

---

# Deep Hydrology: Hourly, Gap-Free Flood Maps Through Joint Satellite and Hydrologic Modelling

---

Tanya Nair   Veda Sunkara   Jonathan Frame   Philip Popien   Subit Chakrabarti  
Cloud to Street  
Brooklyn, NY  
`{tanya,veda,jonathan,philip,subit}@cloudtostreet.ai`

## Abstract

Climate change-driven weather disasters are rapidly increasing in both frequency and magnitude [18]. Floods are the most damaging of these disasters, with approximately 1.46 billion people exposed to inundation depths of over 0.15m, a significant life and livelihood risk [12]. Accurate knowledge of flood-extent for ongoing and historical events facilitates climate adaptation in flood-prone communities by enabling near real-time disaster monitoring to support planning, response, and relief during these extreme events. Satellite observations can be used to derive flood-extent maps directly; however, these observations are impeded by cloud and canopy cover, and can be very infrequent and hence miss the flood completely. In contrast, physically-based inundation models can produce spatially complete event maps but suffer from high uncertainty if not frequently calibrated with expensive land and infrastructure surveys [17, 19]. In this study, we propose a deep learning approach to reproduce satellite-observed fractional flood-extent maps given dynamic state variables from hydrologic models, fusing information contained within the states with direct observations from satellites. Our model has an hourly temporal resolution, contains no cloud-gaps, and generalizes to watersheds across the continental United States with a 6% error on held-out areas that never flooded before. We further demonstrate through a case study in Houston, Texas that our model can distinguish tropical cyclones that caused flooding from those that did not within two days of landfall, thereby providing a reliable source for flood-extent maps that can be used by disaster monitoring services.

## 1 Introduction

Floods caused losses of USD 20 billion globally in 2021 alone, and the relative property loss caused by floods are highest in places with social vulnerability [16], further exacerbating economic disparities. Effective disaster response can mitigate the aftermath of widespread flooding; however, flood managers and disaster response personnel require timely and accurate maps of the maximum or "peak" spatial inundation (flood-extent maps) to mount an appropriate response.

Previous work mapping floods and flood risk falls broadly in two categories: physically-based hydraulic models and satellite-observed segmentation models. Hydraulic models use physical equations that define relationships between streamflow, elevation, slope, soil-type, and human infrastructure to produce flood depth and extent maps. To build high-quality hydraulic flood models however, requires engineering-grade models that capture geographical and infrastructure conditions, a complex and expensive process that is not accessible to resource-limited communities in short time frames. Freely available hydraulic models exist and can be used to generate flood maps, but suffer from oversimplifications such as excluding precipitation-driven or pluvial flooding [4, 20], that reduces the accuracy of their outputs. Publicly available optical and active microwave satellite

imagery enables model-free flood monitoring and can be used to generate flood-extent maps with machine learning methods. However, cloud and canopy cover consistently limit observations during rainfall-driven disasters for optical satellites, and the 6-12 day revisit times of active microwave (cloud-piercing) satellites mean only slow-moving floods can be reliably captured [22].

Emerging studies have leveraged both observed and modeled signals to detect floods, with limitations in geographic generalization and applicability. Dasgupta et al. [2] used coarse soil moisture and precipitation observations to predict inundation given a pre-flood optical satellite image in a case study a single region of interest. However, in a feature analysis they showed no impact of flood forcing information, a counterintuitive result that indicates overfitting to the information in a single region of interest (ROI), where flooding primarily occurred in low-lying areas adjacent to a river. Furthermore, their model requires cloud-free pre-flood and during-flood images to train, which can be impractical to collect for fast-moving or rain-driven events. Guo et al. [5] trained an image translation network to convert static terrain representations to water depth images given a hyetograph. While their results are promising, their study was limited to 18 events in a single catchment, and all training data was simulated using hydraulic models. Other recent works have trained networks to produce the outputs of a hydraulic model given a similar set of inputs; however, these studies have been limited to synthetic flooding scenarios [6, 21], single and/or small study areas [6, 9, 21], or riverine (fluvial) flooding only [6].

**Contribution** In this study, we train a fully convolutional encoder-decoder network on hydrologic model states and auxiliary static layers to produce fractional water extent maps (henceforth referred to as "flood maps") similar to those derived directly from satellite observations. To our knowledge, this is the first work that directly uses hydrologic states and satellite observed water to train a single network in a large, multi-watershed study. Our approach leverages the robustness and scalability of satellite observations with the spatial completeness of physically based inundation models in order to produce hourly, gap-free flood maps, which can enable all communities in the continental United States to respond and adapt to the effects of increasingly prevalent flooding.

## 2 Methods

We trained a fully convolutional encoder-decoder network to regress the fractional water in a 2km x 2km pixel (a grid cell) given two dynamic input layers and three static input layers. Architecture details are provided in Appendix A.1.

**Target flood maps** We selected 6865 satellite observations from 189 flood events from 2000-2021 in the contiguous United States (CONUS) that have been identified by the Dartmouth Flood Observatory [1]. Our selection from 21,017 candidate images was based on how much flood water and how many unobstructed (cloud-free) pixels an image contained. As a training target, we used fractional water predictions from a proprietary water segmentation model, designed to regress fractional water in a 250m x 250m grid cell given a satellite observation.

**Dynamic Inputs** For each satellite observation, we extract two corresponding hydrologic state variables from the U.S. National Water Model (NWM) [14]. First, a 1km resolution volumetric soil moisture taken from the NWM land surface component (NOAH-MP), defined as a ratio of water volume to soil volume. The second hydrologic state we use is a 250m resolution terrain router, representing a depth of water ponded on the surface. The terrain router uses a physical equation to route overland flow across the land surface based on a digital elevation model (DEM). Here it is important to note that the dynamics of flooding are more complex than described with the terrain router. It is possible for a lot of routing to occur on a grid cell without flooding occurring, due to factors including evaporation, soil transmission, and human-made flood infrastructure. NWM dynamic data sources are available hourly. In order to capture temporal information about the flood dynamics, we aggregate the previous 72 hours of terrain routing and soil moisture, and provide these as inputs to the model.

**Static Inputs** For each satellite observation we also extracted three static layers, to provide the model with context about locations that may be more or less prone to flood. We use two DEM-derived layers from Hydrosheds [8]: flow direction, describing the direction of water flow as one of eight equidistant radians, and a flow accumulation layer, describing the number of upstream grid cells. Both

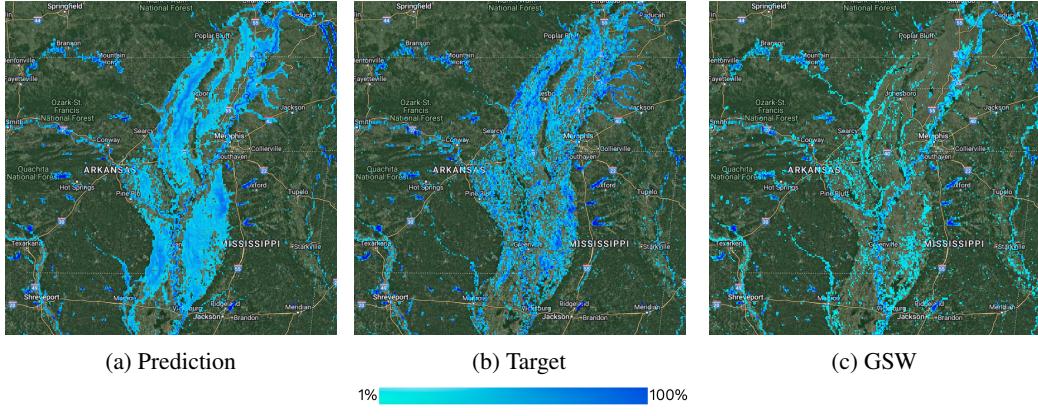


Figure 1: Held-out fractional water prediction of a flood event along the Mississippi River

Hydrosheds layers are provided at 15 arcseconds resolution. Finally, we also use a 30m resolution global surface water layer derived from the the Joint Research Council Global Surface Water data v.1.1 (GSW) [11], which describes the frequency of occurrence of water on a grid cell.

**Splits** We split 6865 examples into three cross-validation folds and a single test set. Splits were made temporally and assessed for geographic balance afterwards. More details are provided in Appendix A.2.

### 3 Results and Discussion

We measure Mean Squared Error (MSE) across cross-validation folds, excluding true-negative pixels (dry land defined as  $y < 0.02$  and  $\hat{y} < 0.02$ ) in this calculation to avoid an artificial suppression of the error, as most pixels in the dataset do not contain water. We find that among pixels where  $\text{GSW} \in [0, 0.02]$  (never-flooded pixels), the network has a MSE of  $4.0e-3 \pm 1.9e-4$ , corresponding to roughly 6% ( $0.25\text{km}^2$ ) error per pixel. Where  $\text{GSW} \in [0.02, 0.30]$  (flooded-before pixels), the network obtains a MSE of  $4.8e-3 \pm 4.1e-4$  (roughly 7% or  $0.27\text{km}^2$  error per pixel). Evaluating on these subgroups demonstrates that the network is not reproducing a trivial solution by re-producing the GSW input. An example of a held-out prediction is provided in Fig. 1 demonstrates the network’s ability to make predictions that align with a direct observation.

To determine whether the network can provide actionable flood evidence, we evaluated whether the network can separate hurricanes and storms in the Houston area based on whether or not they caused flooding within the city. We select five events that struck Houston between 2005-2021: three that caused flooding in the central Houston (Hurricane Harvey, Storms Nicholas and Imelda), and two that did not (Hurricane Laura, Storm Rita). This flood assessment was based on news reports. None of these events were part of the training dataset.

We used the network to make hourly predictions of fractional water within a  $30\text{km} \times 30\text{km}$  region of interest (ROI) centered on Houston. Then, we use the input GSW layer to exclude permanent water grid cells ( $\text{GSW} > 0.30$ ) to compute the Fractional Flooded Area (FFA) at each hour of the ROI throughout the event (Fig. 2). Additionally, we visualize the predicted water extent for Hurricanes Laura and Harvey on their respective peak days (Fig. 3).

The Houston case study demonstrates that this model can detect flooding within one day of a tropical cyclone making landfall. This means that as storms evolve, users can understand where floods are happening in order to appropriately direct emergency resources. In contrast, through a study of five flood events over Houston, we found that it took, on average, 5-13 days to obtain a cloud-free view of 65% of Houston using the MODIS Surface Reflectance product, which has the fastest revisit time among public monitoring satellites of 0.5 days. Instead of waiting for days for an unimpeded satellite observation, and potentially missing damaging floods that recede in less than five days, this model can be used to aid response with localized near real-time flood maps. And unlike geographic data points obtained through scraping social media, such as Twitter, that may be completely unavailable due to power outages during storms, or unreliable in less populated areas, our model can be run independently by response teams outside of the affected region.

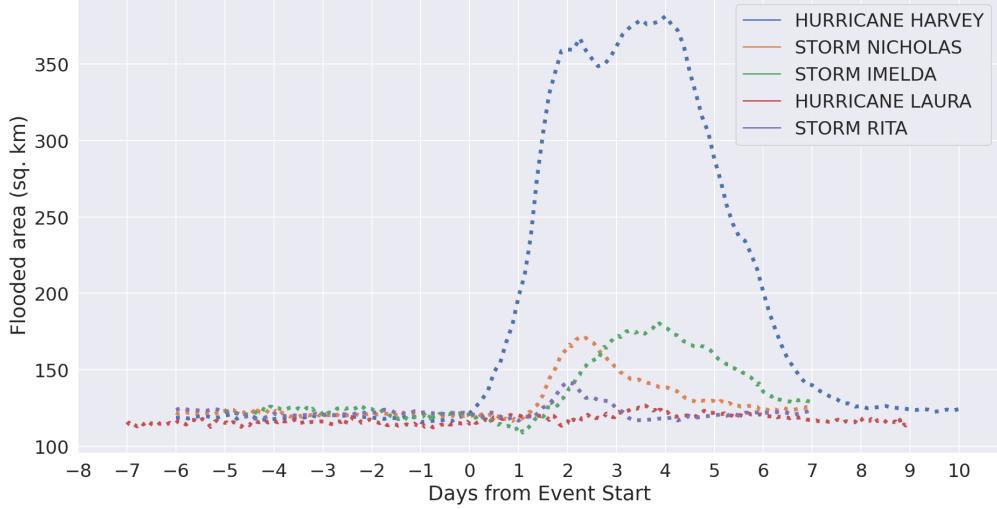


Figure 2: Predicted Fractional Flooded Area for five named storms that struck the Houston area. Harvey, Nicholas, and Imelda caused flood damage, whereas Laura and Rita did not.

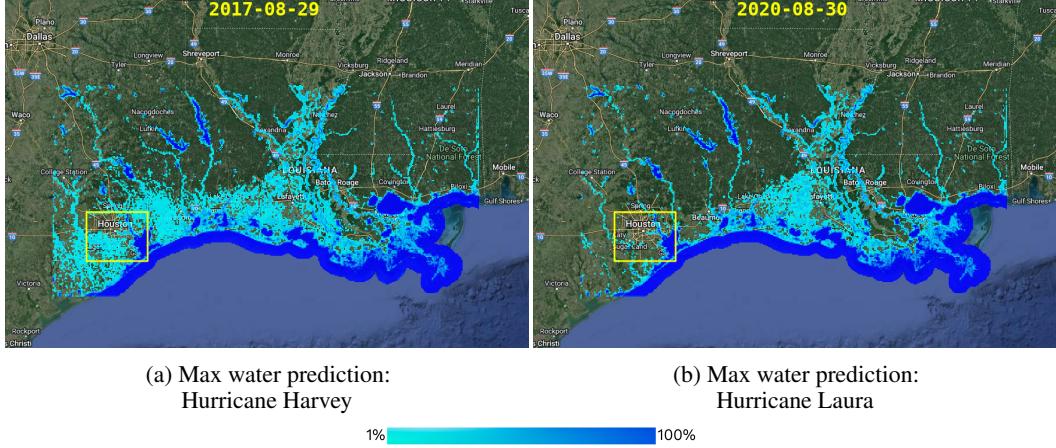


Figure 3: Fractional water predicted during a) Hurricane Harvey and b) Hurricane Laura on their respective peak days. The yellow box indicates the Houston ROI.

## 4 Conclusion

We demonstrate a fractional water segmentation model based on hydraulic state variables and static geographic layers, and showed that this model can be used to predict satellite observed flood maps when no such observation is available, as is often the case during flood disasters. By reproducing satellite observations, our model is able to scale to multiple locations and multiple resolutions, overcoming limitations of physically based models. Future work should explore integrating globally-available data sources, so that the model can be run outside of the US. Additionally, future validation should consider proxy mechanisms to validate the performance of the model during events for which no observation is available, either by comparing to flood-risk maps, physically modelled flood maps, or flood insurance claims. These recommendations are suggested with the ultimate goal of providing robust, gap-free flood maps that can be reliably used by emergency response teams to direct resources during flood disasters. This work presents a scientifically novel step forward in that direction, and hopefully motivates future solutions that leverages all possible information sources to provide timely and reliable information to communities responding to flood disasters.

## References

- [1] G Robert Brakenridge. “Global active archive of large flood events”. In: *Dartmouth Flood Observatory, University of Colorado* (2010).
- [2] Antara Dasgupta, Lasse Hybbeneth, and Björn Waske. “Towards Daily High-resolution Inundation Observations using Deep Learning and EO”. In: *arXiv preprint arXiv:2208.09135* (2022).
- [3] Jia Deng et al. “Imagenet: A large-scale hierarchical image database”. In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee. 2009, pp. 248–255.
- [4] Francesco Dottori et al. “Development and evaluation of a framework for global flood hazard mapping”. In: *Advances in water resources* 94 (2016), pp. 87–102.
- [5] Zifeng Guo et al. “Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks”. In: *Journal of Flood Risk Management* 14.1 (2021), e12684. DOI: <https://doi.org/10.1111/jfr3.12684>.
- [6] Syed Kabir et al. “A deep convolutional neural network model for rapid prediction of fluvial flood inundation”. In: *Journal of Hydrology* 590 (2020), p. 125481. ISSN: 0022-1694. DOI: <https://doi.org/10.1016/j.jhydrol.2020.125481>.
- [7] Diederik P Kingma and Jimmy Ba. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014).
- [8] Bernhard Lehner, Kristine Verdin, and Andy Jarvis. “New global hydrography derived from spaceborne elevation data”. In: *EOS, Transactions American Geophysical Union* 89.10 (2008), pp. 93–94.
- [9] Fumiya Makinoshima et al. “Early forecasting of tsunami inundation from tsunami and geodetic observation data with convolutional neural networks”. In: *Nature Communications* 12 (Apr. 2021). DOI: [10.1038/s41467-021-22348-0](https://doi.org/10.1038/s41467-021-22348-0).
- [10] NASA. *MODIS Grids*. Tech. rep. 2021. URL: [https://modis-land.gsfc.nasa.gov/MODLAND\\_grid.html](https://modis-land.gsfc.nasa.gov/MODLAND_grid.html).
- [11] Jean-François Pekel et al. “High-resolution mapping of global surface water and its long-term changes”. In: *Nature* 540.7633 (2016), pp. 418–422. DOI: [10.1038/nature20584](https://doi.org/10.1038/nature20584).
- [12] Jun Rentschler and Melda Salhab. *People in Harm’s Way : Flood Exposure and Poverty in 189 Countries. Policy Research Working Paper*. Tech. rep. World Bank, Washington, DC, 2020.
- [13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.
- [14] Fernando R. Salas et al. “Towards Real-Time Continental Scale Streamflow Simulation in Continuous and Discrete Space”. In: *Journal of the American Water Resources Association* 54 (1 2018), pp. 7–27. ISSN: 17521688. DOI: [10.1111/1752-1688.12586](https://doi.org/10.1111/1752-1688.12586).
- [15] Mingxing Tan and Quoc V. Le. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”. In: 2019. URL: <https://arxiv.org/abs/1905.11946>.
- [16] Beth Tellman et al. “Using disaster outcomes to validate components of social vulnerability to floods: flood damage and property damage across the USA”. In: *Using disaster outcomes to validate components of social vulnerability to floods: flood damage and property damage across the USA*. SocArXiv, June 2020. DOI: [10.31235/osf.io/byrgu](https://doi.org/10.31235/osf.io/byrgu).
- [17] MA Trigg et al. “The credibility challenge for global fluvial flood risk analysis”. In: *Environmental Research Letters* 11.9 (2016), p. 094014.
- [18] Margareta Wahlstrom et al. *The human cost of weather-related disasters 1995–2015*. Tech. rep. 2015.
- [19] Philip J Ward et al. “Usefulness and limitations of global flood risk models”. In: *Nature Climate Change* 5.8 (2015), pp. 712–715.
- [20] Dai Yamazaki et al. “A physically based description of floodplain inundation dynamics in a global river routing model”. In: *Water Resources Research* 47.4 (2011).
- [21] Xingyu Yan et al. “A Rapid Prediction Model of Urban Flood Inundation in a High-Risk Area Coupling Machine Learning and Numerical Simulation Approaches”. In: *International Journal of Disaster Risk Science* 12.20210610 (2021), p. 903. ISSN: 2095-0055. DOI: [10.1007/s13753-021-00384-0](https://doi.org/10.1007/s13753-021-00384-0).

- [22] Jie Zhao et al. “A large-scale 2005–2012 flood map record derived from ENVISAT-ASAR data: United Kingdom as a test case”. In: *Remote Sensing of Environment* 256 (2021), p. 112338.

## A Appendix

### A.1 Model Architecture

We trained an encoder-decoder UNet-style architecture [13] with an EfficientNet B1 [15] encoder pretrained on Imagenet [3] to minimize the Mean Squared Error (MSE) between predicted fractional water  $\hat{Y}$  and observation  $Y$ , given multi-channel input  $X$ . The model is optimized with Adam [7]. During training, random crops of 352x352, and random flips (horizontal and vertical) were applied as data augmentation.

### A.2 Splits

Fig. 4a) shows the temporal distribution of splits. 6865 examples are split into three cross-validation folds and a single test set. Splits are made temporally, and assessed for geographic balance after the fact (Fig. 4b). The sinusoidal tile grid of the MODIS land products is used to assess geographic balance, with each tile representing  $10^\circ \times 10^\circ$  at the equator [10].

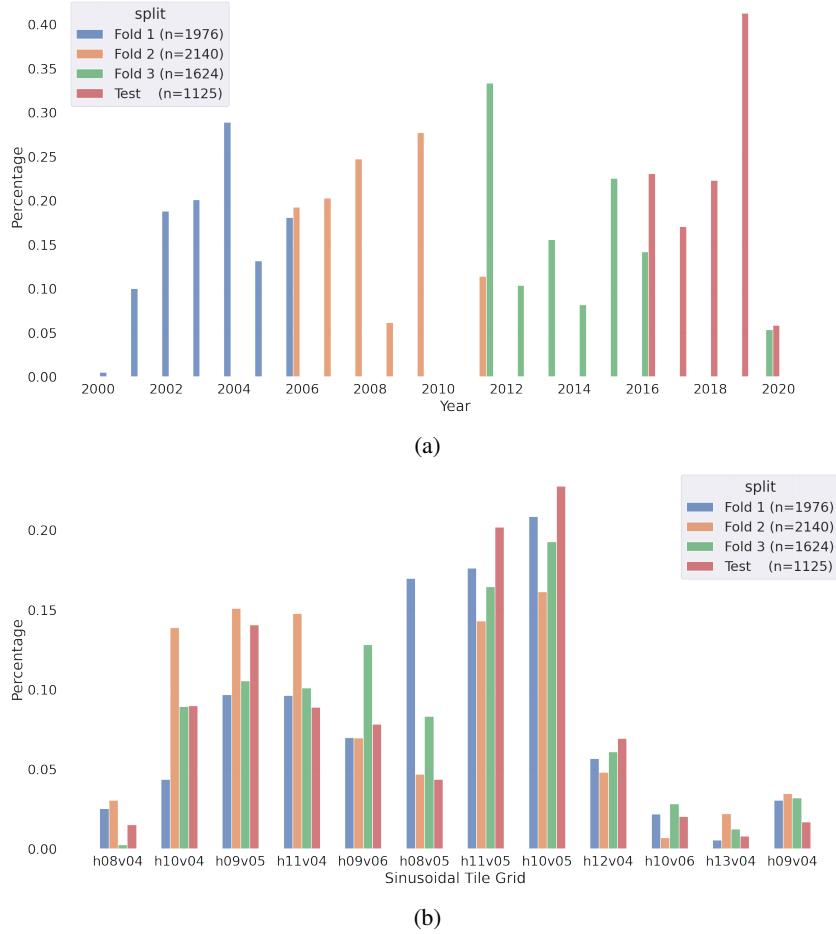


Figure 4: Distribution of splits a) temporally and b) geographically

### A.3 Houston Events Description

Table 1 summarizes the dates of the events used in the Houston case study

Event	Date Range	Caused Flooding?
Hurricane Harvey	2017-08-26 - 2017-09-05	Yes
Storm Nicholas	2021-09-13 - 2021-09-20	Yes
Storm Imelda	2019-09-17 - 2019-09-24	Yes
Hurricane Laura	2020-08-27 - 2020-09-05	No
Storm Rita	2005-09-24 - 2005-10-01	No

Table 1: Summary of events used in Houston case study

### A.4 Satellite Observations During Hurricane Harvey

We visualize satellite observations after Hurricane Harvey’s landfall from the Aqua and Terra satellite constellation (MODIS) as well as predictions from our proprietary satellite segmentation model in Fig. 5. While extreme flooding outside the city is visible four days after the hurricane, it takes 7 days from the event to obtain a clear view of the Houston, at which point damaging flooding may have receded. In contrast, our model will make predictions hourly from the time of the event striking, and within 24 hours it is evident that this storm is a disastrous outlier (Fig. 2).

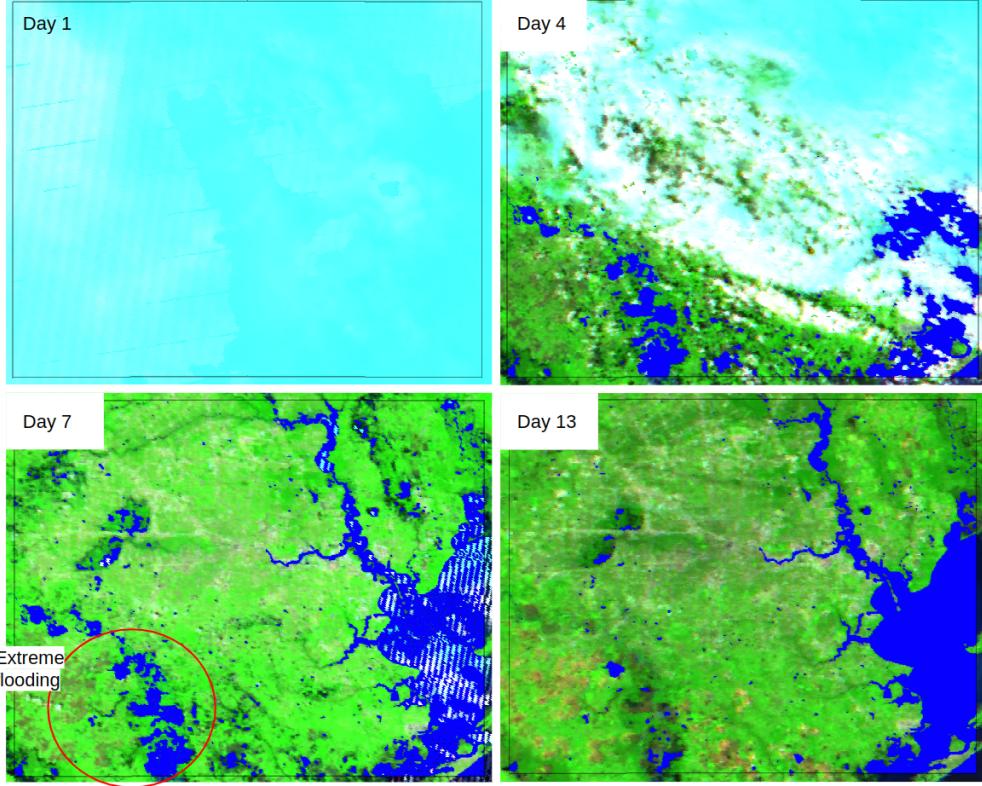


Figure 5: False Colour Composites of Aqua and Terra (MODIS) observations with segmented water overlaid in blue on several days after Hurricane Harvey’s landfall.

### A.5 Proprietary Satellite Segmentation Model

Our proprietary water segmentation model was designed to produce fractional water labels at 250m resolution based on an optical satellite observation. To demonstrate the effectiveness of this model, a test set of 2209 hand-labelled examples were produced from held-out geographies. A summary of the model's performance on this held-out test set is provided in Table 2.

Pixel Subset	MSE (mean +/- std)
Never Flooded Pixels ( $\text{GSW} \in [0, 0.005]$ )	0.014 +/- 0.005
Flooded Before Pixels ( $\text{GSW} \in [0.005, 0.30]$ )	0.021 +/- 0.005

Table 2: Mean Squared Error on the held-out test set