/s area and between 500-1000mb.

We also investigated the importance of the location where moisture is added. The first aspect of location we focused on is the distance between where the moisture is added and the target river basin. Here we vary the location of the western boundary of the inner "d02" domain to

examine its impact on amplified precipitation. This has, to our knowledge, never been assessed though we suspected that it explains, at least in part, some of the differences among model-based estimates. The other aspect of location we examine is whether the moisture is added everywhere in the domain, or only in the path of the storm using the RHM-IVT or RHP-IVT approach (Table 1). This approach can be used in conjunction with RHM, RHP or RHP-ratio and restricts moisture amplification to areas where IVT > 250 kg/m/s. Given that the magnitude of moisture increases in RHP-ratio depends on how much moisture was originally present, we expect that using the IVT criterion to restrict moisture amplification outside of the storm may become unnecessary. We investigated the two above hypotheses by running sensitivity analyses in which RHP-ratio simulations are performed with different western domain boundaries as well as with and without the IVT criterion.

Chap. 4. Table 1. Overview of moisture amplification approaches considered, including existing approaches described in the literature as well as our proposed RHP-ratio method.

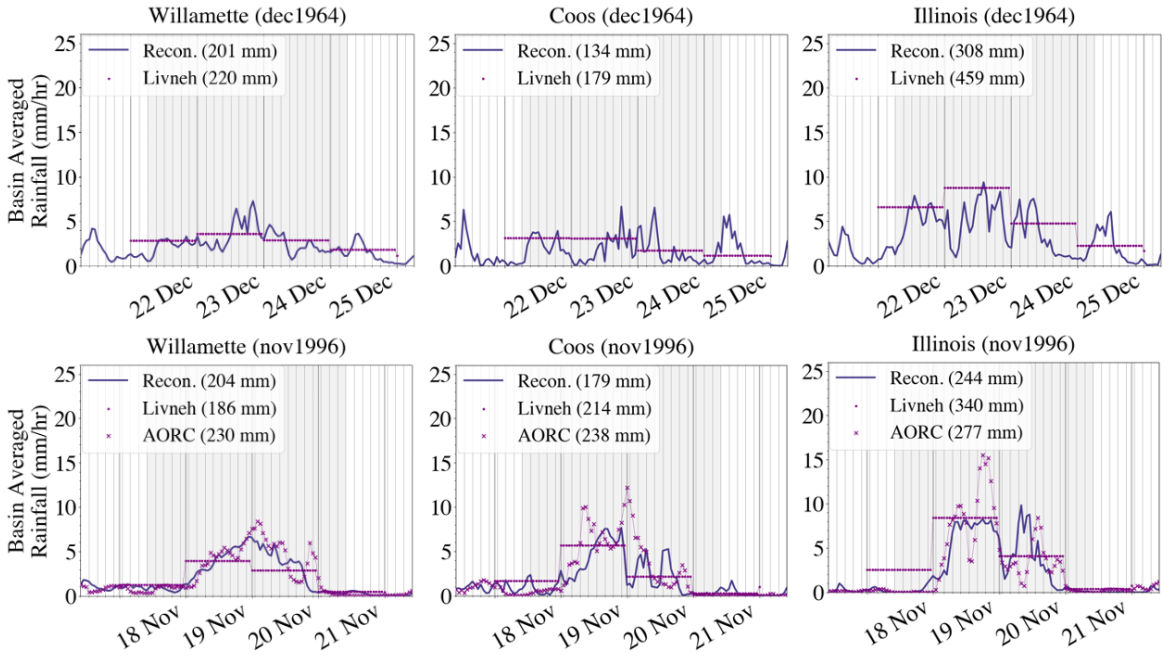
	Moisture Amount	Location Moisture Added	References
Relative humidity maximization (RHM)	Moisture is increased to saturation (i.e., RH=100%).	Everywhere (increase less pronounced where more moisture present).	Ishida et al., 2015
Relative humidity perturbation (RHP, e.g., RHP-3.0)	Moisture is amplified by a uniform multiplier (e.g., 3.0). The largest multiplier we use is 4.0 as it causes near-saturation of the storms (Fig. 6).	Everywhere but increase is more pronounced where more moisture present.	Toride et al., 2019
RHM-IVT or RHP-IVT	Depends on the method it is used in conjunction with.	Restricted to only the path of the AR (where IVT > 250 kg/m/s).	Toride et al., 2019
RHP-ratio	Moisture is amplified by a multiplier (moisture during the storm to the historical maximum).	Everywhere, where increase is a function of the historical maximum.	This study

3. Results

Building on recent advances improving the physical realism of model-based moisture amplification (Toride, 2019), we present a new moisture amplification method called RHP-ratio and describe how it compares with existing approaches. We also address the issue (common to all model-based methods) of case-by-case storm centering versus use of homogeneous regions.

3.a. Model reconstructions of historical storms

The first step in model-based PMP estimation is the identification and reconstruction of severe storms. The quality of our WRF precipitation reconstructions (Fig. 3) appears to be adequate (similar to earlier studies on AR-based PMP, see Tarouilly et al., 2022 in the Feather and Toride et al., 2019 in the Willamette) in the three basins we are considering (Willamette, Coos, Illinois). We compared model reconstructions to the Analysis of Record for Calibration (AORC; Kitzmiller et al., 2018) dataset when available (i.e., for the 1996 storm) and the Livneh dataset (Livneh et al., 2015) for both the 1964 and 1996 storms. Given observational uncertainty in the gridded datasets (evidenced by differences between the AORC and Livneh precipitation totals for the 1996 storm, also see Newman et al., 2015) the observed and reconstructed storm precipitation may be considered similar.



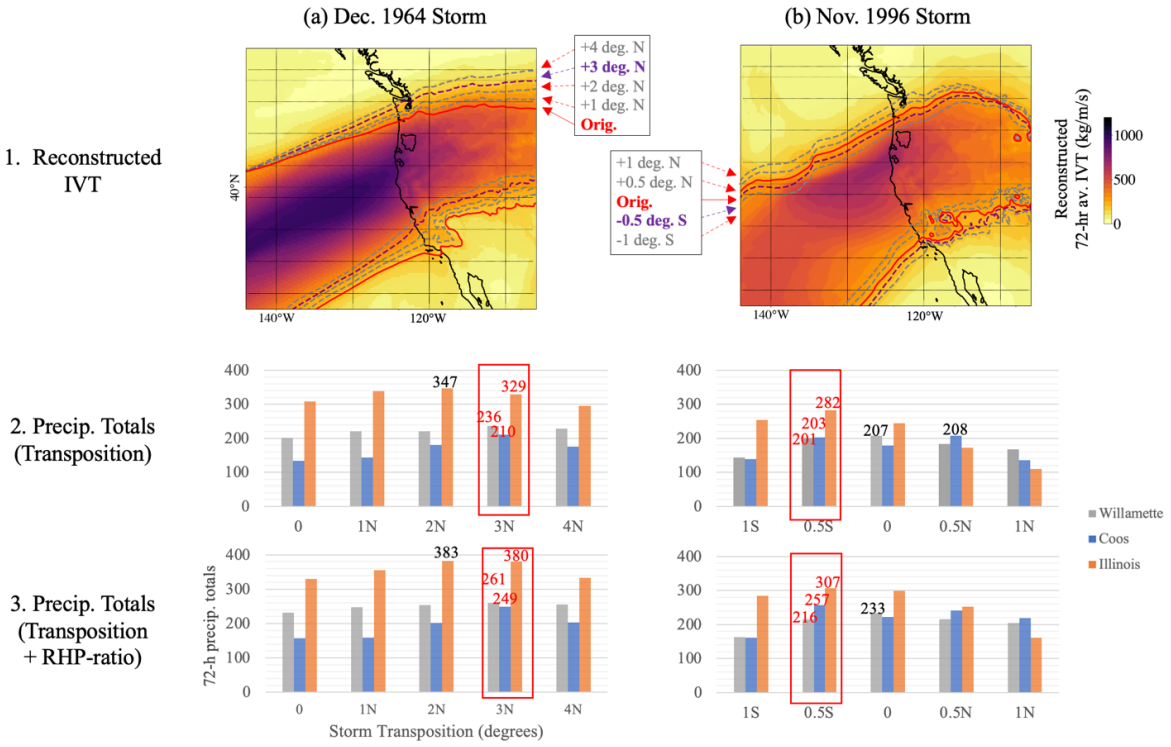
Chap. 4. Fig. 3. Observed (dotted lines) and reconstructed (solid lines) precipitation timeseries with precipitation totals shown in the legend for each dataset. The grayed part of the timeline shows the 72-hour storm period used to calculate precipitation totals. The solid blue line is the model reconstruction, dotted purple lines are observations. Major gridlines are daily and minor gridlines 3-hourly.

3.b. Simultaneous optimization of storm position over multiple river basins

The next step in PMP estimation is storm transposition. Fig. 4 (top row) shows different possible storm positions corresponding to different latitudinal shifts for the two storms of interest in coastal Oregon. We investigated whether it is possible to position a storm in a way that maximizes (or nearly maximizes) precipitation for our three study basins simultaneously so as to reduce the number of simulations required in subsequent steps. The overall optimal storm position (i.e., the storm position that maximizes the sum of precipitation totals over all three basins) is shown on Fig. 4 maps (top row) by the purple outlines and on the Fig. 4 bar charts (middle and bottom rows) by red boxes. Bar charts of precipitation totals (Fig. 4 middle and bottom rows) are shown for both transposition only (middle row) and

transposition combined with moisture amplification (Fig. 4 bottom row). The latter are discussed in Section 3.f. after moisture amplification is implemented.

We find that for the very wide AR storms presented here, the overall optimal position nearly maximizes precipitation over all basins, with some relatively minor decreases (5% at most) in some. Fig. 4 shows that for the storm of 1996, the overall optimal position (0.5-degree shift to the south) led to a small decrease in precipitation totals (from 207 to 201 mm) in the Willamette basin compared to the individual optimal position for that basin (which happens to be the original storm position). The overall optimal position for 1996 also led to a small decrease in the Coos (from 208 to 203 mm) compared to its individual optimal position (a 0.5-degree shift to the north). For the 1964 storm, the overall optimal position (3-degree shift to the north) caused a decrease (from 347 to 329 mm) in the Illinois basin compared to the individual optimal position for that basin (2-degree shift to the North). There were no other decreases. These differences arguably are within the range of uncertainties affecting PMP estimates (see Section 3.a., Tarouilly et al., 2022 for model uncertainty and Tarouilly et al. in review for small sample size uncertainty). If so, it may be acceptable to amplify only the transposed simulation that produced the overall optimal storm position, instead of amplifying a different transposed simulation for each basin. Nonetheless, we amplified all simulations for all river basins to test whether the small decreases in precipitation due to using a single storm position remain acceptable after moisture amplification (Fig. 4, bottom row; Section 3.f.).



Chap. 4. Fig. 4. Different possible AR positions (top row) and resulting 72-hr precipitation totals over the three target basins without (middle row) and with moisture amplification (bottom row; as discussed in Section 3.f.). The contours on the maps show the area where IVT > 250 kg/m/s: the red solid line is the original position of the storm, dotted lines show new position after transposition (the purple dotted line is the overall optimal position). On the bar plots, the red boxes and numbers show precipitation totals for the overall optimal storm position for each basin. The black boxes and numbers highlight basins in which precipitation would be higher using a different storm position.

3.c. Moisture increases implemented in the outer ("d01") modeling domain

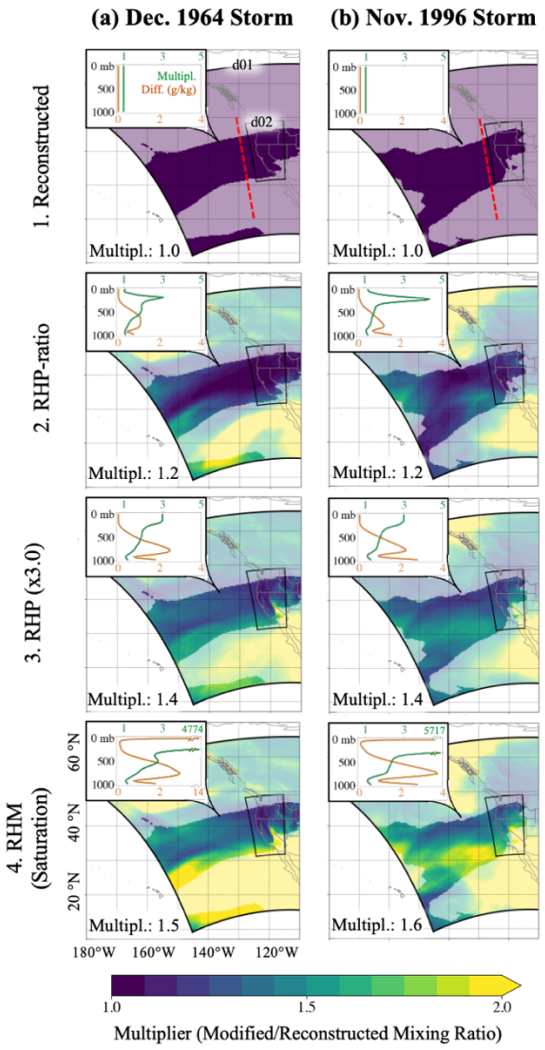
Here we consider the moisture increases produced by RHP (per Toride et al., 2019: multipliers ranging from 1.2-4.0), RHP-ratio (the proposed method; variable multiplier determined by the historical maximum) and RHM (per Ishida et al., 2015: variable multiplier that corresponds to forced saturation) within the outer ("d01") model domain upwind of the study river basins. Note that we examined RHP in simulations with multipliers ranging from 1.2-4.0, but in the interest of space we only show selected RHP results.

We first examine moisture changes in the outer d01 modeling domain that result from the different methods (RHM, RHP and RHP-ratio) (Fig. 5). We averaged the multipliers shown on the Fig. 5 maps over the area in the outer ("d01") domain where IVT > 250 kg/m/s and 500-1000 mb pressure heights. Multipliers were 1.2 for both storms for RHP-ratio, 1.4 for both storms for RHP-3.0 and 1.6 in 1996 and 1.5 in 1964 for RHM. Note that all the multipliers discussed here were capped at values that would cause saturation (Fig. 3). This is why, in the case of RHP-3.0 for example, the multipliers are much lower than the values the name indicates. It is interesting to note that the existing HMR approaches typically capped precipitation increases at 1.5 citing that larger increases may modify storm dynamics (HMR 55A; Hansen et al., 1988; HMR 57; Hansen et al., 1994; and HMR 59; Corrigan et al., 1999), which is very similar to the largest multipliers in our simulations (1.6).

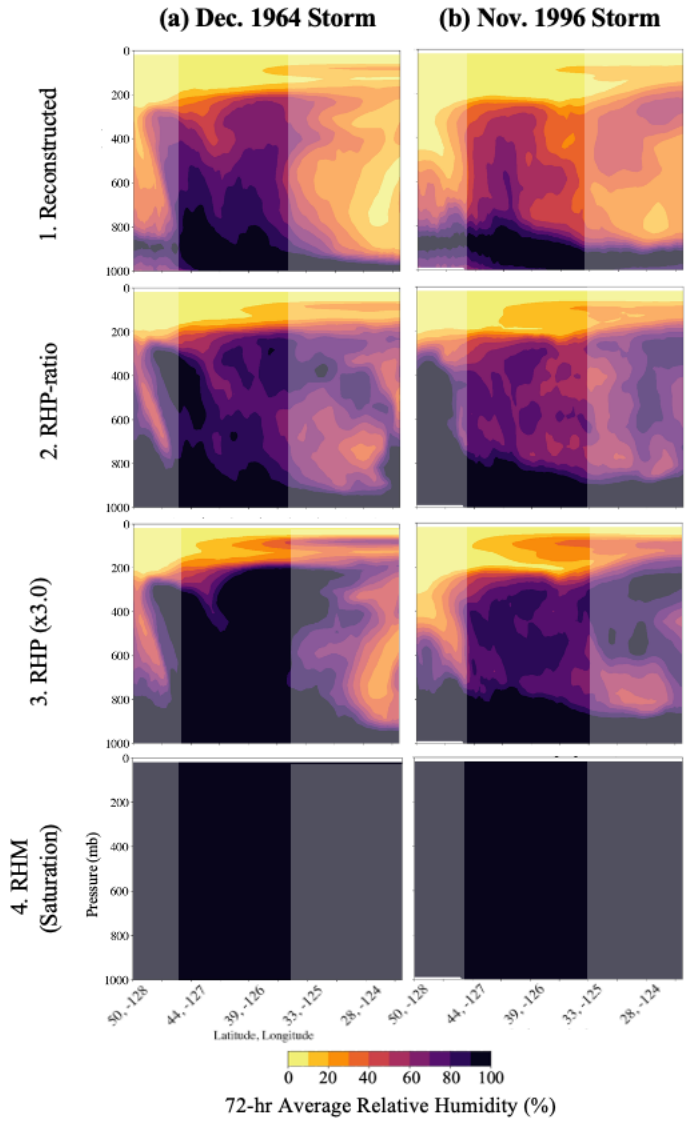
Vertical profiles of moisture changes (shown in the insets, Fig. 5) provide additional information. Multipliers (brown lines, Fig. 5 insets) indicate that all methods caused large moisture increases relative to original moisture amounts at the top of the troposphere, but the increases caused by RHM are much greater than for the other methods: multipliers are up to ~5000 for RHM (which is clearly unrealistic), while they do not exceed 4.0 for the other methods. The absolute increases (green lines, Fig. 5 insets) also highlight that the very large moisture increases at the top of the atmosphere in RHM do not occur in RHP-3.0 or RHP-ratio. In these simulations, the most substantial increases largely occur around 850 mb.

We next examine relative humidity i.e., how close the atmosphere is to saturation, as a result of the moisture increases we implemented (Fig. 6). We took cross-sections along the dashed line shown on Fig. 4 top row maps. We found that gradually increasing multipliers in RHP-1.2-4.0 led to an atmosphere that is gradually closer to saturation, with the largest multipliers producing a cross-section similar to that produced by RHM. We note that even in RHP-3.0 or

RHP-ratio (in which saturation is not the goal), the atmosphere did reach saturation especially between 800-1000 mb, which constrains moisture increases regardless of the multiplier used. RHP-ratio led to the atmosphere being closer to saturation in the northern portion of the domain, likely due to the lower air temperature and saturation vapor pressure. However, the atmosphere at the cross-section was not entirely saturated even for RHM, likely due to feedbacks (e.g., precipitation). Importantly, the December 1964 storm had a more saturated atmosphere than the November 1996 storm before any modifications were made, consistent with our findings above, which show that RHM for the 1964 storm requires a lower multiplier (1.5) than for 1996 (1.6) to reach saturation.



Chap. 4. Fig. 5. Changes in moisture (mixing ratio, g/kg) for different moisture amplification experiments relative to reconstructions. Maps show the multiplier averaged over 72-hrs and 500-1000mb pressure heights. Multipliers for reconstructions (top row) are one i.e., there is no change relative to reconstruction. Maps for RHP-ratio are the same as in Fig. 2 and are shown here again to allow for comparisons with other methods. Insets show the vertical profiles of multipliers in brown and increase amounts in green. Vertical profiles are averaged over the darker (not shaded) area shown on the map (where original IVT > 250 kg/m/s) and the same 72-hr storm duration; note the x-axis scale for RHM.



Chap. 4. Fig. 6. Relative humidity (RH) vertical cross-sections (taken along the red dashed line just upwind of the inner "d02" domain, shown on Fig. 2, upper panels) for different moisture amplification experiments. Darker (not shaded) areas show parts of the profile where IVT > 250 kg/m/s. Note that 100% RH corresponds to a saturated atmosphere. RH for RHM experiments (bottom row) is 100% i.e., the atmosphere is fully saturated everywhere.

3.d. Moisture and precipitation response over the study basins

We now examine moisture and precipitation over the study river basins in response to the moisture forcings in the outer "d01" domain described in Section 3.c. above. Fig. 7 provides an overview of 72-hr atmospheric moisture averages and precipitation totals for each storm

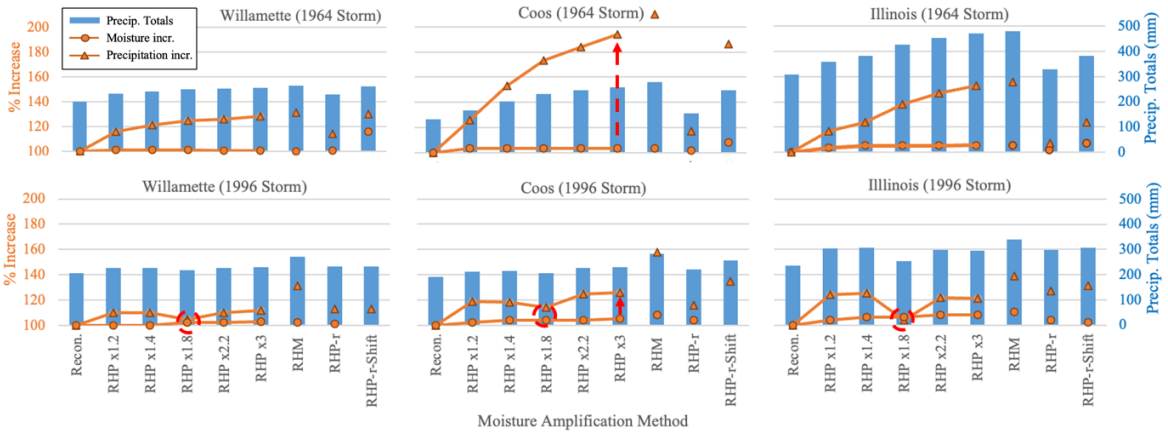
and river basin, while Figs. 8 & 9 additionally show the spatial and temporal patterns of precipitation. Irrespective of the approach, we note that moisture amplification tends to increase precipitation in regions of already high orographic precipitation (Fig. 8) and tends to fill-in periods of relatively low precipitation rather than increasing precipitation peaks (Fig. 9) (also see Tarouilly et al., 2022).

The moisture increases over the river basins were modest (below 10% for all combinations of storms and basins) for RHP 1.2-4.0 and RHP-ratio (Fig. 7, orange lines and circles). Stronger moisture forcing (e.g., RHP-1.4 as opposed to RHP-1.2) tended to produce larger moisture increases. While RHM also always produced modest moisture increases below 10%, the increases were slightly larger for the storm of November 1996 than for December 1964. RH cross-sections (Fig. 6, top row) suggest this is because the storm of December 1964 was closer to saturation prior to any modifications.

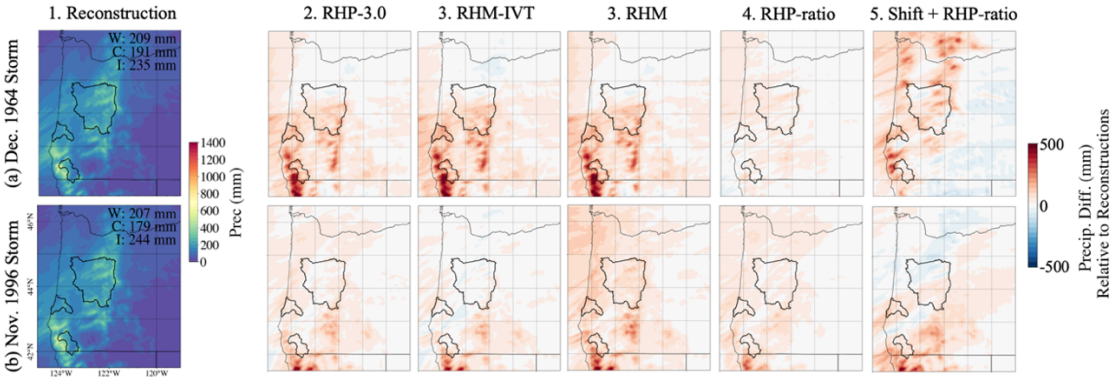
In turn, those modest moisture increases over the basins caused precipitation responses (Fig. 7, orange lines and triangles) that vary but can be large: ranging between 10-54% for RHP-1.2-4.0 and similarly between 31-110% for RHM. The magnitude of these precipitation increases strongly depend on the storm and basin for both RHP 1.2-4.0 and RHM: with RHP-2.2 for example, the moisture increases were 12% (1964) and 13% (1996) over the Coos basin, but the precipitation increase was over 80% in 1964 as opposed to under 30% in 1996 (Fig. 7; see red arrows). We additionally observe non-linearities even for a given storm and basin. For example, the moisture increase was larger, but the precipitation increase lower, with RHP-1.8 than with RHP-1.4, for all river basins during the storm of 1996 (Fig. 7; see red circles).

We further examine the precipitation time series (Fig. 9) to understand the precipitation increases we see during some simulations. Taking the storm of 1996 over the Coos river

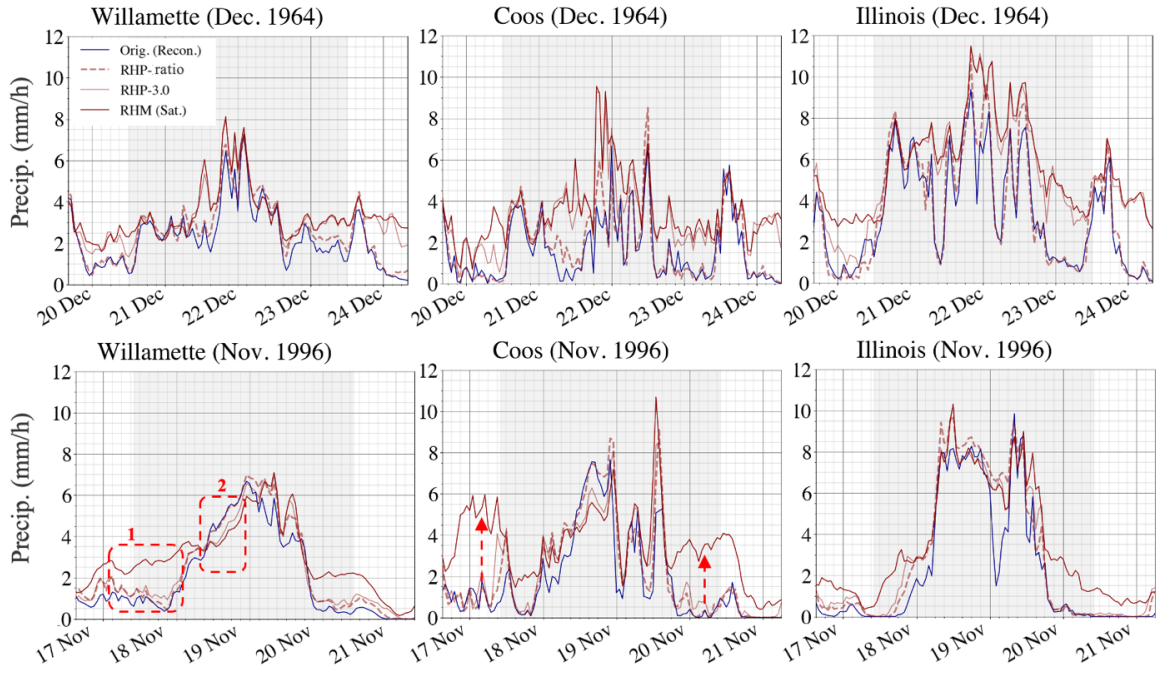
basin as an example (Fig. 9, see arrows), RHM caused a large increase in precipitation before and after the period of highest-intensity precipitation: this is likely due to saturation of the entire domain in RHM, which makes moisture available to precipitate before and after the original period of high-intensity precipitation. We also note that RHM (and to a lesser extent RHP) sometimes produced a large increase in precipitation compared to reconstructions at the beginning of the storm followed by an unintended decrease (Fig. 9, red boxes 1 and 2). This finding may be explained by a large, possibly unrealistic amount of moisture being precipitated at the beginning of the storm that is no longer available later.



Chap. 4. Fig. 7. Moisture increases (orange lines/circles), precipitation increases (orange lines/triangles) and precipitation amounts (blue bars) over the study river basins for different moisture amplification experiments. All values are basin averages (for moisture) or totals (for precipitation) over the 72-hr storm periods.



Chap. 4. Fig. 8. Spatial patterns of precipitation: reconstructions (left) and differences between reconstructions and different moisture amplification experiments (right). The numbers on the reconstructed precipitation maps indicate 72-hr precipitation totals. The maps shown for Shift + RHP-ratio are for the overall optimal position and therefore may not cause a precipitation increase in all individual basins.



Chap. 4. Fig. 9. Basin-averaged precipitation timeseries for reconstructions and different moisture amplification experiments (different colors). The grayed part of the timeline shows the 72-hour storm period used to calculate precipitation totals. Major gridlines are daily and minor gridlines 3-hourly.

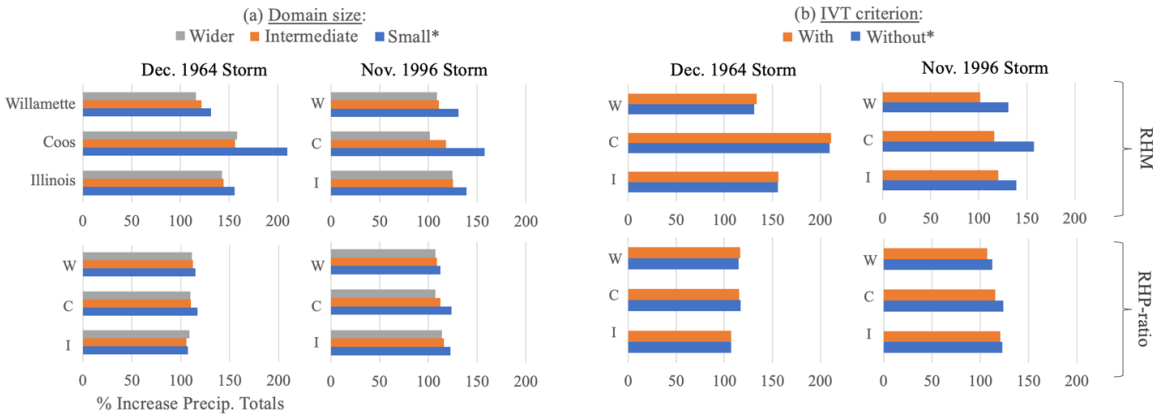
3.e. Decreased importance of added moisture location when using RHP-ratio

The previous section assessed the importance of the magnitude and vertical distribution of moisture increases by RHP-ratio; we now examine the importance of the geographic location of the moisture increases in RHP-ratio. We focus in particular on (a) the distance from the target basin at which the moisture is added, determined by the location of the inner "d02" domain western domain in this case (see Fig. 1 for domain sizes) and (b) where in the domain the moisture increases are implemented, e.g., everywhere as opposed to only in the path of the storm using the IVT > 250 kg/m/s criterion. We examine here whether these considerations have an impact on amplified precipitation totals in RHP-ratio and for comparison, in RHM.

Precipitation totals amplified by RHM are strongly dependent on the location of the inner domain western boundary for the two storms we examined (Fig. 10a, top row). Taking the Coos river basin as the most pronounced example, precipitation increases ranged widely between 58-110% (1964) and 2-57% (1996) depending on domain size. We also note that the intermediate basin size unexpectedly produces less precipitation than the wider one for the Coos basin. When RHP-ratio was used instead of RHM (Fig. 10a, bottom row), the differences in precipitation increases due to domain size were greatly reduced with increases ranging 7-24% in 1996 and 10-17% in 1964 for the Coos basin.

We next examine the dependency of amplified precipitation totals to whether moisture is added everywhere as opposed to only in the path of the storm (called RHM-IVT, if the IVT criterion is applied in addition to the RHM approach as described in Table 1, or RHP-ratio-IVT if applied to RHP-ratio). We observed large differences in precipitation increases

between RHM and RHM-IVT, i.e., strong dependence on the IVT criterion in the 1996 storm (the differences were more limited in 1964) (Fig. 10b). As expected, we no longer observed large differences between RHP-ratio and RHP-ratio-IVT for either storm, i.e., there is no dependence on the IVT criterion when used in addition to the RHP-ratio approach (Fig. 9b). RHP-ratio and the IVT criterion both have the effect of limiting moisture increases to where moisture was present in the original storm, making the IVT criterion no longer necessary if RHP-ratio is used.



Chap. 4. Fig. 10. Percent increase in precipitation relative to reconstructions for different amplified simulations (top row: RHM; bottom row: RHP-ratio). The different colors correspond to different domain sizes as shown on Fig. 1 (a., left panel) or whether the IVT criterion is used (b., right panel). The asterisk in the legend (i.e., the blue bars) indicates the simulations which have been used throughout the paper.

3.f. Transposition combined with RHP-ratio

This section describes the final amplified precipitation totals obtained by the combination of storm transposition and RHP-ratio. Applying RHP-ratio to transposed storms further increases precipitation totals (Fig. 4). The overall optimal storm positions identified in Section 3.a. remain the same after RHP-ratio is applied to the transposed storms (Fig. 4, bottom row). This is an important point as it suggests amplifying storms at all positions is

likely unnecessary, therefore greatly reducing the number of simulations needed. We still find after applying RHP-ratio to transposed storms that precipitation totals are slightly decreased in some river basins when using the overall optimal storm position rather than each basin's individually optimized storm position (from 233 to 216 mm for the storm of 1996 in the Willamette River basin and from 383 to 380 mm for the 1964 storm in the Illinois River basin; there is no longer a decrease in the Coos River basin in 1996) (Fig. 4, bottom row).

4. Discussion

While we reiterate that our goal here was not to produce a PMP estimate, it nonetheless is of interest to place the values we obtained in the context of the observed precipitation totals and existing PMP estimates. We focus on the Willamette River basin as it has been the subject of several previous studies. The largest historical 72-hr precipitation total based on ERA5 is 240 mm for the1964 storm (our WRF reconstruction of the same storm produced a 72-hr total of 207 mm). The largest amplified precipitation totals obtained in this study by the combination of transposition (3-degree shift to the north) and RHP-ratio for the same 1964 storm is 261 mm in 72 hours. Existing previous PMP estimates include 328 mm (model based; Toride et al, 2019) and 427 mm (HMR57; Hansen et al., 1994). While Toride et al. (2019) and this study use similar approaches, the moisture amplification method is slightly different (RHP-IVT vs. RHP-ratio) and there are additional differences in model configuration, domain size and position and forcing datasets. The fact that the Toride et al. (2019) value is only 25% larger than our value, given these methodological differences, speaks to the robustness of model-based PMP estimation.

Regarding the apparent differences between HMR and model-based estimates, we highlight that current model-based methods are based on storm transposition and amplification, while

HMR methods additionally include an enveloping step. Envelopment is the process for selecting the largest values from depth-area-duration curves obtained after the transposition and amplification of different storms; it is meant to address the limited number of storms used in PMP estimation (WMO, 2009). As a result, comparisons should not be made between final HMR and model-based values, but rather it should be recognized that what is commonly referred to as "model-based PMP" is in fact the model-based transposition and amplification steps of PMP estimation. Therefore, if comparisons are to be made, they should be between a given step (e.g., amplification) of HMR and model-based PMP procedures. Unfortunately, such comparisons are difficult to make as the HRM reports typically do not provide intermediate precipitation totals before and after moisture scaling.

We also discuss considerations for possible improvements to the RHP-ratio method. We suggest that capping moisture increases at saturation (if reached) in all model-based methods including RHP-ratio may cause the amplification step to produce lower model-based estimates compared to HMR methods. This is because it may create an artificial upper bound controlled by air temperature if air temperature is not modified. Existing HMR methods on the other hand, use surface dew point temperatures to obtain the ratio by which to scale precipitation: the historical maximum dew point may have occurred on a warmer day than the storm which is being amplified, possibly allowing for larger multipliers. The differences in the multipliers before and after application of the saturation constraint are large (see Fig. 2), suggesting that model-based PMP values could be larger if the air temperature was allowed to increase to release or reduce the saturation constraint.

Besides, we point out that historical maximum moisture amounts used in this study likely approach but may not reach the upper bound of moisture. However, there is currently no basis for selecting different moisture amounts between the historical maximum (as in RHP-

ratio, likely too low) and saturation (as in RHM, likely too high). One way to consider would be to search large global model ensembles for the historical period (such as the CESM2 large ensemble; Danabasoglu et al., 2020) for storms that may have higher moisture amounts than observed ones. While there is no guarantee that these storms would have the highest moisture amounts possible, they are much more rare storms that observed ones as they can be selected from a larger sample, and are therefore likely much closer to the upper bound.

5. Conclusions

Our goal was to improve the applicability of model-based PMP. We did so by addressing two main challenges: the computational feasibility (number of model runs required) of the transposition step for producing regional as opposed to single-basin estimates, and the consistency and defensibility of the moisture amplification step.

Regarding transposition, we showed for the cases we examined (1) that it may be possible to use a single storm position for multiple basins and (2) that the optimal storm position is unchanged by moisture amplification. As a result, several simulations would still be required to identify the optimal storm position, but only one would need to be re-un for moisture amplification, (i.e., this process would not need to be repeated for every basin). If this can be extended to other storms and other basins beyond the three examined here, it would substantially reduce the computational requirements to produce regional model-based PMP estimates and enable the production of statewide estimates.

We additionally found that moisture amplification is a key part of the PMP estimation process: it is the largest source of uncertainty in model-based PMP estimates (e.g., in comparison to model uncertainty, see Tarouilly et al., 2022 and to small sample size uncertainty, see Tarouilly et al. in review) and as such should be the focus on future efforts.

Building on recent attempts to improve the realism of model-based moisture amplification (Toride et al., 2019), we proposed another approach (RHP-ratio) that uses the historical maximum mixing ratio to constrain moisture increases. While we acknowledge that the method as implemented here will require fine-tuning, it provides a strong methodological basis around which model-based PMP could be standardized. Importantly, our experiments (in which larger moisture amounts did not necessarily produce more precipitation) highlighted that the vertical distribution of moisture matters in addition to the moisture amounts. Both this study as well as Toride et al. (2019) additionally suggest that moisture needs to be "optimized" rather than maximized in order to maximize precipitation. It should also be noted that beyond optimizing moisture to produce more precipitation, the realism of the modifications needs to be considered: a large part of the decisions that have made PMP estimation (whether HMR or model-based) somewhat subjective have to do with balancing safety and realism when determining how to amplify storms.

Moisture amplification being a key part of PMP estimation, we recommend further assessment of sensitivities of PMP estimates to any parts of the process that are somewhat subjective. In particular, we have noted that air temperature should possibly be allowed to vary in order to avoid imposing an artificial upper bound on the amounts of moisture the atmosphere can hold. This control that air temperature exerts on moisture amounts via the Clausius-Clapeyron relationship and ultimately on PMP estimates will become particularly important for the estimation of PMP in a warming climate. We additionally recommend assessing whether historical maximum moisture is a sufficient approximation of the upper bound of moisture or whether alternative datasets (e.g., global model large ensembles, such as the CESM2-LE) should inform different moisture amounts to be used for storm amplification. Our results are broadly applicable to mountainous coastal watersheds in the U.S. West affected by AR storms. The feasibility of model-based moisture amplification

needs to be evaluated in other regions, in particular those with convective storms, which have been more challenging for models to reproduce accurately (Mahoney et al., 2021).

We reiterate the importance of comparing new model-based estimates as they are developed both with earlier model-based estimates as well as with currently in-use HMR estimates. We noted that despite methodological differences, our values are not entirely different (within 16%) from other model-based studies, pointing to some degree of robustness in the model-based approach. Due to the lack of detail provided in the HMRs, it may be necessary to re-estimate values produced by the HMR moisture amplification step for comparison with model-based estimates, but this would allow to conclude whether model-based estimates really are lower than HMR ones, and if so, identify why. It must be remembered that though they have rarely been quantified, uncertainties likely exist both in the HMR procedures as well as in the observation data that they rely on. Besides, the return period of PMP estimates varies widely, between 10^5-10^9 years (M. Schaefer, personal communication, Dec. 2023): comparisons of model-based PMP estimates with precipitation frequency estimates may be a starting point for explaining the discrepancies and possibly informing more consistent PMP estimates.

The National Research Council recommended in 1994 not to change HMR PMP methodology for “lack of a clearly better alternative” but identified high-resolution numerical modeling as an avenue for improvement. Thirty years later, this work demonstrates that NWP modeling has provided a tool to better understand the dynamics of extreme events, precipitation response to moisture in particular, and to better reflect their role in more physically-based PMP procedures.

Chapter 5: Conclusions

At the outset, my main objective in this dissertation was to develop guidance for updating PMP estimates that incorporate all available storm data, reflect current knowledge of precipitation processes and that can be applied in a changing climate. I focused on model-based approaches to PMP estimation, which are emerging as a credible alternative to HMR methods once several key sources of uncertainty are addressed. My overarching goal has been to characterize and reduce, where possible, known sources of uncertainty that affect model-based methods. Ultimately, the key question I addressed is: could a PMP storm could produce more precipitation than model-based estimates suggest? Or are we over-designing for unrealistically large events? Answering these questions is a necessary step before incorporating model-based approaches into long-overdue updating of regional PMP estimates.

In my view, this dissertation makes the following original contributions. I found that model uncertainty and uncertainty related to the small sample of severe historical storms have a modest (~10%) impact on PMP estimates (Chapters 2 and 3). In so doing, I also demonstrated the benefits of using large model ensembles to better understand very rare events (Chapter 3). Additionally, I showed that uncertainty related to moisture amplification had a much larger impact (closer to 30%) on PMP estimates and should therefore be the focus of ongoing efforts to improve model-based PMP guidelines (Chapter 4). The RHP-ratio approach that I propose in Chapter 4 is an important step in that direction as it provides a more physically-based approach to PMP estimation than earlier methods by grounding moisture increases in physical observations of extreme moisture amounts (climatology).

While there are many ways besides moisture maximization to increase storm precipitation in a model-based PMP framework, none have been shown to be entirely suitable so far. Such approaches include maintaining the boundary conditions that produced peak precipitation for the entire storm duration, but as with fully saturating the atmosphere, it is difficult to justify whether doing so is realistic. Attempts have been made at enhancing horizontal winds but it has been difficult to determine how, and existing experiments have not produced substantial precipitation increases. There may be a very large number of atmospheric variables to adjust simultaneously in order to maximize precipitation and knowledge of which variables would need to be modified and how is currently lacking. One avenue to explore may be the select the highest-efficiency (instead of highest-precipitation) storms from the historical record and then moisture maximize them.

My work has shown that moisture (and likely other variables) needs to be "optimized" rather than maximized in order to maximize precipitation. Even if optimizing or maximizing moisture (or another variable) produces more precipitation, whether that modification is realistic needs to be considered. The large number of decisions that have made PMP estimation somewhat subjective ultimately have to do with balancing safety and realism. While numerical weather prediction models do not tell us the upper bounds of any variable, the underlying model equations follow the laws of physics and large ensembles of model runs can help us to determine where upper bounds may lie. The storms from such large ensembles, some of which are more severe than ever observed, can be used either to inform how to adjust historical storms or may be used directly as the PMP storm.

Updating PMP estimates using model-based methods will at some point require rigorous comparisons with HMR estimates and differences will need to be explained. Final HMR and final model-based values are not directly comparable as the steps involved are not exactly the

same. For example, the HMRs perform enveloping of precipitation curves for different areas and durations in addition to transposition and moisture maximization; this has not been done in model-based methods. Comparisons of values from e.g., the moisture maximization step of HMR and model-based approaches would be useful, but the HMRs rarely contain information about intermediate values. It must be remembered that though they have rarely been quantified, uncertainties likely exist both in the HMR procedures as well as in the observations on which they rely. The equivalent return period of PMP estimates varies widely, between 10^5-10^9 years according to the National Research Council (1994), such that comparisons between PMP and precipitation frequency estimates would also be needed.

The value of developing improved model-based PMP estimation procedures lies both in their importance for estimating the PMF for historical conditions, and in providing a mechanism for doing so in a future climate. The atmospheric simulations produced using the WRF model to obtain PMP estimates can be used directly to force a distributed hydrologic model and obtain the PMF. The atmospheric simulation for a given storm could be combined with multiple antecedent conditions scenarios to obtain a range of model-based PMF estimates. The model-based approach also lends itself to re-evaluating PMP and PMF in a warming climate, this could be done either using a pseudo-global warming approach i.e., repeating my experiments with higher air temperature at the model boundaries (Brogli et al., 2023) or instead applying model-based PMP methods directly to storms from climate projections. In addition to being an interesting theoretical question ("Are floods bounded?"), understanding precipitation extremes is evolving as a question of key significance for the safety and resilience of infrastructure in a changing climate.

REFERENCES

Abbs, D. J. (1999). A numerical modeling study to investigate the assumptions used in the calculation of probable maximum precipitation. *Water Resources Research*, *35*(3), 785–796.

Berner, J., Shutts, G. J., Leutbecher, M., & Palmer, T. N. (2009). A Spectral Stochastic Kinetic Energy Backscatter Scheme and Its Impact on Flow-Dependent Predictability in the ECMWF Ensemble Prediction System. *Journal of the Atmospheric Sciences*, *66*(3), 603–626. https://doi.org/10.1175/2008JAS2677.1

Berner, J., Ha, S. Y., Hacker, J. P., Fournier, A., & Snyder, C. (2011). Model Uncertainty in a Mesoscale Ensemble Prediction System: Stochastic versus Multiphysics Representations. *Monthly Weather Review*, *139*(6), 1972–1995. https://doi.org/10.1175/2010MWR3595.1

Berner, J., Fossell, K. R., Ha, S. Y., Hacker, J. P., & Snyder, C. (2015). Increasing the Skill of Probabilistic Forecasts: Understanding Performance Improvements from Model-Error Representations. *Monthly Weather Review*, *143*(4), 1295–1320. https://doi.org/10.1175/MWR-D-14-00091.1

Brogli, R., Heim, C., Mensch, J., Sørland, S. L., & Schär, C. (2023). The pseudo-global-warming (PGW) approach: methodology, software package PGW4ERA5 v1.1, validation, and sensitivity analyses. *Geosci. Model Dev*, *16*, 907–926. https://doi.org/10.5194/gmd-16-907-2023

Bruyère, C. L., Done, J. M., Holland, G. J., & Fredrick, S. (2014). Bias corrections of global models for regional climate simulations of high-impact weather. *Climate Dynamics*, *43*(7–8), 1847–1856. https://doi.org/10.1007/S00382-013-2011-6/FIGURES/9

Caldwell, P., Chin, H. N. S., Bader, D. C., & Bala, G. (2009). Evaluation of a WRF dynamical downscaling simulation over California. *Climatic Change*, *95*(3–4), 499–521. https://doi.org/10.1007/S10584-009-9583-5

Cannon, F., Weihs, R., Steinhoff, D. F., Papadopoulus, C., Kawzenuk, B., Mulrooney, P., Zheng, M., Ao, P., Cobb, A., Wilson, A. M., Martin, A. C., Reynolds, D., Subramanian, A. C., Delle Monache, L., & Ralph, F. M. (2022). Precipitation Forecast Skill and Uncertainty Over California Watersheds in a High-Resolution Ensemble. *Monthly Weather Review. In Review*.

Cao, Q., Mehran, A., Ralph, F. M., & Lettenmaier, D. P. (2019). The Role of Hydrological Initial Conditions on Atmospheric River Floods in the Russian River Basin. *Journal of Hydrometeorology*, *20*(8), 1667–1686. https://doi.org/10.1175/JHM-D-19-0030.1

Chapman, W. E., Subramanian, A. C., Delle Monache, L., Xie, S. P., & Ralph, F. M. (2019). Improving Atmospheric River Forecasts With Machine Learning. *Geophysical Research Letters*, *46*(17–18), 10627–10635. https://doi.org/10.1029/2019GL083662

Chen, L. C., & Bradley, A. A. (2006). Adequacy of using surface humidity to estimate atmospheric moisture availability for probable maximum precipitation. *Water Resources Research*, *42*(9). https://doi.org/10.1029/2005WR004469

Chen, X., & Hossain, F. (2016). Revisiting extreme storms of the past 100 years for future safety of large water management infrastructures. *Earth's Future*, *4*(7), 306–322. https://doi.org/10.1002/2016EF000368

Chen, X., & Hossain, F. (2018). Understanding model-based probable maximum precipitation estimation as a function of location and season from atmospheric reanalysis. *Journal of Hydrometeorology*. https://doi.org/10.1175/JHM-D-17-0170.1

Cobb, A., Delle Monache, L., Cannon, F., & Ralph, F. M. (2021). Representation of Dropsonde-Observed Atmospheric River Conditions in Reanalyses. *Geophysical Research Letters*, *48*(15), 93357. https://jglobal.jst.go.jp/en/detail?JGLOBAL_ID=202102275008651677

Corrigan, P., Fenn, D. D., Kluck, D. R., & Vogel, J. L. (1999). *Hydrometeorological Report No. 59. Probable Maximum Precipitation for California.*

Danabasoglu, G., Lamarque, J. F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., … Strand, W. G. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*(2). https://doi.org/10.1029/2019MS001916

English, J. M., Turner, D. D., Alcott, T. I., Moninger, W. R., Bytheway, J. L., Cifelli, R., & Marquis, M. (2021). Evaluating Operational and Experimental HRRR Model Forecasts of Atmospheric River Events in California. *Weather and Forecasting*, *36*(6), 1925–1944. https://doi.org/10.1175/WAF-D-21-0081.1

Enzel, Y., Ely, L. L., House, P. K., Baker, V. R., & Webb, R. H. (1993). Paleoflood evidence for a natural upper bound to flood magnitudes in the Colorado River Basin. *Water Resources Research*, *29*(7), 2287–2297. https://doi.org/10.1029/93WR00411

Federal Energy Regulatory Commission (FERC), 2001. Determination of the probable maximum flood (Chap. VIII). In Engineering Guidelines for the Evaluation of Hydropower Projects. Washington (DC): United States Department of Energy, p. 121.

Fish, M. A., Wilson, A. M., & Ralph, F. M. (2019). Atmospheric River Families: Definition and Associated Synoptic Conditions. *Journal of Hydrometeorology*.

Gangrade, S., Kao, S. C., Naz, B. S., Rastogi, D., Ashfaq, M., Singh, N., & Preston, B. L. (2018). Sensitivity of Probable Maximum Flood in a Changing Environment. *Water Resources Research*. https://doi.org/10.1029/2017WR021987

Gangrade, S., Kao, S. C., Dullo, T. T., Kalyanapu, A. J., & Preston, B. L. (2019). Ensemble-Based Flood Vulnerability Assessment for Probable Maximum Flood in a Changing Environment. *Journal of Hydrology*, *576*, 342–355. https://doi.org/10.1016/J.JHYDROL.2019.06.027

Gershunov, A., Shulgina, T., Clemesha, R. E. S., Guirguis, K., Pierce, D. W., Dettinger, M. D., Lavers, D. A., Cayan, D. R., Polade, S. D., Kalansky, J., & Ralph, F. M. (2019). Precipitation regime change in Western North America: The role of Atmospheric Rivers. *Scientific Reports 2019 9:1*, *9*(1), 1–11. https://doi.org/10.1038/s41598-019-46169-w

Grell, G. A., Dudhia, J., & Stauffer, D. (1994). A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5) (No. NCAR/TN-398+STR). University Corporation for Atmospheric Research. doi:10.5065/D60Z716B

Grell, G. A., & Dévényi, D. (2002). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. Geophysical Research Letters, 29(14), 38-1-38–4. https://doi.org/10.1029/2002GL015311

Hansen, E., Fenn, D., Schreiner, L., Stodt, R., & Miller, J. (1988). *Hydrometeorological Report No. 55A: Probable Maximum Precipitation Estimates- United States Between the Continental Divide and the 103rd Meridian*.

Hansen, E., Fenn, D., Corrigan, P., Vogel, J., Schreiner, J., & Stodt, R. (1994). *Hydrometeorological Report No. 57: Probable Maximum Precipitation - Pacific Northwest States, Columbia River (including portions of Canada), Snake River and Pacific Coastal Drainages*.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/QJ.3803

Holland, G. J., Done, J., Bruyere, C., Cooper, C., & Suzuki-Parker, A. (2010). Model Investigations of the Effects of Climate Variability and Change on Future Gulf of Mexico Tropical Cyclone Activity. *Proceedings of the Annual Offshore Technology Conference*, *2*, 1659–1671. https://doi.org/10.4043/20690-MS

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. Monthly Weather Review, 134(9), 2318–2341. https://doi.org/10.1175/MWR3199.1

Ishida, K., Kavvas, M L, Asce, F., Jang, ; S, Chen, ; Z Q, Asce, A. M., Ohara, ; N, & Anderson, M. L. (2014). Physically Based Estimation of Maximum Precipitation over Three Watersheds in Northern California: Atmospheric Boundary Condition Shifting. *Journal of Hydrologic Engineering*, *20*(4), 04014052. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001026

Ishida, K., Kavvas, M. L., Jang, S., Chen, Z. Q., Ohara, N., & Anderson, M. L. (2015). Physically Based Estimation of Maximum Precipitation over Three Watersheds in Northern California: Relative Humidity Maximization Method. *Journal of Hydrologic Engineering*, *20*(10), 04015014. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001175

Ishida, K., Kavvas, M. L., & Jang, S. (2015). *Comparison of Performance on Watershed-Scale Precipitation between WRF and MM5*. https://doi.org/10.1061/9780784479162.095

Jankov, I., Bao, J. W., Neiman, P. J., Schultz, P. J., Yuan, H., & White, A. B. (2009). Evaluation and Comparison of Microphysical Algorithms in ARW-WRF Model Simulations of Atmospheric River Events Affecting the California Coast. *Journal of Hydrometeorology*, *10*(4), 847–870. https://doi.org/10.1175/2009JHM1059.1

Jiménez, P. A., Dudhia, J., González-Rouco, J. F., Navarro, J., Montávez, J. P., & García-Bustamante, E. (2012). A Revised Scheme for the WRF Surface Layer Formulation. Monthly Weather Review, 140(3), 898–918. https://doi.org/10.1175/MWR-D-11-00056.1

Kim, Y. K., Kim, S. M., & Tachikawa, Y. (2020). Analyzing Uncertainty in Probable Maximum Precipitation Estimation With Pseudoadiabatic Assumption. *Water Resources Research*, *56*(9), e2020WR027372. https://doi.org/10.1029/2020WR027372

Kingsmill, D. E., Neiman, P. J., Moore, B. J., Hughes, M., Yuter, S. E., & Ralph, F. M. (2013). Kinematic and Thermodynamic Structures of Sierra Barrier Jets and Overrunning Atmospheric Rivers during a Landfalling Winter Storm in Northern California. *Monthly Weather Review*. https://doi.org/10.1175/MWR-D-12-00277.1

Kitzmiller, D. H., Wu, W., Zhang, Z., Patrick, N., Tan, X., Kitzmiller, D. H., Wu, W., Zhang, Z., Patrick, N., & Tan, X. (2018). The Analysis of Record for Calibration: A High-Resolution Precipitation and Surface Weather Dataset for the United States. *AGUFM*, *2018*, H41H-06. https://ui.adsabs.harvard.edu/abs/2018AGUFM.H41H..06K/abstract

Knutti, R., & Sedláček, J. (2013). Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change*, *3*(4), 369–373. https://doi.org/10.1038/NCLIMATE1716

Krantz, W., Pierce, D., Goldenson, N., & Cayan, D. (2021). *Memorandum on Evaluating Global Climate Models for Studying Regional Climate Change in California*.

Kunkel, K. E., Karl, T. R., Easterling, D. R., Redmond, K., Young, J., Yin, X., Hennon, P., Karl, T. R., Easterling, D. R., Redmond, K., Young, J., Yin, X., & Hennon, P. (2013). Probable maximum precipitation and climate change. *Geophysical Research Letters*, *40*(7), 1402–1408. https://doi.org/10.1002/GRL.50334

Leung, R. L., & Qian, Y. (2009). Atmospheric rivers induced heavy precipitation and flooding in the western U.S. simulated by the WRF regional climate model. *Geophysical Research Letters*, *36*(3). https://doi.org/10.1029/2008GL036445

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., Maurer, E. P., & Lettenmaier, D. P. (2013). A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions. *Journal of Climate*, *26*(23), 9384–9392. https://doi.org/10.1175/JCLI-D-12-00508.1

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., Cayan, D. R., & Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950-2013. *Scientific Data*. https://doi.org/10.1038/sdata.2015.42

Lundquist, J., Hughes, M., Gutmann, E., & Kapnick, S. (2019). Our Skill in Modeling Mountain Rain and Snow is Bypassing the Skill of Our Observational Networks. *Bulletin of the American Meteorological Society*, *100*(12), 2473–2490. https://doi.org/10.1175/BAMS-D-19-0001.1

Mahoney, K. M., Jackson, D. L., Neiman, P., Hughes, M., Darby, L., Wick, G., White, A., Sukovich, E., & Cifelli, R. (2016). Understanding the Role of Atmospheric Rivers in Heavy Precipitation in the Southeast United States. *Monthly Weather Review*, *144*(4), 1617–1632. https://doi.org/10.1175/MWR-D-15-0279.1

Mahoney, K., McColl, C., Hultstrand, D. M., Kappel, W. D., McCormick, B., & Compo, G. P. (2021). Blasts from the past: Reimagining historical storms with model simulations to modernize dam safety and flood risk assessment. *Bulletin of the American Meteorological Society*, 1–35. https://doi.org/10.1175/BAMS-D-21-0133.1

Martin, A., Ralph, F. M., 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. *Journal of Hydrometeorology*, *19*(7), 1097–1113. https://doi.org/10.1175/JHM-D-17-0098.1

MGS. (2005). Stochastic Modeling of Extreme Floods on the American River at Folsom Dam Flood-Frequency Curve Extension. www.hec.usace.army.mil

Michaelis, A. C., Martin, A. C., Fish, M. A., Hecht, C. W., & Ralph, F. M. (2021). Modulation of Atmospheric Rivers by Mesoscale Frontal Waves and Latent Heating: Comparison of Two U.S. West Coast Events. *Monthly Weather Review*, *149*(8), 2755–2776. https://doi.org/10.1175/MWR-D-20-0364.1

Micovic, Z., Schaefer, M. G., & Taylor, G. H. (2015). Uncertainty analysis for Probable Maximum Precipitation estimates. *Journal of Hydrology*, *521*, 360–373. https://doi.org/10.1016/J.JHYDROL.2014.12.033

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. Journal of Geophysical Research: Atmospheres, 102(D14), 16663–16682. https://doi.org/10.1029/97JD00237

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line:

Comparison of One- and Two-Moment Schemes. Monthly Weather Review, 137(3), 991–1007. https://doi.org/10.1175/2008MWR2556.1

National Research Council, 1994. Estimating Bounds on Extreme Precipitation Events: A Brief Assessment. *Washington, DC: The National Academies Press*. https://doi.org/10.17226/9195.

Neiman, P. J., Hughes, M., Moore, B. J., Martin Ralph, F., & Sukovich, E. M. (2013). Sierra Barrier Jets, Atmospheric Rivers, and Precipitation Characteristics in Northern California: A Composite Perspective Based on a Network of Wind Profilers. *Monthly Weather Review*, *141*(12), 4211–4233. https://doi.org/10.1175/MWR-D-13-00112.1

Newman, A. J., Clark, M. P., Craig, J., Nijssen, B., Wood, A., Gutmann, E., Mizukami, N., Brekke, L., & Arnold, J. R. (2015). Gridded Ensemble Precipitation and Temperature Estimates for the Contiguous United States. *Journal of Hydrometeorology*, *16*(6), 2481–2500. https://doi.org/10.1175/JHM-D-15-0026.1

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., & Xia, Y. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. Journal of Geophysical Research, 116(D12), D12109. https://doi.org/10.1029/2010JD015139

O'Connor, J. E., Grant, G. E., & Costa, J. E. (2002). The geology and geography of floods. In P. House, R. Webb, V. Baker, & D. Levish (Eds.), *Ancient Floods Modern Hazards* (Vol. 5, pp. 359–385). https://www.webofscience.com/wos/woscc/full-record/WOS:000180349700021?SID=5B2wiCWrdGEcZA6Grq3

Ohara, N., Asce, A. M., Kavvas, ; M L, Asce, F., Kure, ; S, Chen, ; Z Q, Jang, ; S, Tan, E., & Asce, M. (2011). Physically Based Estimation of Maximum Precipitation over American

River Watershed, California. *Journal of Hydrologic Engineering*, *16*(4), 351–361. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000324

Ohara, N., Kavvas, M. L., Anderson, M. L., Chen, Z. Q., & Ishida, K. (2017). Characterization of extreme storm events using a numerical model-based precipitation maximization procedure in the feather, Yuba, and American River Watersheds in California. *Journal of Hydrometeorology*, *18*(5), 1413–1423. https://doi.org/10.1175/JHM-D-15-0232.1

Palmer, T., Buizza, R., Doblas-Reyes, F., Jung T, Leutbecher, M., Shutts, G., Steinheimer, M., & Weisheimer, A. (2009). *Stochastic parametrization and model uncertainty | ECMWF Technical Memoranda Number 598*. https://doi.org/0.21957/ps8gbwbdv

Rahimi, S., Huang, L., Goldenson, N., Risser, M., Feldman, D., Lebo, Z., Norris, J., Dennis, E., Thackeray, C., & Hall, A. (2023). Mean-state biases in CMIP6 GCMs are a predictor of precipitation biases in dynamical downscaling. *Geophysical Research Letters*.

Ralph, F. M., Rutz, J. J., Cordeira, J. M., Dettinger, M., Anderson, M., Reynolds, D., Schick, L. J., & Smallcomb, C. (2019). A scale to characterize the strength and impacts of atmospheric rivers. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/BAMS-D-18-0023.1

Reeves, H. D., Lin, Y. L., & Rotunno, R. (2008). Dynamic Forcing and Mesoscale Variability of Heavy Precipitation Events over the Sierra Nevada Mountains. *Monthly Weather Review*, *136*(1), 62–77. https://doi.org/10.1175/2007MWR2164.1

Risley, J. (2004). *Floods of November 1996 through January 1997 in the Umpqua River Basin, Oregon*. https://pubs.usgs.gov/fs/2004/3134/pdf/fs2004-3134.pdf

Rodgers, K. B., Lee, S. S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., Edwards, J., Kim, J. E., Simpson, I. R., Stein, K., Stuecker, M. F., Yamaguchi, R.,

Bódai, T., Chung, E. S., Huang, L., Kim, W. M., Lamarque, J. F., Lombardozzi, D. L., Wieder, W. R., & Yeager, S. G. (2021). Ubiquity of human-induced changes in climate variability. *Earth System Dynamics*, *12*(4), 1393–1411. https://doi.org/10.5194/ESD-12-1393-2021

Shutts, G. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, *131*(612), 3079–3102. https://doi.org/10.1256/QJ.04.106

Shutts, G., Leutbecher, M., Weisheimer, A., Stockdale, T., Isaksen, L., & Bonavita, M. (2011). *Representing model uncertainty: stochastic parametrizations at ECMWF | ECMWF Newsletter Number 129*. https://doi.org/10.21957/fbqmkhv7

Simpson, I. R., Bacmeister, J., Neale, R. B., Hannay, C., Gettelman, A., Garcia, R. R., Lauritzen, P. H., Marsh, D. R., Mills, M. J., Medeiros, B., & Richter, J. H. (2020). An Evaluation of the Large-Scale Atmospheric Circulation and Its Variability in CESM2 and Other CMIP Models. *Journal of Geophysical Research: Atmospheres*, *125*(13), e2020JD032835. https://doi.org/10.1029/2020JD032835

Skamarock, W. C., & Klemp, J. B. (2008). A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal of Computational Physics*, *227*(7), 3465–3485. https://doi.org/10.1016/J.JCP.2007.01.037

Smith, J. A., & Baeck, M. L. (2015). "Prophetic vision, vivid imagination": The 1927 Mississippi River flood. *Water Resources Research*, *51*(12), 9964–9994. https://doi.org/10.1002/2015WR017927

Su, Y., & Smith, J. A. (2021). An Atmospheric Water Balance Perspective on Extreme Rainfall Potential for the Contiguous US. *Water Resources Research*, *57*(4). https://doi.org/10.1029/2020WR028387

Su, Y., Smith, J. A., & Villarini, G. (2023). The Hydrometeorology of Extreme Floods in the Lower Mississippi River. *Journal of Hydrometeorology*, *24*(2), 203–219. https://doi.org/10.1175/JHM-D-22-0024.1

Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization. Monthly Weather Review, 136(12), 5095–5115. https://doi.org/10.1175/2008MWR2387.1

Tiedtke, M. (1989). A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models. Monthly Weather Review, 117(8), 1779–1800. https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2

Toride, K., Iseri, Y., Warner, M. D., Frans, C. D., Duren, A. M., England, J. F., & Kavvas, M. L. (2019). Model-Based Probable Maximum Precipitation Estimation: How to Estimate the Worst-Case Scenario Induced by Atmospheric Rivers? *Journal of Hydrometeorology*, *20*(12), 2383–2400. https://doi.org/10.1175/JHM-D-19-0039.1

Trinh, T., Iseri, Y., Diaz, A. J., Snider, E. D., Anderson, M. L., & Kavvas, M. L. (2021). Maximization of Historical Storm Events over Seven Watersheds in Central/Southern Sierra Nevada by Means of Atmospheric Boundary Condition Shifting and Relative Humidity Optimization Methods. *Journal of Hydrologic Engineering*, *27*(3), 04021051. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002159

USWB. (1961). *Hydrometeorological Report No. 36. Interim Report. Probable Maximum Precipitation in California.*

Waananen, A; Harris, D; Williams, R. (1965). *Floods of December 1964 and January 1965 in the Far Western States* (Issue January). https://pubs.usgs.gov/wsp/1866a/report.pdf

Warner, M. D., Mass, C. F., & Salatheé, E. P. (2012). Wintertime Extreme Precipitation Events along the Pacific Northwest Coast: Climatology and Synoptic Evolution. *Monthly Weather Review*, *140*(7). https://doi.org/10.1175/MWR-D-11-00197.1

World Meteorological Organization (2009) Manual for Estimation of Probable Maximum Precipitation, 3rd edition, WMO - No. 1045, Geneva, ISBN 978-92-63-11045-9

Yang, L., & Smith, J. (2018). Sensitivity of Extreme Rainfall to Atmospheric Moisture Content in the Arid/Semiarid Southwestern United States: Implications for Probable Maximum Precipitation Estimates. *Journal of Geophysical Research: Atmospheres*, *123*(3), 1638–1656. https://doi.org/10.1002/2017JD027850

Zhao, W., Smith, J. A., & Bradley, A. A. (1997). Numerical simulation of a heavy rainfall event during the PRE-STORM Experiment. *Water Resources Research*, *33*(4), 783–799. https://doi.org/10.1029/96WR03036