Geophysical Research Letters

RESEARCH LETTER
10.1029/2024GL109424

Special Collection:
Advancing Interpretable AI/ML Methods for Deeper Insights and Mechanistic Understanding in Earth Sciences: Beyond Predictive Capabilities

Rapid Inundation Mapping Using the US National Water Model, Satellite Observations, and a Convolutional Neural Network

Jonathan M. Frame[1,2], Tanya Nair[1], Veda Sunkara[1], Philip Popien[1], Subit Chakrabarti[1], Tyler Anderson[1], Nicholas R. Leach[1], Colin Doyle[1], Mitchell Thomas[1], and Beth Tellman[1]

[1]Floodbase, Brooklyn, NY, USA, [2]University of Alabama, Tuscaloosa, AL, USA

Abstract Rapid and accurate maps of floods across large domains, with high temporal resolution capturing event peaks, have applications for flood forecasting and resilience, damage assessment, and parametric insurance. Satellite imagery produces incomplete observations spatially and temporally, and hydrodynamic models require tradeoffs between computational efficiency and accuracy. We address these challenges with a novel flood model which predicts surface water area from the U.S. National Water Model using a convolutional neural network (NWM-CNN). We trained NWM-CNN on 780 flood events, at a 250 m resolution with an RMSE of 4.58% on held out validation geographies. We demonstrate NWM-CNN across California during the 2023 atmospheric rivers, comparing predictions against Sentinel-1 mapped flood observations. We compared historical predictions from 1979 to 2023 to flood damage reports in Sacramento County, California. Results show that NWM-CNN captures inundation extent better than the Height Above Nearest Drainage (HAND) approach (25%–36% RMSE, respectively).

Plain Language Summary We use machine learning (ML) to map floods quickly and accurately over large areas, which can help with predicting flooded extent, understanding impact, and aiding flood insurance and response. On their own, satellite images don't catch everything because they may be obscured or unavailable at the peak of a flood. Computer models that predict floods require a trade-off between speed, accuracy, and resolution. Our solution uses ML to learn from the U.S. National Water Model and satellite images from past floods to predict how much of an area will be covered in water. We demonstrate this on floods in California in 2023 caused by atmospheric rivers, and we looked back at floods in Sacramento County from 1979 to 2023. We compared our method to another commonly used model and found ours was more accurate, making it a promising tool for future flood mapping and response planning.

Key Points:

- Convolution neural networks (CNN) are suitable for rapid modeling of surface water dynamics for large-scale inundation mapping
- We deploy a CNN for continuous flood mapping across all of California during the devastating 2023 atmospheric river (AR) events
- Inundation extent across Sacramento is more accurately predicted with CNN than the Height Above Nearest Drainage (HAND)

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. M. Frame,
jmframe@ua.edu

Citation:

Frame, J. M., Nair, T., Sunkara, V., Popien, P., Chakrabarti, S., Anderson, T., et al. (2024). Rapid inundation mapping using the US National Water Model, satellite observations, and a convolutional neural network. *Geophysical Research Letters*, *51*, e2024GL109424. https://doi.org/10.1029/2024GL109424

Received 25 MAR 2024
Accepted 18 JUL 2024

Author Contributions:

Conceptualization: Jonathan M. Frame, Subit Chakrabarti, Beth Tellman
Data curation: Jonathan M. Frame, Tanya Nair, Veda Sunkara, Philip Popien, Tyler Anderson, Nicholas R. Leach

1. Introduction

Floods are primarily mapped using either hydrodynamic models or remote sensing observations. Satellite imagery produces incomplete inundation observations spatially and temporally, and hydrodynamic models suffer from epistemic and aleatory uncertainty and computational limitations. We address these challenges with a novel ML-based flood modeling approach using the U.S. National Water Model (NWM) (Cosgrove et al., 2024; Salas et al., 2018) to predict *satellite-observed inundation extents*. This model provides a spatially and temporally complete estimate of a NRT flood extent across the contiguous United States (CONUS) and a complete reanalysis from the period from 1979 to 2023. The primary objective and contribution of this paper is to introduce this novel approach to estimate surface water dynamics over large spatial and temporal domains, without the need for satellite imagery as an input. This is a first-of-its-kind for machine learning (ML) assisted flood models. We demonstrate our model on the California 2023 Atmospheric River (AR) flood events. We compare our model to the U.S. National Water Center's (NWC) current approach to flood inundation mapping, Height Above Nearest Drainage (HAND) (Aristizabal et al., 2023; Liu et al., 2018; Zheng et al., 2018). We then demonstrate a historical reanalysis of flooding across Sacramento, CA from 1979 through 2023.

1.1. Satellite Observations Are a Powerful but Incomplete Tool to Map Floods

Satellite images can produce accurate flood maps across large spatial domains (Tellman et al., 2021). Radar can detect surface water even when clouds are present (Zhao et al., 2021) while optical sensors image the earth daily at 5–500 m resolutions. Earth observations of flood inundation improve disaster response, aid and relief (Ho

Formal analysis: Jonathan M. Frame, Tanya Nair, Veda Sunkara, Tyler Anderson, Nicholas R. Leach, Mitchell Thomas
Funding acquisition: Beth Tellman
Investigation: Jonathan M. Frame, Tanya Nair, Veda Sunkara, Philip Popien, Tyler Anderson, Nicholas R. Leach, Colin Doyle
Methodology: Jonathan M. Frame, Tanya Nair, Veda Sunkara, Philip Popien, Nicholas R. Leach
Project administration: Subit Chakrabarti, Colin Doyle, Beth Tellman
Resources: Subit Chakrabarti, Tyler Anderson
Software: Jonathan M. Frame, Tanya Nair, Veda Sunkara, Philip Popien, Nicholas R. Leach
Supervision: Jonathan M. Frame, Philip Popien, Subit Chakrabarti, Colin Doyle, Beth Tellman
Validation: Jonathan M. Frame, Tanya Nair, Veda Sunkara, Nicholas R. Leach
Visualization: Jonathan M. Frame, Tanya Nair
Writing – original draft: Jonathan M. Frame, Beth Tellman
Writing – review & editing: Jonathan M. Frame, Tanya Nair, Veda Sunkara, Philip Popien, Subit Chakrabarti, Tyler Anderson, Nicholas R. Leach, Colin Doyle, Mitchell Thomas, Beth Tellman

et al., 2021; Schumann et al., 2018; Tellman et al., 2022). Satellite based flood observations are regularly used for hydraulic flood model intercomparison (Bernhofen et al., 2018, 2022; Trigg et al., 2016) and model validation (Molinari et al., 2019; P. Bates, 2023). ML has enabled accurate automated delineation of flood extent from satellite imagery (Bentivoglio et al., 2022; Bonafilia et al., 2020; Hänsch et al., 2022; Jakubik et al., 2023; Wieland et al., 2023). Yet satellites provide an incomplete observation of maximum flood extent (Bauer-Marschallinger et al., 2022; Du et al., 2018; Jensen & Mcdonald, 2019; Li et al., 2021; Shastry et al., 2023; Tulbure et al., 2022). Thus, approaches to fill in gaps between sensors to map peak inundation over large spatial and temporal domains are needed, which we offer here.

1.2. Large Spatial Domain Hydrology and Inundation Modeling Has Inadequate NRT Spatial Accuracy

Flood models based on surface water dynamics can provide predictions at peak inundation moments. These models can make gap-free flood forecasts, project flooding for the past with reanalysis data, or estimate inundation change in future climates. Setup and computational time to run simulations for NRT flood hazard assessment at large scales an ongoing challenge (van den Bout et al., 2023). Most operational or NRT hydrodynamic models estimate fluvial inundation, but real-world damaging flood events are often compound (Guan et al., 2023) or multi-form (Kruczkiewicz et al., 2022). Pluvial inundation is challenging for many modeling methods, particularly HAND, which requires a pre-defined nearby flowpath (Aristizabal et al., 2023). Hazard models are often based on scenarios taken from the historical record or future climate scenarios, not generated in NRT from current conditions (P. D. Bates et al., 2021). Continental or Global Scale models that operate in NRT (Alfieri et al., 2018) are typically discharge predictions (G. Nearing et al., 2024), or matched with previously processed inundation extents, not dynamic inundation responses (Dottori et al., 2017).

1.3. ML Improves Modeling of Surface Water Dynamics

Many hydrology problems, including operational applications, have been successfully addressed with ML (G. S. Nearing et al., 2020; Nevo et al., 2022; Frame, 2022). Bentivoglio et al. (2022) reviewed 58 ML flood modeling and inundation mapping methods, showing promising results, but emphasizing a need for real-time warnings and generalization to "unseen case studies." Our ML model is trained with a large sample of flood scenarios to learn to generate continuous dynamic flood maps, without satellite imagery as input, and generalizes to new locations and scenarios.

Convolutional neural networks (CNN) are well suited for flood mapping Zhou et al. (2022). Dasgupta et al. (2022) saw good results training a CNN to predict flooding on one event, but noted that "ways to incorporate the rainfall and antecedent catchment conditions upstream should be prioritized." Du et al. (2021) were able to predict satellite-like inundation using precipitation and satellite imagery, but noted performance was highly reliant on the timing of satellite capture relative to peak flood, and cannot generalize to new locations. Merging satellite observations with hydrodynamic models has been approached with data assimilation, but this also suffers from temporal availability of satellite data (Jafarzadegan et al., 2021). Our approach applies a CNN model trained with antecedent catchment conditions (from the NWM), on many satellite-observed flood events, under a wide variety of terrain conditions (Supplemental A1 in Supporting Information S1), to predict flood extents dynamically with no satellite input data at inference time. We refer to this as "NWM-CNN", and here we demonstrate NRT flood mapping on an unseen case study across a large domain.

NWM-CNN differs from surrogate modeling, which involves training an ML model to act like a process-based model. Surrogate models have been demonstrated to predict pluvial flooding from spatially highly variable rainfall events (Hofmann & Schüttrumpf, 2021; Zahura & Goodall, 2022). Guo et al. (2021) proposed a surrogate model for urban flood mapping, but warned of the limitation of surrogate models for generalization to new scenarios, and points to training on observational data as a solution. Our approach solves these problems. NWM-CNN makes predictions of observed flooding, as would be seen by a satellite with a complete spatial view at peak flood time, which is distinct from surrogate modeling approaches.

1.4. The Extreme 2023 Flood Season in California

California was hit by series of 31 ARs during the first half of the 2023 water year (Toohey, 2023). The highly intense rainfall of these events is a major source of flooding across California (Zou et al., 2023). There were 955 flood related reports logged by the National Weather Service. The estimated $5–7 billion US Dollars (USD) in

property losses was the most damaging flood event recorded in California. Less than a quarter of the losses ($0.5 to $1.5 billion estimated) was insured due to low NFIP (24%), and residential property take up rates (1%–8%) (Guy Carpenter, 2023).

We used the 2023 California Floods to demonstrate NWM-CNN because of its widespread spatial extent, and compound pluvial and fluvial causes of inundation. We compare NWM-CNN to the NWC flood inundation mapping (FIM) methodology (NWM-HAND). Our results demonstrate a promising approach to fill in gaps in the incomplete satellite record by leveraging widely available continental scale hydrologic model inputs from the NWM, showing the applicability of NWM-CNN for large regions for both NRT monitoring and historical reanalysis.

2. Methods

2.1. Model and Data

We summarize NWM-CNN here with more details in Supplemental A in Supporting Information S1. We use the NWM as the hydrological foundation for predicting the resulting surface water extent observable from Sentinel-2. NWM-CNN does not depend on satellite observations *as the input*, but is trained to predict surface water extent measured from optical satellite images. The model was trained with a loss function that only counts cloud-free pixels. From this procedure, the model learns to predict cloud-free surface water extent as seen by a satellite. The training set includes 780 flood events, with corresponding Sentinel-2 images from 2015 to 2022. These samples span a diverse spectrum of urbanization, surface water types, and geographies. We used a proprietary Floodbase CNN to identify surface water extent, which was trained on hand labels. The surface water identification model has a Critical Success Index (CSI; also known as Intersection over Union) of 76.3% (s.d. 3.3%) on never-before-flooded areas and 88.6% (s.d. 4.2%) on previously flooded areas (Supplemental Table A2 in Supporting Information S1). We use a CNN to take advantage of the spatial distribution of the NWM hydrologic states to predict the resulting the spatial distribution of surface water. Inputs to NWM-CNN include soil moisture and the mass state in the terrain router. We also include static inputs from three sources: a digital elevation model (Lehner et al., 2008), a global surface water raster (Pekel et al., 2016) and an annual agricultural land use map (USDA National Agricultural Statistics Service, 2023).

We trained a fully convolutional encoder-decoder network (Ronneberger et al., 2015) to predict the percent surface water area per pixel (PSWApp; as estimated by Sentinel-2) at 250 m resolution, and at the hour and date the satellite image was available. We aggregate 72 hours of terrain routing and soil moisture, and provide these as inputs to the model. All data, including surface water inundation, is resampled to a 250 m resolution. The model was trained in three folds of data, withholding a fourth fold as a held out test set, averaging a performance of 4.58 root mean square error (RMSE) and standard deviation (s.d) 2.07% across geographies (Supplemental Table A1 in Supporting Information S1).

2.1.1. U-Net Architecture

We use a U-Net architecture with an EfficientNet-B1 encoder. This version of a CNN allows features at different scales (through successive re-sampling) to be used for prediction of a class label at each pixel (Ronneberger et al., 2015), which is a desirable output for mapping surface water. This architecture makes an estimate of the value of each pixel in the output image from the whole of the input images.

Contracting Path (EfficientNet-B1 Encoder).

For each layer l of the encoder, context from the input features is propagated to successive feature maps that are downsampled through learnable convolution operations. Through the training process, the model learns appropriate weights for downsampled data to represent the surface water extent from hydrologic states from the input features. As many contracting architectures exist, our choice of the EfficientNet-B1 encoder is based on its ability to compress information in the model efficiently, reducing feature redundancy (Tan & Le, 2019). The contracting equations, based on an MBConv block, are described as follows:

$$
\begin{aligned}
C_{l,1} &= Swish(BatchNorm(x_l * W_{l,1})) \\
DwC_{l,2} &= Swish(BatchNorm((C_{l,1} * W_{l,2}))) \\
P_{l,3} &= AvgPool2d(C_{l,2}) \\
C_{l,4} &= Swish(P_{l,3} * W_{l,4}) \\
C_{l,5} &= Sigmoid(C_{l,4} * W_{l,5}) \\
M_{l,6} &= C_{l,2} * C_{l,5} \\
C_{l,7} &= BatchNorm(M_{l,6} * W_{l,7})
\end{aligned} \tag{1}
$$

where x_l is the input to layer l. $C_{l,i}$ are the feature maps from convolutional operations in the layer, $DwC_{l,2}$ are feature maps learned from a depthwise convolution. Batch Normalization, Swish, and Sigmoid functions are applied after convolutions stabilize training by facilitating gradients to propagate through the network M_l is the feature map multiplying with a channel attention mechanism $P_{l,3}$ through $C_{l,5}$ which facilitates the model to learn relationships between its different input layers (i.e., relationships between the dynamic and static inputs).

Expansive Path (Decoder).

For each layer l in the decoder, the feature map is upsampled by combining the corresponding map from the contracting path. The upsampling eventually results in features of the same resolution of the inputs. Skip connections provide information directly from the encoder to the convolutions in the decoder, by which the decoder not only has the compressed relevant features, but also has the higher resolution features. The expansion equations are:

$$
\begin{aligned}
U_l &= UpSample(B_{l-1}) \\
C'_{l,1} &= ReLU\left(U_l * W'_{l,1}\right) \\
C'_{l,2} &= ReLU\left(C'_{l,1} + C_{l-1,8} * W'_{l,2}\right)
\end{aligned} \tag{2}
$$

where $C'_{l,1}$ and $C'_{l,2}$ are feature maps in the decoder and + indicates the concatenation operation.

Final Output.

The output can be represented as:

$$
PSWApp = Clip\left(C'_{final}, 0, 1\right) \tag{3}
$$

where PSWApp is the resulting image of percent surface water area per pixel.

2.2. Anomalous Surface Water Area (ASWA)

NWM-CNN predicts percent surface water area, regardless if that extent is part of a permanent water body or a damaging flood. We consider the surface water across different spatial domains delineated by Hydrologic Unit Codes (HUC). We normalize the mean value across the HUC by subtracting out the lowest values during a defined time period within the individual HUC regions. This provides a means of comparing surface water across different boundaries with distinct surface water conditions. We refer to this as anomalous surface water area (ASWA).

Consider $PSWA$ as the percent of surface water across the entire prediction domain represented as a scalar (e.g., $\sum PSWApp$) and $PSWA_1$, $PSWA_2$, …, $PSWA_n$ as the corresponding time series, where $PSWA_t$ represents the tth image. The average pixel value of an image $PSWA_t$ is denoted as $\overline{PSWA}_t$, and the image with the minimum average pixel value is denoted as $PSWA_{\min}$. For our interest in flood characteristics, we specifically look at ASWA, or the amount of surface water above the defined baseline, $PSWA_{\min}$.

$$
ASWA_t = \overline{PSWA}_t - \overline{PSWA}_{\min} \tag{4}
$$

where $\overline{PSWA}_t$ is calculated as:

$$\overline{PSWA}_t = \frac{1}{N}\sum_{x,y} PSWApp_{(x,y)} \quad (5)$$

where N represents the total number of pixels in each image, (x, y) represents the coordinates of a pixel in the image, and $\overline{PSWA}_{\min}$ is calculated as:

$$\overline{PSWA}_{\min} = \min\left\{\overline{PSWA}_1, \overline{PSWA}_2, \ldots, \overline{PSWA}_n\right\} \quad (6)$$

2.3. Model Application

We applied NWM-CNN to the 2023 AR events across California. This application was chosen to demonstrate the temporal and spatial completeness of our flood model, as well as the accuracy of the model during peak flooding conditions.

2.3.1. Time Series Across California

ASWA is a spatial aggregate for HUC regions across California for the time period October 2022 through May 2023. We used the rasterstats (Perry, 2015) package in Python to run zonal statistics to calculate the mean PSWApp value across the HUC region (United States Geological Survey, 2023).

2.3.2. Spatial Mapping Example: Comparison Against Satellite Observations and Pixel-Wise Analysis

We demonstrate the ability of NWM-CNN to map surface water extent during a flood event by comparing to a satellite observation-based flood map in an analysis domain that spans two HUC 6 scale catchments in Sacramento. We use a composite map from Sentinel-1, a radar-based sensor not blocked by clouds, from January 6th–13th 2023 to capture maximum inundation. Our results are composed of the maximum PSWApp value across the date range to capture maximum inundation for the event. We chose the domain for the comparison as the bounding box encompassing the HUC 8 watersheds with the highest magnitude of anomalous flooded area. We use the bounding box around HUC 8 18020104 because that also happens to capture the majority of HUC 8 18020158. We used several imperfect metrics to compare pixels at 250 m resolution from NWM-HAND vs. NWM-CNN across our domain (a) square root of the mean squared error (RMSE), (b) precision, recall (Hit Rate), and FI scores, and (c) CSI, False Alarm Ratio (FAR), and Error Bias (EB). For RMSE, we re-sample the Sentinel-1 flood map from 10 m resolution to 250 m resolution using the mean pixel value, yielding the percent of 250 m pixels. RMSE is an ideal metric for the CNN model which produces a continuous output, but not for NWM-HAND, which produces a flood extent map. We also present results excluding pixels that were inundated prior to the specific AR event in our RMSE analysis. Precision, recall and F1 scores, as well as commonly use flood model performance metrics of CSI, EB, and FAR (P. D. Bates et al., 2021; Bernhofen et al., 2018) are ideal for NWM-HAND, which is a binary map. In order to calculate the binary metrics we included pixel values greater than one as "true" otherwise "false."

2.3.3. Historical Retrospective Run

We ran NWM-CNN for the NWM retrospective dates 1979 through 2022 for the Sacramento area, which has a high risk of flooding for a metropolitan area, with 29 severe flood events between 1950 and 2015 (Sacramento County Department of Water Resources, 2016). Annual water year peak ASWA was then qualitatively compared to historical flooding events in the Sacramento area, specifically a 30 km radius circle centered at Sacramento's city hall. Finally, we cross referenced these values with the damage estimates listed in the National Center for Environmental Information (NCEI) storm events database, which includes floods that occurred after the 1996 water year (Murphy, John D., 2021).

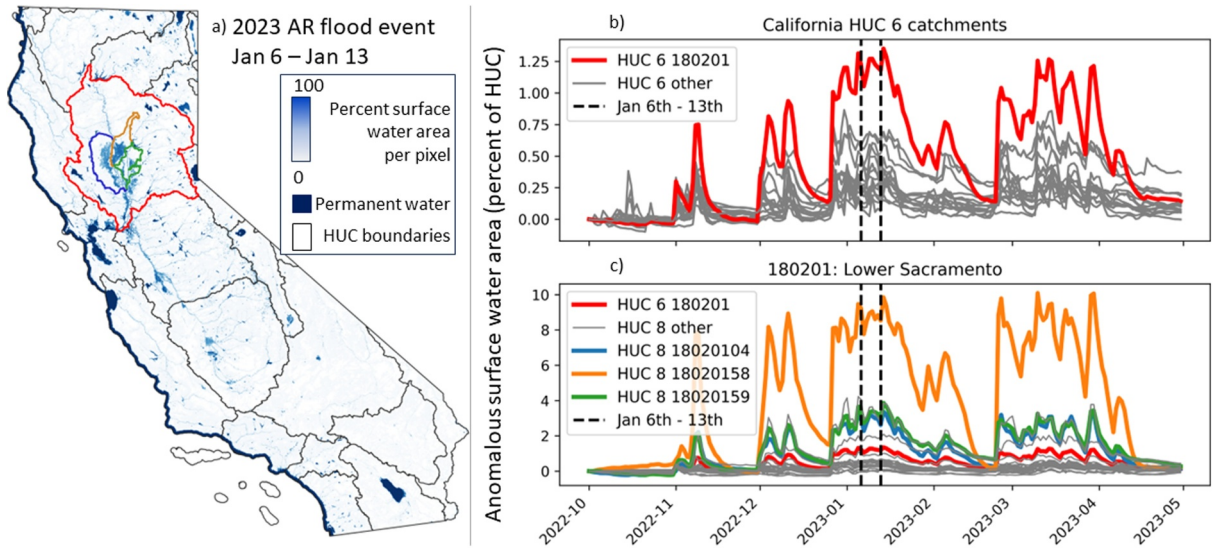

Figure 1. (a) California statewide snapshot of surface water predicted by NWM-CNN from a January 2023 AR event. (b) Summarized surface water areas across all of California at the HUC 6 scale. (c) Summarized surface water area at the HUC 8 scale across the Lower Sacramento catchment area, the HUC 6 catchment with the highest anomalous surface water response.

3. Results and Discussion

3.1. Continuous Monitoring of Surface Water Across California Throughout the 2023 AR Events

We present the surface water response across California during the 2023 AR event. Figure 1 shows a statewide snapshot of surface water predicted by NWM-CNN during one of the major AR events in January 2023. The predictions are gap-free at 250 m pixel resolution. Figure 1 also shows a time series of predictions aggregated to HUC catchments across the state. The HUC 6 catchment with the highest anomalous surface water response, by far, is the Lower Sacramento (180,201, up to 1.25% ASWA). Within 180,201 there is a wide variation of surface water responses at the HUC 8 scale, with the largest coming from 18,020,158 (up to 10% ASWA), which is highlighted along with 18,020,104 and 18,020,159 in Figure 1.

These results visually demonstrate the clustering of ARs that are relatively common across California (Slinskey et al., 2023). During these clustering of events, NRT monitoring (and forecasting) of potential flooding conditions becomes critical, as the sequence of events can (temporally) compound to produce unusually large magnitude flooding (Bowers et al., 2023). NWM-CNN is computationally capable of producing NRT and forecasted estimates of flooding at hourly time steps across CONUS.

3.2. Comparison Against Satellite Observations and Pixel-Wise Analysis

We present a snapshot of the modeled surface water extent produced across California during the January 2023 ARs, with a visual comparison against a satellite-observed map of the maximum inundated area observed in the state with the Sentinel-1 sensor inclusive of January 6th, 11th and 13th, 2023. Figure 2 shows these maps plotted in the Lower Sacramento River Basin. In subplot "d: Overlay" the satellite observations are plotted with transparency showing false negatives of the model (transparent red) the true positives of the model (purple) and the false negatives of the model (blue). False positive predictions are made in the upstream portions of this image, and false negative predictions are made in the downstream portion of this domain. Both the model and observation have about the same number of low value pixels (< 5 PSWApp). NWM-CNN over predicts the number of pixels between 5 and 50 PSWApp, but under predicts the number of pixels above 50 PSWApp. NWM-HAND method in black under predicts observed surface water.

Using the 250 m pixel values representing the PSWApp, we calculated an RMSE of 25% for NWM-CNN and 36% for NWM-HAND from within the analysis bounding box shown above in Figure 2. Table 1 shows the results excluding pixels that were shown to be inundated prior to the event, and pixels that result in "true negative" (where the observation and the models predict zero PSWA).

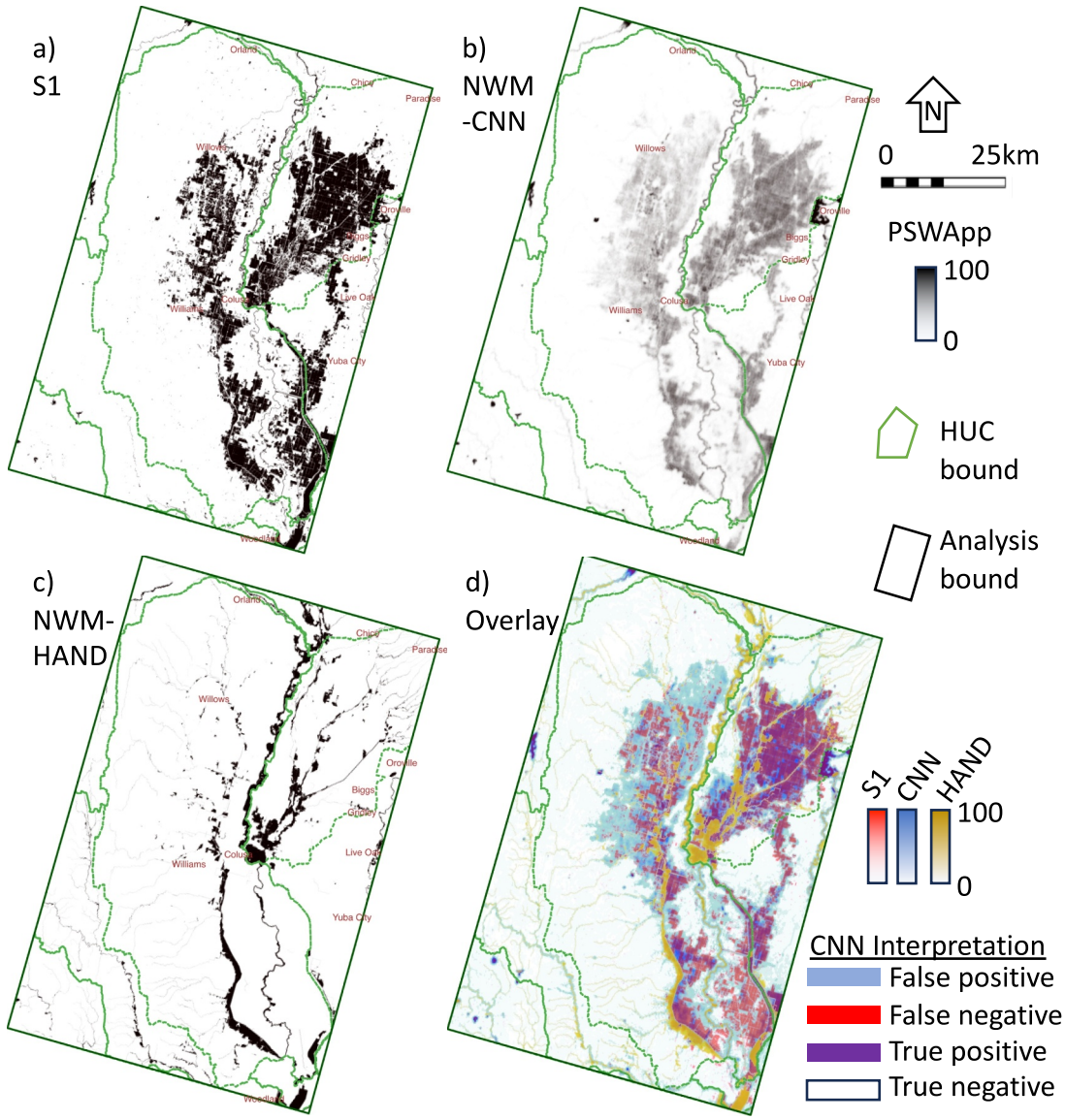

Figure 2. Comparison of the mapped model results. Individual greyscale maps shown in subplot (a) satellite-observation, (b) NWM-CNN and (c) NWM-HAND. And overlay of all three maps in subplot (d) where the blue represents our NWM-CNN, the satellite-observed surface water extent map is shown in transparent red, and the NWM-HAND results are shown in orange. With this color scheme, NWM-CNN false positives appear blue, true positives appear magenta and false negatives appear red.

CSI values of 0.7–0.8 are considered "good" for small, locally built flood models (P. D. Bates et al., 2021). For models that are making forecasts, without the assimilation of flood observations, NWM-CNN CSI value of 0.58 is reasonably good performance, especially considering this model provides rapid inundation maps from across CONUS with no data collection overhead, low computational cost and no fine-tuning required. For instance, Wing et al. (2019) report a CSI value of 0.57 for their model applied to Houston, TX, during Hurricane Harvey forced with NWM streamflow forecasts. The NWM-CNN CSI could be as high as 0.66 for this event, if the threshold of PSWA is optimized to 5% instead of held at zero (see sensitivity analysis, Supplemental C1 in Supporting Information S1). NWM-CNN has a relatively high EB, but a relatively low FAR. The NWM-CNN tends to overestimate extent, but underestimate individual pixel values. This means that while it predicts many events, a good portion of these predictions are indeed correct.

NWM-CNN outperforms the NWM-HAND method, and has closer to the CSI metrics for the 100-year flood plain reported from Fathom's US Flood model validation test in Iowa (CSI: 0.84). P. D. Bates et al. (2021) model

Table 1
Model Performance Statistics for Sacramento During January 23 Atmospheric River Event

Metric	NWM-CNN	NWM-HAND
RMSE All pixels	25%	36%
RMSE Ignoring pre-event water	21%	28%
RMSE Ignoring pre-event water and dry	23%	60%
Precision	0.60	0.45
Recall	0.94	0.25
F1	0.73	0.32
Critical success index	0.58	0.19
False Alarm Ratio	0.40	0.55
Error Bias	1.57	0.56

accounts for local infrastructure directly in their model architecture, which is not easily scalable to the large domain for which NWM-CNN was designed to run in NRT. While direct comparisons are elusive given many flood model evaluations report CSIs for return periods outputs (e.g., (P. D. Bates et al., 2021; Bernhofen et al., 2018; Trigg et al., 2016) and not discrete events (and perhaps, not a good metric for continuous data in NWM-CNN), we conclude the CSI for a NRT model (the NWM-CNN presented here) over a large area is performing reasonably well, but with room for improvement. Fine tuning the threshold for distinguishing "flood" versus "Not Flood" from NWM-CNN PSWApp values in either individual pixels or in specific regions is recommended with further analysis and consideration of local conditions (see Supplemental C1 in Supporting Information S1).

These results demonstrate a computationally efficient and reliable Flood Inundation Mapping (FIM) product that is directly informed by the NWM. At the time of this writing in 2024, NWS is operationalizing a FIM product based on NWM-HAND (Glaudemans, 2023), available in four states (Texas, Louisiana, New York and Pennsylvania). Further investment is being made to expand Flood Inundation Mapping services nation-wide (National Oceanic and Atmospheric Administration, 2023a, 2023b). Improvements to these flood mapping efforts could be made using ML (e.g., the CNN method proposed here) over current HAND approaches.

3.3. Retrospective Analysis of Flood History

Figure 3 shows annual maximum ASWA (Anomalous Surface Water Area) (%) across the Sacramento analysis domain for the complete retrospective period of the NWM (1980–2022). Also included on this plot is damage data from NCEI, plotted from 1997 onward. Most years (90%) with damages above zero correspond to a maximum ASWA over the median (1.8%), with 2016 as a notable exception. Most years (9 out of 10) where NWM-CNN predicted a higher than average (median) ASWA also corresponds to a year with flood damages, with 2022 as a notable exception. NWM-CNN predicted 4 years with above average (median) ASWA that do not correspond to NCEI damage claims, possibly indicating a tenancy for overestimation. Of the five highest water years predicted

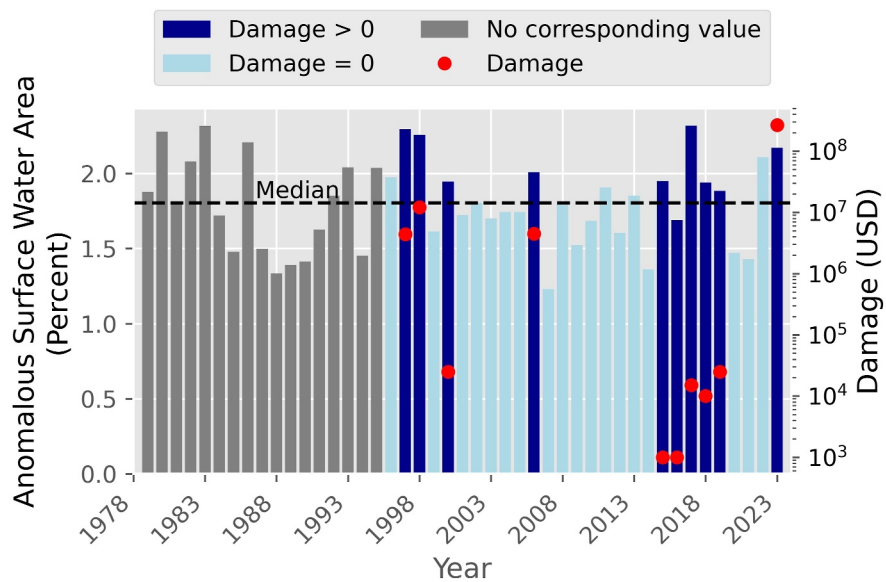

Figure 3. Annual (water year) maximum anomalous surface water area (%) across the Sacramento analysis domain. Gray colored bars from 1979 to 1996 do not have corresponding data in the Storm Events Database, while blue bars from 1996 onward were references against that database for Sacramento County, and include the total of estimated property and crop damage.

by NWM-CNN from 1987 onwards (> 3.5% ASWA), when damage data is available, four are the highest estimated damage events from NCEI (all exceeding $1 M USD, 1997, 1998, 2006, and 2023).

Historically, damaging flooding events include (1980, 1982 and 1983) (Sacramento County Department of Water Resources, 2016) and (1986, 1995, 1997, 2006) (Sacramento County, Accessed in 2023). The flood of 1986 is reported as one of the most severe events, and even though NWM-CNN predicts a high flood year, this result is likely an underestimate, as levee failure caused major flooding (Sacramento County Department of Water Resources, 2016), which cannot be captured by NWM-CNN. Peak annual ASWA is highest in 2017, which is the result of a series of ARs which struck California in January and February 2017 (California Nevada River Forecast Center, 2017), although 2017 corresponds to a low estimate of property and crop damage.

4. Conclusion

CNN-based models are well suited to fuse satellite imagery and dynamic hydrological models for gap-free rapid mapping of flooding over large spatial and temporal domains. Our model (NWM-CNN) is trained to predict the flood characteristics that are observable by satellite images from the relatively high-resolution gridded state values from the NWM. The limitation of the model is that errors or biases in satellite-based surface water observations will propagate and be learned by the model, but the benefit is that since satellite images are not used as a dynamic input, the model does not suffer from optical-obscurities or low revisit times normally plaguing satellite-based inundation mapping. Critically, this means NWM-CNN captures peak flooding that satellite sensors, except in extremely rare cases, will inevitably miss. NWM-CNN makes predictions that spatially match with a test satellite image, but pixel-by-pixel the predictions tend to under-represent the higher magnitude values. The visual results shown in Figure 2 show a generally good spatial correspondence between the model and satellite observations. NWM-CNN RMSE of 25% indicates a reasonable prediction, as compared to an NWM-HAND RMSE of 36%.

Future work is ongoing to improve NWM-CNN. Here we are demonstrating results with the minimally sufficient input data, with two dynamic inputs and three static inputs. Streamflow forecasting models, for instance, have been shown to make the best predictions with 14 dynamic inputs and dozens of static inputs (G. Nearing et al., 2024). Additional dynamic inputs could improve the timing and magnitude of the flood signal by incorporating streamflow and dynamic satellite inputs with higher resolution sensors. Additional static inputs could improve the spatial distribution of flood water, and including time varying inputs could account for non-stationary conditions. Future research aims to develop an approach that scales globally beyond CONUS.

The rapid run time over large spatial and temporal scale, and the gap-free predictions, mean NWM-CNN is useful in a variety of applications. NWM-CNN is also suited for short-term ensemble forecasting, matching the forecast times of the NWM, because it can produce inundation maps using NWM inputs. The model is ideal for index-based or parametric insurance applications, because it can produce a long and consistent time series (from 1979) to price an insurance product and a NRT output to serve as a trigger or strike. Ultimately, NWM-CNN demonstrates that the role of satellite data in inundation mapping needs to move beyond mere calibration, validation, parameterization, or even data assimilation with physically based inundation models. ML effectively leverages both the benefits of satellite observations and continuity of dynamic hydrologic states variables to complement each other and overcome the weakness inherent in each.

Data Availability Statement

Data are provided at HydroShare: https://www.hydroshare.org/resource/8b76906c4b604c458fbcb5ea7c8c0be7/ (Frame, 2024a). Data are available for use under a Creative Commons Attribution Non Commercial 4.0 International license. Analysis code for results presented in this paper is available from Zenodo Under the following DOI: 10.5281/zendo.13153247 (Frame, 2024b). Height Above Nearest Drainage results are generated with public data from the continental flood inundation mapping (CFIM) framework (Liu et al., 2020). The NWM-CNN, as well are the CNNs used to derive Sentinel maps, are proprietary models by Floodbase and are not available for use, except in specific institutions for academic purposes for collaboration or mutual benefit, email support@floodbase.com for inquiries.

Acknowledgments
This study was funded by Floodbase. Brian Cosgrove aided this study by answering many questions about the National Water Model data and retrospective runs.

References

Alfieri, L., Cohen, S., Galantowicz, J., Schumann, G. J.-P., Trigg, M. A., Zsoter, E., et al. (2018). A global network for operational flood risk reduction. *Environmental Science & Policy*, *84*, 149–158. Retrieved 2020-03-11, from https://doi.org/10.1016/j.envsci.2018.03.014

Aristizabal, F., Salas, F., Petrochenkov, G., Grout, T., Avant, B., Bates, B., et al. (2023). Extending height above nearest drainage to model multiple fluvial sources in flood inundation mapping applications for the u.s. national water model. *Water Resources Research*, *59*(5). https://doi.org/10.1029/2022WR032039

Bates, P. (2023). Fundamental limits to flood inundation modelling. *Nature Water*, *10*(47), 566–567. https://doi.org/10.1038/s44221-023-00106-4

Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., et al. (2021). Combined modeling of US fluvial, pluvial, and coastal flood hazard under current and future climates. *Water Resources Research*, *57*(2). Retrieved 2021-05-13, from. https://doi.org/10.1029/2020WR028673

Bauer-Marschallinger, B., Cao, S., Tupas, M. E., Roth, F., Navacchi, C., Melzer, T., et al. (2022). Satellite-based flood mapping through bayesian inference from a sentinel-1 sar datacube. *Remote Sensing*, *15*, 3673. Retrieved from https://doi.org/10.3390/rs14153673

Bentivoglio, R., Isufi, E., Jonkman, S. N., & Taormina, R. (2022). Deep learning methods for flood mapping: A review of existing applications and future research directions. *Hydrology and Earth System Sciences*, *26*(16), 4345–4378. Retrieved from https://doi.org/10.5194/hess-26-4345-2022

Bernhofen, M. V., Cooper, S., Trigg, M., Mdee, A., Carr, A., Bhave, A., et al. (2022). The role of global data sets for riverine flood risk management at national scales. *Water Resources Research*, *58*(4). Retrieved 2022-04-08, from. https://doi.org/10.1029/2021WR031555

Bernhofen, M. V., Whyman, C., Trigg, M. A., Sleigh, P. A., Smith, A. M., Sampson, C. C., et al. (2018). A first collective validation of global fluvial flood models for major floods in Nigeria and Mozambique. *Environmental Research Letters*, *13*(10), 104007. Retrieved 2018-11-06, from. https://doi.org/10.1088/1748-9326/aae014

Bonafilia, D., Tellman, B., Anderson, T., & Issenberg, E. (2020). Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. In *The IEEE/CVF conference on computer vision and pattern recognition (CVPR) workshops* (p. 11). Retrieved from https://doi.org/10.1109/CVPRW50498.2020.00113

Bowers, C., Serafin, K. A., Tseng, K.-C., & Baker, J. W. (2023). *Atmospheric river sequences as indicators of hydrologic hazard in present and future climates*. Authorea Preprints.

California Nevada River Forecast Center. (2017). Heavy precipitation events California and northern Nevada january and february 2017. Retrieved from https://www.cnrfc.noaa.gov/storm_summaries/janfeb2017storms.php. Accessed November 2023.

Cosgrove, B., Gochis, D., Flowers, T., Dugger, A., Ogden, F., Graziano, T., et al. (2024). NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling. *JAWRA Journal of the American Water Resources Association*, *60*(2), 247–272. Retrieved from https://doi.org/10.1111/1752-1688.13184

Dasgupta, A., Hybbeneth, L., & Waske, B. (2022). Towards daily high-resolution inundation observations using deep learning and eo. *arXiv preprint arXiv*, *2208*, 09135.

Dottori, F., Kalas, M., Salamon, P., Bianchi, A., Alfieri, L., & Feyen, L. (2017). An operational procedure for rapid flood risk assessment in europe.

Du, J., Kimball, J. S., Galantowicz, J., Kim, S.-B., Chan, S. K., Reichle, R., et al. (2018). Assessing global surface water inundation dynamics using combined satellite information from smap, amsr2 and landsat. *Remote Sensing of Environment*, *213*, 1–17. https://doi.org/10.1016/j.rse.2018.04.054

Du, J., Kimball, J. S., Sheffield, J., Pan, M., Fisher, C. K., Beck, H. E., & Wood, E. F. (2021). Satellite flood inundation assessment and forecast using smap and landsat. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 6707–6715. https://doi.org/10.1109/JSTARS.2021.3092340

Frame, J. M. (2022). Deep learning for operational streamflow forecasts a long short-term memory network rainfall-runoff module for the u.s. national water model. PhD Dissertation, University of Alabama, Department of Geologic Sciences.

Frame, J. M. (2024a). *Data: Model results for California ar 2023 nwmcnn*. HydroShare. Retrieved from http://www.hydroshare.org/resource/8b76906c4b604c458fbcb5ea7c8c0be7

Frame, J. M. (2024b). Zenodo. Retrieved from https://doi.org/10.5281/zenodo.13153247. Software: Analysis code for California ar 2023 nwmcnn

Glaudemans, M. (2023). Updated:. *Soliciting comments on experimental flood inundation mapping (fim) services through september 30, 2024*. National Weather Service Headquarters. Retrieved from https://viewer.weather.noaa.gov/water(PublicInformationStatement23-55Updated)

Guan, X., Vorogushyn, S., Apel, H., & Merz, B. (2023). Assessing compound pluvial-fluvial flooding: Research status and ways forward. *Water Security*, *19*(February), 100136. Retrieved from https://doi.org/10.1016/j.wasec.2023.100136

Guo, Z., Leitão, J. P., Simões, N. E., & Moosavi, V. (2021). Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks. *Journal of Flood Risk Management*, *14*(1). https://doi.org/10.1111/jfr3.12684

Guy Carpenter. (2023). Post event report: California atmospheric river. Retrieved from https://www.guycarp.com/insights/2023/02/California-AR-02-08.html

Hänsch, R., Arndt, J., Lunga, D., Gibb, M., Pedelose, T., Boedihardjo, A., et al. (2022). Spacenet 8-the detection of flooded roads and buildings. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 1472–1480).

Ho, J. C., Tellman, B., Vu, W., Bienvenu, J. D., N'diaye, P. I., Weber, S., et al. (2021). From cloud to refugee camp: A satellite-based flood analytics case-study in Congo-brazzaville. In G. J.-P. Schumann (Ed.), *Earth observation for flood applications* (pp. 131–145). Elsevier.

Hofmann, J., & Schüttrumpf, H. (2021). Floodgan: Using deep adversarial learning to predict pluvial flooding in real time. *Water (Switzerland)*, *13*(16), 2255. https://doi.org/10.3390/w13162255

Jafarzadegan, K., Abbaszadeh, P., & Moradkhani, H. (2021). Sequential data assimilation for real-time probabilistic flood inundation mapping. *Hydrology and Earth System Sciences*, *25*(9), 4995–5011. https://doi.org/10.5194/hess-25-4995-2021

Jakubik, J., Chu, L., Fraccaro, P., Gomes, C., Nyirjesy, G., Bangalore, R., et al. (2023). Prithvi-100M. https://doi.org/10.57967/hf/0952

Jensen, K., & Mcdonald, K. (2019). Surface water microwave product series version 3: A near-real time and 25-year historical global inundated area fraction time series from active and passive microwave remote sensing. *IEEE Geoscience and Remote Sensing Letters*, *16*(9), 1402–1406. https://doi.org/10.1109/lgrs.2019.2898779

Kruczkiewicz, A., Cian, F., Monasterolo, I., Di Baldassarre, G., Caldas, A., Royz, M., et al. (2022). Multiform flood risk in a rapidly changing world: What we do not do, what we should and why it matters. *Environmental Research Letters*, *17*(8), 081001. Retrieved 2023-09-28, from. https://doi.org/10.1088/1748-9326/ac7ed9

Lehner, B., Verdin, K., & Jarvis, A. (2008). New global hydrography derived from spaceborne elevation data. *Eos, Transactions, American Geophysical Union*, *89*(10), 93–104. https://doi.org/10.1029/2008eo100001

Li, W., Yang, C., Peng, Y., & Zhang, X. (2021). A multi-cooperative deep convolutional neural network for spatiotemporal satellite image fusion. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 10174–10188. https://doi.org/10.1109/JSTARS.2021.3113163

Liu, Y., Maidment, D., Tarboton, D., Zheng, X., & Wang, S. (2018). A cybergis integration and computation framework for high-resolution continental-scale flood inundation mapping. *JAWRA Journal of the American Water Resources Association*, *54*(4), 770–784. https://doi.org/10.1111/1752-1688.12660

Liu, Y., Tarboton, D., & Maidment, D. (2020). version 0.21 (20200601). Oak Ridge National Laboratory Leadership Computing Facility. Retrieved from https://doi.org/10.13139/ORNLNCCS/1630903 Created: 5/27/2020, 6:47:37 AM; Published: 5/27/2020, 11:15:25 AM. Height above nearest drainage (hand) and hydraulic property table for conus

Molinari, D., De Bruijn, K. M., Castillo-Rodríguez, J. T., Aronica, G. T., & Bouwer, L. M. (2019). Validation of flood risk models: Current practice and possible improvements. *International Journal of Disaster Risk Reduction*, *33*(May 2018), 441–448. Retrieved from https://doi.org/10.1016/j.ijdrr.2018.10.022

Murphy, J. D. (2021). Storm data preparation: National weather service instruction 10-1605. [Computer software manual]. Retrieved from https://www.nws.noaa.gov/directives/sym/pd01016005curr.pdf. Accessed November 2023.

National Oceanic and Atmospheric Administration. (2023a). *Biden-harris administration announces $80 million through investing in America agenda to improve flood prediction capabilities*. NOAA Communications. Retrieved from https://www.noaa.gov/news-release/biden-harris-administration-announces-80-million-to-improve-water-predication-capabilities. Accessed: December 2023.

National Oceanic and Atmospheric Administration. (2023b). Bipartisan infrastructure law. Retrieved from https://www.noaa.gov/infrastructure-law. Accessed December 2023.

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., et al. (2024). Global prediction of extreme floods in ungauged watersheds. *Nature*, *627*(8004), 559–563. Retrieved from https://doi.org/10.1038/s41586-024-07145-1

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., et al. (2020). What role does hydrological science play in the age of machine learning? *Water Resources Research*, *57*(3). Retrieved from https://doi.org/10.1029/2020wr028091

Nevo, S., Morin, E., Rosenthal, A. G., Metzger, A., Barshai, C., Weitzner, D., et al. (2022). Flood forecasting with machine learning models in an operational framework. *Hydrology and Earth System Sciences*, *26*(15), 4013–4032. Retrieved from https://doi.org/10.5194/hess-26-4013-2022

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes.

Perry, M. T. (2015). rasterstats: Geospatial raster summary statistics in Python. Retrieved from https://github.com/perrygeo/python-rasterstats. Accessed 02 11 2023.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 234–241).

Sacramento County. (2023). Region's flooding history. Retrieved from https://waterresources.saccounty.gov/stormready/Pages/Region%27s-Flooding-History.aspx

Sacramento County Department of Water Resources. (2016). 2016 sacramento countywide local hazard mitigation plan update. Retrieved from https://waterresources.saccounty.gov/stormready/Pages/Local-Hazard-Mititagtion-Report.aspx. Accessed November 2023.

Salas, F. R., Somos-Valenzuela, M. A., Dugger, A., Maidment, D. R., Gochis, D. J., David, C. H., et al. (2018). Towards real-time continental scale streamflow simulation in continuous and discrete space. *Journal of the American Water Resources Association*, *54*(1), 7–27. https://doi.org/10.1111/1752-1688.12586

Schumann, G., Brakenridge, G., Kettner, A., Kashif, R., & Niebuhr, E. (2018). Assisting flood disaster response with earth observation data and products. *A critical assessment*, *10*(8), 1230. Retrieved 2018-10-07, from. https://doi.org/10.3390/rs10081230

Shastry, A., Carter, E., Coltin, B., Sleeter, R., McMichael, S., & Eggleston, J. (2023). Mapping floods from remote sensing data and quantifying the effects of surface obstruction by clouds and vegetation. *Remote Sensing of Environment*, *291*, 113556. https://doi.org/10.1016/j.rse.2023.113556

Slinskey, E. A., Hall, A., Goldenson, N., Loikith, P. C., & Norris, J. (2023). Subseasonal clustering of atmospheric rivers over the western United States. *Journal of Geophysical Research: Atmospheres*, *128*(22), e2023JD038833. https://doi.org/10.1029/2023jd038833

Tan, M., & Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. Retrieved from https://arxiv.org/abs/1905.11946

Tellman, B., Lall, U., Islam, A. S., & Bhuyan, M. A. (2022). Regional index insurance using satellite-based fractional flooded area. *Earth's Future*, *10*(3), e2021EF002418. https://doi.org/10.1029/2021ef002418

Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., et al. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, *596*(7870), 80–86. Retrieved from https://doi.org/10.1038/s41586-021-03695-w

Toohey, G. (2023). Volcano? Climate change? Bad luck? Why California was hit with 31 atmospheric river storms. *Los Angeles Times*. Retrieved from https://www.latimes.com/california/story/2023-04-11/californias-wild-winter-of-atmospheric-rivers. Accessed: 21 08 2023.

Trigg, M., Birch, C., Neal, J., Bates, P., Smith, A., Sampson, C., et al. (2016). The credibility challenge for global fluvial flood risk analysis. *Environmental Research Letters*, *11*(9), 094014. https://doi.org/10.1088/1748-9326/11/9/094014

Tulbure, M. G., Broich, M., Perin, V., Gaines, M., Ju, J., Stehman, S. V., et al. (2022). Can we detect more ephemeral floods with higher density harmonized landsat sentinel 2 data compared to landsat 8 alone? *ISPRS Journal of Photogrammetry and Remote Sensing*, *185*, 232–246. Retrieved from https://doi.org/10.1016/j.isprsjprs.2022.01.021

United States Geological Survey. (2023). Watershed boundary dataset. Retrieved 2023-11, from https://www.usgs.gov/national-hydrography/watershed-boundary-dataset (Accessed November 2023)

USDA National Agricultural Statistics Service. (2023). Published crop-specific data layer. Retrieved from https://croplandcros.scinet.usda.gov/. USDA-NASS. Accessed: 2023; Verified: 2023.

van den Bout, B., Jetten, V., van Westen, C., & Lombardo, L. (2023). A breakthrough in fast flood simulation. *Environmental Modelling & Software*, *168*, 105787. Retrieved from https://doi.org/10.1016/j.envsoft.2023.105787

Wieland, M., Martinis, S., Kiefl, R., & Gstaiger, V. (2023). Semantic segmentation of water bodies in very high-resolution satellite and aerial images. *Remote Sensing of Environment*, *287*, 113452. Retrieved from https://doi.org/10.1016/j.rse.2023.113452

Wing, O. E., Sampson, C. C., Bates, P. D., Quinn, N., Smith, A. M., & Neal, J. C. (2019). A flood inundation forecast of hurricane harvey using a continental-scale 2d hydrodynamic model. *Journal of Hydrology X*, *4*, 100039. Retrieved from https://doi.org/10.1016/j.hydroa.2019.100039

Zahura, F. T., & Goodall, J. L. (2022). Predicting combined tidal and pluvial flood inundation using a machine learning surrogate model. *Journal of Hydrology: Regional Studies*, *41*, 101087. Retrieved from https://doi.org/10.1016/j.ejrh.2022.101087

Zhao, M., Olsen, P. A., & Chandra, R. (2021). Seeing through clouds in satellite images. arxiv.org. Retrieved from http://arxiv.org/abs/2106.08408

Zheng, X., Tarboton, D., Maidment, D., Liu, Y., & Passalacqua, P. (2018). River channel geometry and rating curve estimation using height above the nearest drainage. *JAWRA Journal of the American Water Resources Association*, *54*(4), 785–806. https://doi.org/10.1111/1752-1688.12661

Zhou, Y., Wu, W., Nathan, R., & Wang, Q. J. (2022). Deep learning-based rapid flood inundation modeling for flat floodplains with complex flow paths. *Water Resources Research*, *58*(12). https://doi.org/10.1029/2022WR033214

Zou, X., Cordeira, J. M., Bartlett, S. M., Kawzenuk, B., Roj, S., Castellano, C. M., et al. (2023). Mesoscale and synoptic scale analysis of narrow cold frontal rainband during a landfalling atmospheric river in California during january 2021. Retrieved from https://doi.org/10.22541/essoar.168677226.69319241/v1