Rapid inundation mapping using the US National Water Model, satellite observations, and a convolutional neural network

Jonathan M. Frame[1], 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], Beth Tellman[1]

[1]Floodbase, Brooklyn, NY

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).

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 250m 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. Historically, we compared the data from 1979-2023 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% to 36% RMSE, respectively).

Plain Language Summary

We use machine learning 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 can miss parts of the flood or aren't available at the peak of a flood. Computer models that predict floods require a trade-off between speed, accuracy and resolution. Our solution uses a machine learning method to combine satellite images and data from the U.S. National Water Model that learns 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 when 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.

1 Introduction

Floods impact more people than any other hazard and economic loss from flood damage is increasing (Allen et al., 2018). Flood losses globally were 82 billion US Dollars (USD) in 2021, and another 50 billion USD in 2022 (Bevere & Finucane, 2022). Accurate knowledge of flood extent for ongoing and historical events helps facilitate climate adaptation in flood-prone communities by enabling near real-time (NRT) disaster monitoring to support response and relief during extreme events, and financial protection such as insurance to recover from them.

Floods are primarily mapped using one of two approaches: hydrodynamic models or through remote sensing observations. Both fail to capture maximum inundation extents accurately for distinct reasons; satellite imagery produces incomplete inundation observations spatially and temporally, and hydrodynamic models suffer from tradeoffs in computational efficiency and accuracy. We address these challenges with a novel flood modeling strategy which trains outputs from a hydrologic model (U.S. National Water Model; NWM) (Salas et al., 2018; Cosgrove et al., 2024) on satellite observed inundation extents represented as percent surface water area that can substitute direct satellite observations hourly across the CONtiguous United States (CONUS) from 1979-2023. The primary objective and contribution of this paper is to present a novel approach to estimate surface water dynamics over large spatial and temporal domains, which we demonstrate using the California 2023 (AR) Flood event, with a special focus on Sacramento. In Sacramento, we compare our model to the U.S. National Water Center's (NWC) current approach to flood inundation mapping, Height Above Nearest Drainage (Aristizabal et al., 2023; Liu et al., 2018; Zheng et al., 2018).

1.1 Satellite observations are a powerful but incomplete tool to map floods

Satellite images are used to produce accurate flood maps across large spatial domains, and at high spatial and temporal resolutions (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 5-500m resolutions. Earth observations of flood inundation improve disaster response (Schumann et al., 2018), rapid aid assessment and financing from assistance relief programs (Ho et al., 2021), and access to financial recovery through insurance (Tellman et al., 2022). The International Charter: Space and Major Disasters (`https://disasterscharter.org`; accessed November 2023) enables governments and satellite providers to rapidly map floods and share data for major global events to improve flood response. Satellite based flood observations are regularly used for hydraulic flood model intercomparison (Trigg et al., 2016; Bernhofen et al., 2022, 2018) and model validation (Molinari et al., 2019; P. Bates, 2023). Machine learning has enabled accurate automated delineation of flood extent from satellite imagery (Bonafilia et al., 2020; Wieland et al., 2023; Hänsch et al., 2022; Jakubik et al., 2023).

Yet even when combining multiple sensors together (Li et al., 2021; Tulbure et al., 2022), satellites provide an incomplete observation of maximum flood extent due to vegetation or cloud blockage (Shastry et al., 2023) and flood water can recede before an observation. Even with radar sensors, capturing the peak extent of the event is challenging (Bauer-Marschallinger et al., 2022), and the side looking angle of radar makes urban observations a challenge due to occlusion by buildings. Water in riparian forest and under canopy can only be detected in longer wavelength L-band microwave sensors, at course resolutions (Jensen & Mcdonald, 2019; Du et al., 2018). 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 spatially and temporally complete predictions at peak inundation moments. Unlike satellite observations, these models can make gap-free flood forecasts, project flooding for the past with reanalysis data, or estimate inundation change in future climates. Hydrodynamic models require an enormous amount of setup and computational time to run, making simulations for NRT flood hazard assessment at large scales (Van den Bout et al., 2023) an ongoing challenge. Most operational or NRT hydrodynamic models are well suited to estimate fluvial inundation from riverbank overflows, but real world damaging flood events are often compound (Guan et al., 2023) or multi-form (Kruczkiewicz et al., 2022) with rainfall, riverbank, infrastructure failure, storm surge, or other compound influences. Pluvial inundation is challenging for many modeling methods, particularly the Height Above Nearest Drainage (HAND) method, which requires a pre-defined nearby flowpath for an inundation prediction using the discharge output (Aristizabal et al., 2023). Hazard models (P. D. Bates et al., 2021) which include coastal, pluvial, and riverine flooding are often based on scenarios taken from the historical record or future climate scenarios, not generated in NRT from current conditions. Continental or Global Scale models that operate in NRT (Alfieri et al., 2018) are typically discharge predictions (e.g. ECMWF's Glofas and EEFAS models (Dottori et al., 2017)). ECMWF's GLOFAS model translates discharge predictions into spatial extents via a lookup table and catalogues of previously processed inundation extents, and is not dynamically modeled (Dottori et al., 2017).

1.3 Deep learning improves modeling of surface water dynamics

Many hydrology and water resources problems have been successfully addressed with deep learning (G. S. Nearing et al., 2020; Nevo et al., 2022; Frame, 2022). Flood forecasting and monitoring that primarily relies on streamflow (G. Nearing et al., 2024) will suffer during pluvial and compound events, which are responsible for damaging floods (Guan et al., 2023). Merging satellite observations with hydrodynamic models has been approached with data assimilation, but suffers from temporal availability of satellite data issues described above (Jafarzadegan et al., 2021). The general strategy of our deep learning model is to train on a large sample of flood scenarios to learn to generate continuous accurate flood maps without the need for intensive runtime computations or time consuming curation of local data sources.

Most approaches to deep learning for flood mapping rely on convolutional neural networks (CNN). Guo et al. (2021) proposed a CNN for urban flood mapping, but warned that a CNN model should not be trained on one catchment area only. Zhou et al. (2022) trained a CNN to predict a continuous flood inundation extent from point-based water level data. Dasgupta et al. (2022a) 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." Our approach, previously introduced by Nair et al. (2022), 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. We refer to this as "NWM-CNN"

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 in California (Zou et al., 2023). The 2023 flooding affected a large portion of the state. There were 955 flood, flash flood, or debris flow reports logged by the National Weather Service (NWS), and several levees broke along the Consumes River and Pajaro River. The Salinas river overflow cutoff transportation access to the Monterrey Peninsula. The estimated 5-7 billion USD in property losses was the most damaging flood event recorded in California history (the second being flooding in Jan/March 1995, 2 billion loss inflated adjusted). 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%) (Carpenter, 2023).

We use 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 (height above nearest drainage; 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. We use the NWM as the hydrological foundation for predicting the resulting surface water extent observable from Sentinel-2. 780 flood events, with corresponding Sentinel-2 images were selected by sampling from 2015-2022 across a gradient of urbanization, surface water, and geographies (Inland, Coastal Atlantic North, Coastal Atlantic South, Coastal Pacific, Coastal Gulf, and Inland) with surface water estimates using a convolutional neural network (CNN) trained on handlabels from Floodbase, with a Critical Success Index (CSI; also known as Intersection over Union score) of 76.3% (s.d. 3.3%) on never before flooded areas and 88.6% (s.d 4.2.%) on previously flooded areas (Table A2). 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., 2016a) 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 250m 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 250m resolution. The model was trained in 3 folds of data, withholding a 4th fold as a held out test set, averaging a performance of 4.58 RMSE (s.d 2.07%) across geographies (Table A1).

2.1.1 U-Net Architecture

We specifically 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 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:

$$\begin{aligned}
C_{l,1} &= Swish(BatchNorm(x_l * W_{l,1})) \\
DwC_{l,2} &= Swish(BatchNorm((C_{l,1} * W_{l,2})) \\
P_{l,4} &= AvgPool2d(C_{l,3}) \\
C_{l,5} &= Swish(P_{l,4} * W_{l,5}) \\
C_{l,6} &= Sigmoid(C_{l,5} * W_{l,6}) \\
M_{l,7} &= C_{l,3} * C_{l,6} \\
C_{l,8} &= BatchNorm(M_{l,7} * W_{l,8})
\end{aligned} \tag{1}$$

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 $P_{l,4}$ through $C_{l,6}$ which facilitates the model to learn relationships between its different input layers (ie. 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:

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

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:

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

where PSWApp is the resulting image of surface water area percentages 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 scales 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 $PSWA_1, PSWA_2, \ldots, PSWA_n$ as the corresponding time series, where $PSWA_t$ represents the t-th image. The average pixel value of an image $PSWA_t$ is denoted as $PS\bar{W}A_t$, 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, $PSWA_{min}$.

$$ASWA_t = PS\bar{W}A_t - PS\bar{W}A_{min} \tag{4}$$

where $PS\bar{W}A_t$ is calculated as:

$$P\bar{SW}A_i = \frac{1}{N}\sum_{x,y} PSWApp_i(x,y) \tag{5}$$

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

$$P\bar{SW}A_{min} = \min\left\{P\bar{SW}A_1, P\bar{SW}A_2, \ldots, P\bar{SW}A_n\right\} \tag{6}$$

.

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 6^{th} - 13^{th} 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 250m resolution from NWM-HAND versus NWM-CNN across our domain i) square root of the mean squared error (RMSE), ii) precision, recall (A.K.A. Hit Rate), and FI scores, and iii) CSI, False Alarm Ratio (FAR), and Error Bias (EB). For RMSE, we re-sample the Sentinel-1 flood map from 10 meter resolution to 250 meter resolution using the mean pixel value, yielding the percent of 250-meter 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 calculated the binary metrics we include pixel values greater than zero as "true" and pixel values of zero as "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 PSWA was then qualitatively compared to historical flooding events in the Sacramento area, specifically a $30km$ 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), which includes floods that occurred after the 1996 water year (Murphy, John D., 2021).

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 250m 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 (180201, up to 1.25% ASWA). Within 180201 there is a wide variation of surface water responses at the HUC 8 scale, with the largest coming from 18020158 (up to 10% ASWA), which is highlighted along with 18020104 and 18020159 in Figure 1.

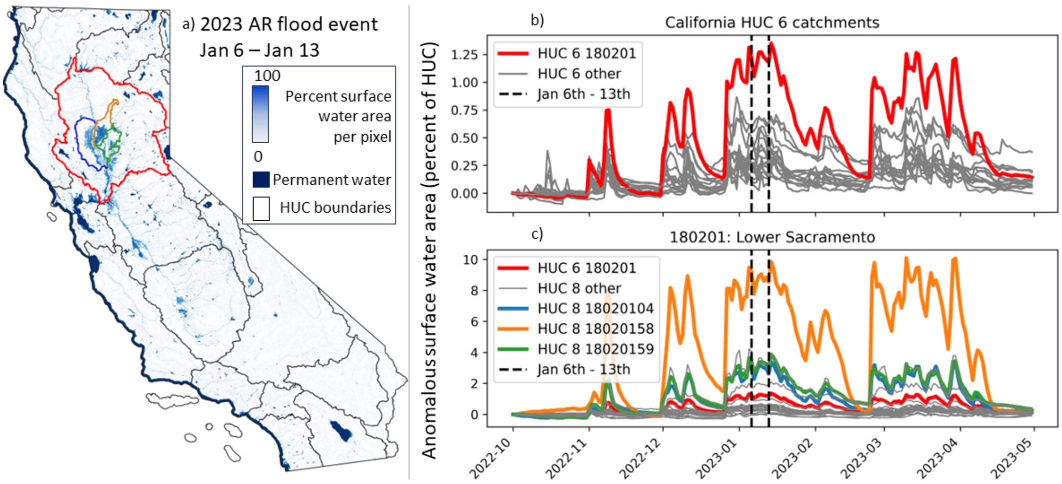

Figure 1. A: California statewide snapshot of surface water predicted by NWM-CNN from a January 2023 AR event. B: Summarized surface water areas across all of California at the HUC 6 scale. C: Summarized surface water area at the HUC 8 scale across the Lower Sacramento catchment area, the HUC 6 catchment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

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 modelled 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 this figure the satellite observations are plotted with 50% transparency, which shows the 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 to 50 PSWApp, but under predicts the number of pixels above 50 PSWApp. NWM-HAND method in black under predicts observed surface water.

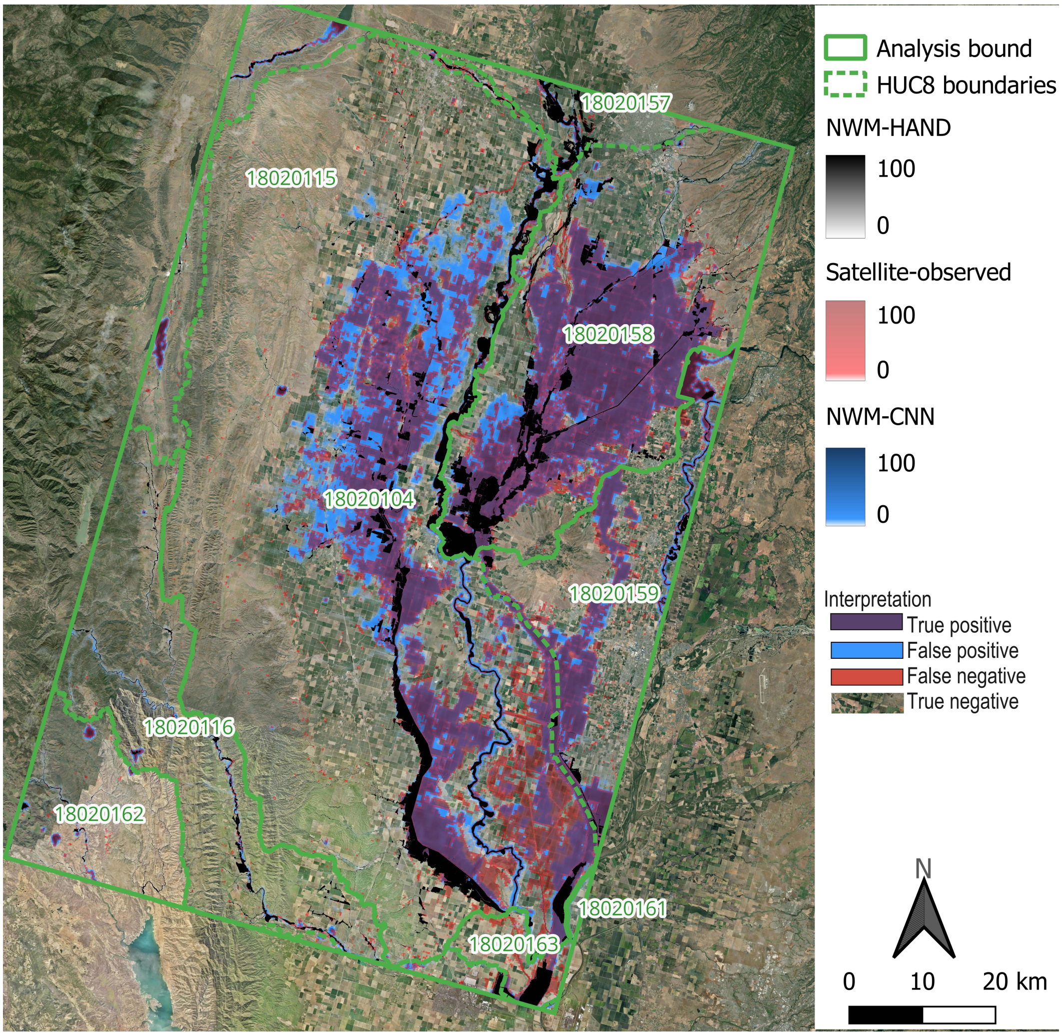

Figure 2. Direct comparison of the mapped model results, as PSWApp, where the blue represents our NWM-CNN, the satellite-observed surface water extent map is shown in transparent red, and the NWM-HAND results are shown in black. With this color scheme, NWM-CNN false positives appear blue, true positives appear magenta and false negatives appear red.

Using the 250m 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).

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 ob-

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

Table 1. Model performance statistics for Sacramento during January 23 Atmospheric River event.

servations, 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 0 (see sensitivity analysis, **??**). 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 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; Trigg et al., 2016; Bernhofen et al., 2018) 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" vs "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 B).

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 machine learning (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 predicts a relatively high maximum ASWA also corresponds to a year with flood damages, with 2022 as a notable exception. Of the five highest water years predicted 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 million dollars (USD), 1997, 1998, 2006, and 2023).

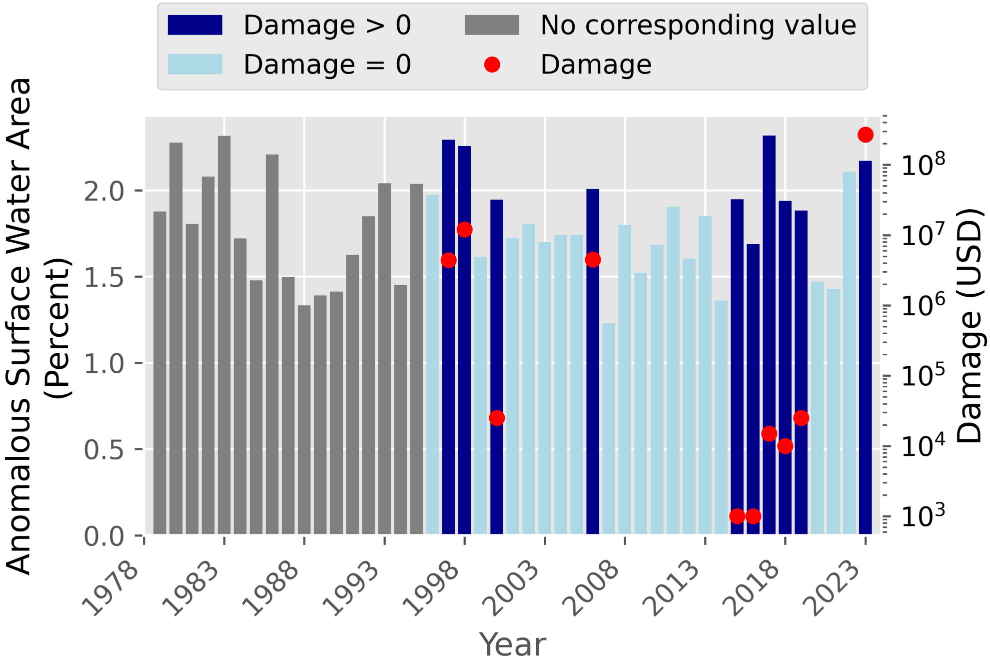

Figure 3. Annual (water year) maximum anomalous surface water area (%) across the Sacramento analysis domain. Grey colored bars from 1979 to 1996 do not have corresponding data in the Storm Events Database, while blue bars from 1996 onward were references against that database for Sacramento County,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

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 can not 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 (e.g., magnitude, timing, extent, and relative damage) 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. Future research aims to develop an approach that scales globally beyond CONUS.

The rapid run time over large spatial and temporal scale, along with the gap-free nature of inundation predictions spatially and temporally, 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. Machine learning 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.

Open Research Section

Data are provided at HydroShare:
`https://www.hydroshare.org/resource/dbf8e4c2a39a4c228db867b04f9c21ed/`.
Analysis code for results presented in this paper is available on GitHub:
`https://github.com/jmframe/NWM_CNN_california_AR_2023`.
DOI numbers for both the data and code will be generated upon article acceptance.

References

Alfieri, L., Cohen, S., Galantowicz, J., Schumann, G. J.-P., Trigg, M. A., Zsoter, E., ... Salamon, P. (2018). A global network for operational flood risk reduction. , *84*, 149–158. Retrieved 2020-03-11, from `https://linkinghub.elsevier.com/retrieve/pii/S1462901117312637` doi: 10.1016/j.envsci.2018.03.014

Allen, M., Dube, O., Solecki, W., Aragón-Durand, F., Cramer, W., Humphreys, S., ... others (2018). Special report: Global warming of 1.5 c. *Intergovernmental Panel on Climate Change (IPCC)*.

Aristizabal, F., Salas, F., Petrochenkov, G., Grout, T., Avant, B., Bates, B., ... Judge, J. (2023, 5). 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*. doi: 10.1029/2022WR032039

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

Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., ... Krajewski, W. F. (2021). Combined modeling of US fluvial, pluvial, and coastal flood hazard under current and future climates. , *57*(2). Retrieved 2021-05-13, from `https://onlinelibrary.wiley.com/doi/10.1029/2020WR028673` doi: 10.1029/2020WR028673

Bauer-Marschallinger, B., Cao, S., Tupas, M. E., Roth, F., Navacchi, C., Melzer, T., ... Wagner, W. (2022). Satellite-based flood mapping through bayesian inference from a sentinel-1 sar datacube. *Remote Sensing*(15). Retrieved from https://www.mdpi.com/2072-4292/14/15/3673 doi: 10.3390/rs14153673

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

Bernhofen, M. V., Whyman, C., Trigg, M. A., Sleigh, P. A., Smith, A. M., Sampson, C. C., ... Winsemius, H. C. (2018). A first collective validation of global fluvial flood models for major floods in nigeria and mozambique. , *13*(10), 104007. Retrieved 2018-11-06, from http://stacks.iop.org/1748-9326/13/i=10/a=104007?key=crossref.d44bc2d3a14ca7860ccd8e2743346ebb doi: 10.1088/1748-9326/aae014

Bevere, L., & Finucane, J. (2022, Sep). *Flood: new risk-based pricing capabilities, new opportunities to close protection gaps* (Tech. Rep.). Retrieved from https://www.swissre.com/institute/research/sigma-research/Economic-Insights/flood-new-opportunities.html

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 http://openaccess.thecvf.com/content_CVPRW_2020/html/w11/Bonafilia_Sen1Floods11_A_Georeferenced_Dataset_to_Train_and_Test_Deep_Learning_CVPRW_2020_paper.html doi: 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*.

Brakenridge, G. R. (2010). Global active archive of large flood events. *Dartmouth Flood Observatory, University of Colorado*.

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

2023)

Carpenter, G. (2023). *Atmospheric river events.*

Cosgrove, B., Gochis, D., Flowers, T., Dugger, A., Ogden, F., Graziano, T., … Zhang, Y. (2024, jan). NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling. *JAWRA Journal of the American Water Resources Association*. Retrieved from `https://onlinelibrary.wiley.com/doi/10.1111/1752-1688.13184` doi: 10.1111/1752-1688.13184

Dasgupta, A., Hybbeneth, L., & Waske, B. (2022a). Towards daily high-resolution inundation observations using deep learning and eo. *arXiv preprint arXiv:2208.09135*.

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

de Bruijn, J. A., de Moel, H., Jongman, B., de Ruiter, M. C., Wagemaker, J., & Aerts, J. C. (2019). A global database of historic and real-time flood events based on social media. *Scientific data*, *6*(1), 311.

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., … Watts, J. D. (2018). Assessing global surface water inundation dynamics using combined satellite information from smap, amsr2 and landsat. *Remote sensing of environment*, *213*, 1–17.

Frame, J. M. (2022, 8). 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., Kratzert, F., Raney, A., Rahman, M., Salas, F. R., & Nearing, G. S. (2021). Post-processing the national water model with long short-term memory networks for streamflow predictions and model diagnostics. *Journal of the American Water Resources Association*, 1-21. Retrieved from `https://onlinelibrary.wiley.com/doi/epdf/10.1111/1752-1688.12964` doi: 10.1111/1752-1688.12964

Glaudemans, M. (2023, September 26). *Updated: Soliciting comments on experi-*

mental flood inundation mapping (fim) services through september 30, 2024. National Weather Service Headquarters, Silver Spring, MD. Retrieved from https://viewer.weather.noaa.gov/water (Public Information Statement 23-55 Updated)

A global database of historic and real-time flood events based on social media. (2023, August). Retrieved from https://www.ncdc.noaa.gov/stormevents/details.jsp (Accessed: February 2023)

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT press.

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 doi: 10.1016/j.wasec.2023.100136

Guo, Z., Leitão, J. P., Simões, N. E., & Moosavi, V. (2021, 3). Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks. *Journal of Flood Risk Management*, *14*. doi: 10.1111/jfr3.12684

Hänsch, R., Arndt, J., Lunga, D., Gibb, M., Pedelose, T., Boedihardjo, A., ... Bacastow, T. M. (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., ... Glinskis, E. (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.

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

Jakubik, J., Chu, L., Fraccaro, P., Gomes, C., Nyirjesy, G., Bangalore, R., ... Granger, E. (2023, August). *Prithvi-100M.* doi: 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.

Kruczkiewicz, A., Cian, F., Monasterolo, I., Di Baldassarre, G., Caldas, A., Royz, M., ... Van Aalst, M. (2022). Multiform flood risk in a rapidly changing world: what we do not do, what we should and why it matters. , *17*(8), 081001. Retrieved 2023-09-28, from https://iopscience.iop.org/article/10.1088/1748-9326/ac7ed9 doi: 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*, 93-104.

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. doi: 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.

Liu, Y., Tarboton, D., & Maidment, D. (2020). *Height above nearest drainage (hand) and hydraulic property table for conus - version 0.21 (20200601).* Oak Ridge National Laboratory Leadership Computing Facility. Retrieved from https://cfim.ornl.gov/data/vis/v0.2.1/ (Created: 5/27/2020, 6:47:37 AM; Published: 5/27/2020, 11:15:25 AM) doi: 10.13139/ORNLNCCS/1630903

Long, J., Shelhamer, E., & Darrell, T. (2014, 11). Fully convolutional networks for semantic segmentation. *arxiv.org*. Retrieved from http://arxiv.org/abs/1411.4038

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 doi: 10.1016/j.ijdrr.2018.10.022

Murphy, John D. (2021, July 26). Storm Data Preparation: National Weather Service Instruction 10-1605 [Computer software manual]. United States. Retrieved from https://www.nws.noaa.gov/directives/sym/pd01016005curr.pdf (Accessed: November 2023)

Nair, T., Sunkara, V., Frame, J., Popien, P., & Chakrabarti, S. (2022). Deep hydrology : Hourly , gap-free flood maps through joint satellite and hydrologic modelling. *Neurips*. Retrieved from https://s3.us-east-1.amazonaws.com/climate-change-ai/papers/neurips2022/20/paper.pdf

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

National Oceanic and Atmospheric Administration. (2023b, September 28). *Bipartisan infrastructure law.* https://www.noaa.gov/infrastructure-law. (Accessed: December 2023)

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., … Matias, Y. (2024, 3). Global prediction of extreme floods in ungauged watersheds. *Nature*, *627*, 559-563. Retrieved from https://www.nature.com/articles/s41586-024-07145-1 doi: 10.1038/s41586-024-07145-1

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., … Gupta, H. V. (2020). What role does hydrological science play in the age of machine learning? *Water Resources Research*. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2020WR028091 doi: 10.1029/2020wr028091

Nevo, S., Morin, E., Rosenthal, A. G., Metzger, A., Barshai, C., Weitzner, D., … Matias, Y. (2022, 8). Flood forecasting with machine learning models in an operational framework. *Hydrology and Earth System Sciences*, *26*, 4013-4032. Retrieved from https://hess.copernicus.org/articles/26/4013/2022/ doi: 10.5194/hess-26-4013-2022

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

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016b). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422. doi: 10.1038/nature20584

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

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. (Accessed in 2023). *Region's flooding history.* https://waterresources.saccounty.gov/stormready/Pages/Region%27s-Flooding-History.aspx. (Sacramento's risk of flooding is the greatest of any major city in the country. Over the past few decades, our area has experienced significant, sometimes devastating, flooding. The most notable flooding occurred in 1986, 1995, 1997, 2006, and 2017. The Sacramento Area Flood Control Agency identifies Sacramento as the nation's greatest metropolitan flood risk.)

Sacramento County Department of Water Resources. (2016). *2016 sacramento countywide local hazard mitigation plan update.* 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., ... Noman, N. (2018). Towards real-time continental scale streamflow simulation in continuous and discrete space. *Journal of the American Water Resources Association*, *54*, 7-27. doi: 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 http://www.mdpi.com/2072-4292/10/8/1230 doi: 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.

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.

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.

Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., . . . Slayback, D. A. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, *596*, 80-86. Retrieved from http://dx.doi.org/10.1038/s41586-021-03695-w doi: 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., . . . others (2016). The credibility challenge for global fluvial flood risk analysis. *Environmental Research Letters*, *11*(9), 094014.

Tulbure, M. G., Broich, M., Perin, V., Gaines, M., Ju, J., Stehman, S. V., . . . Betbeder-Matibet, L. (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://www.sciencedirect.com/science/article/pii/S0924271622000338 doi: 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.* https://croplandcros.scinet.usda.gov/. USDA-NASS, Washington, DC. (Accessed: 2023; Verified: 2023)

Van den Bout, B., Jetten, V., Cees, J. v. W., & Lombardo, L. (2023). A breakthrough in fast flood simulation. *Preprint. Under review*. doi: https://doi.org/10.31223/X5MM2Z

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://www.sciencedirect.com/science/article/pii/S0034425723000032 doi: 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://www.sciencedirect.com/science/article/pii/S2589915519300239 doi: https://doi.org/10.1016/j.hydroa.2019.100039

Xie, Q., Luong, M.-T., Hovy, E., & Le, Q. V. (2020, June). Self-training with noisy student improves imagenet classification. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition (cvpr).*

Yang, Z.-L., Cai, X., Zhang, G., Tavakoly, A. A., Jin, Q., Meyer, L. H., & Guan, X. (2011). *The community noah land surface model with multi-parameterization options (noah-mp) technical description.* Retrieved from https://www.jsg.utexas.edu/noah-mp/files/Noah-MP_Technote_v0.2.pdf

Zhao, M., Olsen, P. A., & Chandra, R. (2021, 6). 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.

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

Zou, X., Cordeira, J. M., Bartlett, S. M., Kawzenuk, B., Roj, S., Castellano, C. M., . . . Ralph, F. M. (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 doi: 10.22541/essoar.168677226.69319241/v1