
Memory to Map: Improving Radar Flood Maps With Temporal Context and Semantic Segmentation

Veda Sunkara¹ Nicholas R. Leach¹ Siddha Ganju²

Abstract

Global flood risk has increased due to worsening extreme weather events and human migration into growing flood-prone areas. Accurate, high-resolution, and near-real time flood maps can address flood risk by reducing financial loss and damage. We propose *Memory To Map*, a novel machine learning approach that utilizes bi-temporal context to improve flood water segmentation performance for Sentinel-1 imagery. We show that the inclusion of unflooded context for the area, or “memory,” allows the model to tap into a “prior state” of pre-flood conditions, increasing performance in geographic regions in which single-image radar-based flood mapping methods typically underperform (e.g. deserts). We focus on accuracy across different biomes to ensure global performance. Our experiments and novel data processing technique show that the confluence of pre-flood and permanent water context provides a 21% increase in mIoU over the baseline overall, and over 87% increase in deserts.

1. Introduction

Climate change-driven flood disasters are rapidly increasing in frequency and magnitude (UNDRR, 2015). More people are affected by flooding than any other climate disaster (Hallegatte et al., 2017; CRED, 2019), hindering sustainable development. Research consistently shows that relative property loss caused by floods are highest in places of social vulnerability (UNDRR, 2019; Ceola et al., 2014), exacerbating economic disparities (Tellman et al., 2020). Readily available remote sensing satellite data allows for scalable, low cost, near-real time disaster monitoring and mapping in order to support response and relief. Multidisciplinary approaches to flood detection algorithms, leveraging conventional remote sensing, hydrology, and machine learning, are paramount to using satellite imagery for effective flood detection (Mateo-Garcia et al., 2021).

Imagery from Sentinel-1, ESA’s synthetic aperture radar

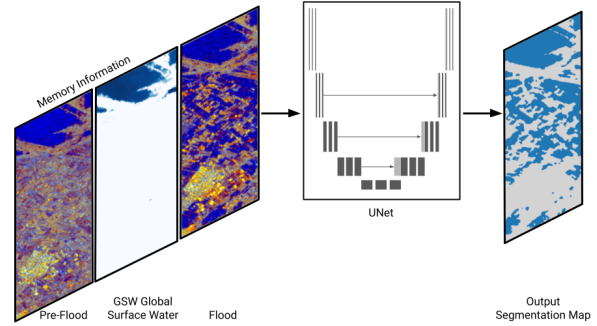


Figure 1. *Memory To Map* inference pipeline. The model combines multi-temporal contextual information, (pre-flood and GSW) with flooded radar imagery to improve segmentation performance for Sentinel-1. We hypothesize that the subtle changes in backscatter between the memory context and the flood radar imagery could potentially be exploited by the model to improve performance.

(SAR) satellite, is commonly used for flood mapping as it is free, easily accessible, cloud penetrating (i.e. the image is not affected by the presence of clouds or other atmospheric conditions), and systemically acquired (Potin et al., 2014). However, the accuracy of SAR-derived flood maps is often reduced due to multiple land surfaces - such as sandy river banks, deserts, recently cleared fields, and terrain shadows - appearing similar to flood waters in backscatter images.

Memory To Map introduces a pre-flood image of the scene and a permanent water layer (derived from the the Joint Research Council Global Surface Water data v1.1 (hereafter “GSW”) (Pekel et al., 2016)) as input channels to the network which provide unflooded context, thereby leveraging change detection when segmenting water. Our approach resulted in significant improvements over the baseline, as well as unique improvements in deserts and other areas that are particularly difficult for flood mapping using SAR. To the best of our knowledge, this is the first work of its type tested globally with accuracy measured across different biomes, motivated by the frequent flooding in different biomes and the necessity of globally performant models. The framework introduced here for training and

validating globally-applicable models is a key contribution of this work. Our approach can also serve as a template for using temporal information in semantic segmentation to detect flooding in SAR images.

We demonstrate a way to utilize accessible data sources to improve semantic segmentation of floods by 21% mean intersection over union (mIoU) overall, creating a model for generating high quality flood maps across domains. We improved performance in uniquely challenging regions such as deserts, in which we found a 87% increase in mIoU, making a flood mapping network useful for global applications. The use of a globally-distributed dataset for training and validation constitutes a major benefit of this model and substantially broadens the applicability of our results. As shown by the reliance of governments and aid agencies on flood maps derived from satellite imagery (Ho et al., 2021), these maps enable better global preparedness for and robust response to increasingly common climate disasters.

2. Related Work

SAR-based flood segmentation approaches often rely on the assumption that inundated areas have lower backscatter than the rest of the scene (Chini et al., 2017; Martinis et al., 2015). While this is often true, inundation may actually increase backscatter compared to unflooded conditions in certain situations, such as flooded vegetation (Tsyganskaya et al., 2018; Zhang et al., 2020). Approaches relying on phase information rather than backscatter have been used for flood detection (Shen et al., 2019), but these methods are often very sensitive to non-target change and can be difficult to implement and interpret.

There are few SAR-based flood segmentation methods that have been demonstrated to be effective globally (Mateo-Garcia et al., 2021; Bonafilia et al., 2020; Rambour et al., 2020). Quantifying our model’s accuracy in different biomes may help users to understand its applicability to specific regions. Many others have discussed performance differences of SAR-based flood mapping approaches in specific geographic areas and conditions (Giustarini et al., 2013; Zhang et al., 2020; Tsyganskaya et al., 2018). The use of RESOLVE biomes (Dinerstein et al., 2017) allows us to analyze chips based on their biogeoclimatic similarity without the need for subjective grouping. Furthermore, we believe the use of biomes as a sampling stratum for developing globally-applicable machine learning models has value beyond the scope of this work, and may be worth consideration for other computer vision tasks which must perform well across many geographic areas.

Recent work in deep learning has leveraged recurrent neural networks for bi-temporal change detection in using satellite data (Lyu et al., 2016; Ma et al., 2019), as well as multimodal inputs that leverage context for semantic label-

ing (Audebert et al., 2018). Our work combines elements of both of these approaches, leveraging a standard segmentation architecture with multimodal temporal information to improve the accuracy of water maps.

3. Memory To Map

We describe the three main contributions of our work:

1. A novel data processing technique that combines flooded, pre-flood, and GSW data to produce state of the art segmentation results.
2. A U-Net training scheme to leverage change information and context to segment floods.
3. Utilization of global biome information to assess balance of model performance across geoclimactic regions, with a focus on desert biomes.

3.1. Dataset

We describe the generation of the three distinct inputs to our network; permanent water from GSW, pre-flood, and flooded chips. We use the term “chips” to indicate 512x512 pixel crops of images, in accordance with domain terminology.

Flooded Chips: Candidate flood events were selected from the Dartmouth Flood Observatory (Brakenridge, 2010) and filtered based on the following criteria: (1) flood events were discarded unless there were Sentinel-1 and Sentinel-2 images within four days of each other for which at least 50% of the Sentinel-2 image overlapped with the Sentinel-1 image. This was required so that Sentinel-2, which carries a multi-spectral sensor sensitive to visible and infrared wavelengths, could be used to support the creation of high-quality hand labels; the Sentinel-2 images were not used in this study beyond the generation of hand labels. (2) The Sentinel-2 image was required to have less than 50% cloud cover. (3) The dataset was balanced to remove over-representation of common biomes (e.g. Tropical & Subtropical Grasslands, Savannas & Shrublands) by removing flood events based on their biome (Dinerstein et al., 2017). These flood events were then divided into 512x512 chips from a total of 125 flood events. For each chip, a team of domain scientists hand-labeled segmentation masks by cross-referencing the Sentinel-1 and corresponding Sentinel-2 chips.

Pre-Flood Chips: To generate pre-flood chips, we conducted a search based on the original Sentinel-1 image for each flooded chip. We searched for a pre-flood Sentinel-1 image meeting three criteria: (1) the pre-flood image must be acquired from the same orbit direction as the flood image from which the flooded chip was derived (i.e. the ascending or descending pass); (2) the image must be acquired from the same relative orbit number; and (3) the pre-flood

image must be acquired 14-35 days prior to the flooded image. This date range was chosen to allow for at least one Sentinel-1 same-orbit acquisition while exceeding the global median flood duration to increase the likelihood that the image was acquired in an unflooded condition (Najibi & Devineni, 2017). If multiple images meet all three of these criteria, the image acquired nearest in time to the flooded image was selected as the pre-flood image. If no images meeting these criteria were found, the chip was discarded. Out of the total 3513 available chips, pre-flood images for 2250 of them were found and the others were not used.

Permanent Water Chips: For each chip, we created a permanent water chip based on GSW data. To do so, we cropped the Global Surface Water occurrence data to the extent of each of our Sentinel-1 chips. The occurrence data was then resampled to 10-meter resolution.

3.2. Training Methodology

We had 2250 total chips from 125 flood events, divided into 1820 training and 430 test. We devised testing and training splits by testing random seeds to optimize for even distribution of flood water per biome and chips per flood event. We used an Adam optimizer and trained on GeForce RTX 3090 GPU. We modify the U-Net (Yakubovskiy, 2020) to accept additional inputs by increasing the number of input channels and train with a batch size of 32, initial learning rate of 0.001, patience of 8, random cropping data augmentations, pretrained timm_efficientnet_b1 backbone (Wightman, 2019), and test with an identical network.

4. Results

We trained and benchmarked our network with the flooded chip, flooded and GSW chips, flooded and pre-flood chips, and lastly flooded, GSW, and pre-flood chips. As shown in Table 1, the network performs best overall when supplied with both the GSW and pre-flood chips. We measured model performance both with overall mIoU and overall F1 Score.

Model Inputs	Flooded	Flooded Pre-Flood	Flooded GSW	Flooded GSW Pre-Flood
mIoU	0.62	0.66	0.72	0.75
F1 Score	0.77	0.80	0.84	0.87

Table 1. Model performance across input channels. The model that includes flooded, GSW, and pre-flood chips perform consistently better than models with fewer inputs. Results tallied for all biomes.

To assess geographic variation in model performance, we assigned each chip a biome based on the RESOLVE biome corresponding to the majority of its area. Of the biomes included, those with the lowest average Sentinel-1 backscatter are Deserts & Xeric Shrublands (henceforth Deserts) and

Tropical & Subtropical Grasslands, Savannas & Shrublands. As shown in Figure 2, just including a pre-flood chip did not improve model performance in these biomes; however including GSW data substantially improved model performance in Deserts. The model which included the flooded, pre-flood, and GSW chips was the best-performing model in the majority of the evaluated biomes. For a selection of predictions for each biome, see Appendix Figure 4.

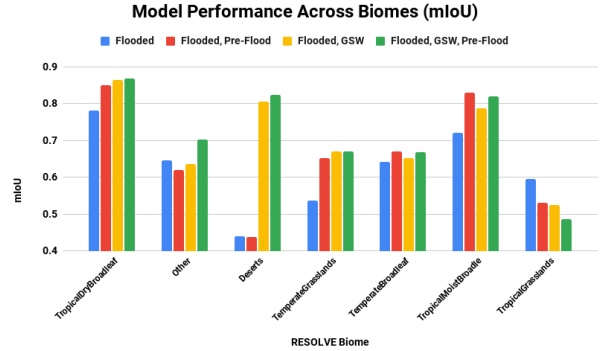


Figure 2. Comparison of four models across different RESOLVE biomes. The biomes are arranged from least to most represented in the data². The model trained on flooded, GSW, and pre-flood chips performed better on Deserts than the model trained on just flooded chips by 87% mIoU.

We sought to improve model performance in uniquely challenging biomes for SAR such as Deserts. To assess this, we trained all four of our networks on only Deserts, all biomes except Deserts, and all biomes, and in all cases tested only on Deserts. As shown in Figure 3, networks trained on only Deserts had the best performance, while networks with flooded and GSW and pre-flood, GSW, and flooded chips performed comparably well when trained on all biomes. The network provided with pre-flood, GSW, and flooded chips trained on all biomes except Deserts had the best performance compared with models with fewer inputs.

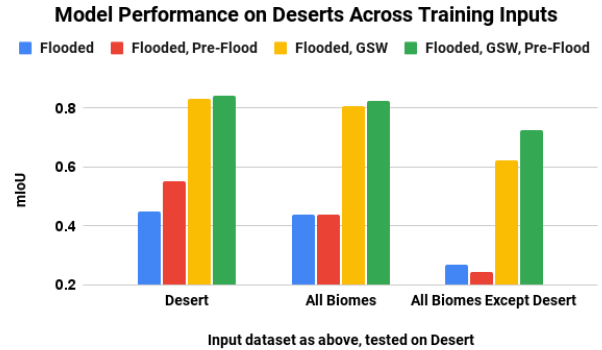


Figure 3. Four models trained on Deserts, every biome except Deserts, and every biome, and tested on Deserts.

5. Analysis

Given the challenges with segmenting water from Sentinel-1 SAR imagery, we developed a segmentation network and algorithm that leveraged context to generate better flood segmentation masks at the high-resolution scale needed to support disaster resilience. We performed a novel cross-biome global analysis, and found that our method significantly improved network performance in uniquely challenging low backscatter regions such as deserts.

While we found that simply providing the model with permanent water data significantly improved performance over segmentation with a single flooded chip, providing permanent water and a pre-flood chip provided the best results. This is possibly because the model is able to leverage the combination of the permanent water mask and the visual representation of the permanent water state (i.e. the pre-flood chip) to then better interpret unique water/land boundaries.

Notably, we found that providing the network with the pre-flood chip and flooded chip only slightly improved performance over simply providing the flooded chip. This is likely due to the network requiring more context for the distinction between the two input chips.

5.1. Biome Analysis

As expected, the models which only included a flooded chip performed poorly in biomes with low average backscatter (Figure 2). For Desert chips, models which included GSW data performed better, and the model which included both GSW and a pre-flood chip was the best performing. The inclusion of the GSW data is likely helping the model distinguish between deserts and open water, which have similar backscatter intensities. These results suggest that information about permanent water is important for SAR-based water segmentation in areas with low average backscatter.

Models provided with flooded, pre-flood, and GSW data, when trained on all biomes except Deserts, perform best on Deserts compared to models with fewer inputs (Figure 3). The models with both forms of context are best able to generalize to unseen biomes. This suggests that, when developing a network with a limited amount of training data representing a key challenge area, providing pre-flood and GSW context together can boost performance. The ability to generalize is necessary for deployment, especially when training data from the main application or specific

geographic regions may be missing.

The approach described in this paper has some important limitations. Generating context data for each chip involves extra data preparation and processing. Selecting appropriate pre-flood images requires some assumptions about the duration of floods that may not always be true; for example, a “pre-flood” image acquired 14 days before the flooded image may still contain flooding. Optimal selection of pre-flood images may also vary among geographic areas and flood types. Additionally, this model is only usable if there is a valid pre-flood image available; simply using GSW yields significant performance boosts and is universally accessible.

As the results of this study show, the inclusion of context did not always improve the model’s performance; in Tropical & Subtropical Grasslands, Savannas & Shrublands, the model which included both GSW and pre-flood chips performed worse than the model with only flooded chips. Changes in the backscattering properties of these grasslands due to rainfall create spatially-varying changes in the backscatter between the pre-flood and flooded images, potentially explaining this dip in performance.

6. Discussion and Conclusion

Memory To Map is a prototype for utilizing contextual information to improve the performance of semantic segmentation of flood water in Sentinel-1 SAR Imagery. Improving the accuracy and reliability of these flood maps allows for near-real time flood monitoring even in adverse weather conditions, a common precedent to extreme flood events. These near-real time maps can then be used for disaster relief and response, as well as for planning for necessary infrastructure for climate resilience. Our novel biome-based analysis demonstrates the applicability of this methodology to global domains. Providing context to segmentation networks results in a significant performance boost in areas with relatively low backscatter, such as deserts, making the algorithm globally performant.

Future work should include a more rigorous exploration of context inputs. To emphasize change detection based on seasons, a pre-flood image could be constructed as a pixelwise statistical measure (such as average or z-score) of backscatter values taken at even intervals across the previous year. Additional contextual information can also be constructed based on publicly available datasets, such as pixelwise masks containing biome, land cover, agricultural, and population density information. Topographic information could also provide the network with context for water drainage and movement patterns over time. As the successes of this approach vary across biomes, using an ensemble approach to decide which model is best for detecting flood water in a scene may provide the best results globally.

²The biome labels refer to Tropical & Subtropical Dry Broadleaf Forests (59 chips); any other biomes (110 chips); Deserts & Xeric Shrublands (128 chips); Temperate Grasslands, Savannas & Shrublands (192 chips); Temperate Broadleaf & Mixed Forests (389 chips); Tropical & Subtropical Moist Broadleaf Forests (633 chips); and Tropical & Subtropical Grasslands, Savannas & Shrublands (731 chips).

Reliable and high-resolution flood detection methods allow for the development of effective disaster relief systems. These system can provide critical services to give governments and aid organizations the resources to rebuild, replant crops, access international aid, and more. The described novel data processing technique produces segmentation models that are performant regardless of biogeography and generalize effectively, creating consistent flood maps that can be deployed to mitigate harm to climate vulnerable communities, regardless of their location.

References

- Audebert, N., Le Saux, B., and Lefèvre, S. Beyond rgb: Very high resolution urban remote sensing with multi-modal deep networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 140:20–32, 2018. ISSN 0924-2716. doi: <https://doi.org/10.1016/j.isprsjprs.2017.11.011>. URL <https://www.sciencedirect.com/science/article/pii/S0924271617301818>. Geospatial Computer Vision.
- Bonafilia, D., Tellman, B., Anderson, T., and Issenberg, E. Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.
- Brakenridge, G. R. Global active archive of large flood events. *Dartmouth Flood Observatory, University of Colorado*, 2010.
- Ceola, S., Laio, F., and Montanari, A. Satellite nighttime lights reveal increasing human exposure to floods worldwide. *Geophysical Research Letters*, 41(20):7184–7190, 2014. doi: 10.1002/2014gl061859.
- Chini, M., Hostache, R., Giustarini, L., and Matgen, P. A hierarchical split-based approach for parametric thresholding of sar images: Flood inundation as a test case. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12): 6975–6988, 2017. doi: 10.1109/TGRS.2017.2737664.
- CRED. 2018: Extreme weather events affected 60 million people. Centre for Research on the Epidemiology of Disasters CRED, Jan 2019. URL <https://www.cred.be/2018-review-disaster-events>.
- Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N. D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., and et al. Ecoregion-based approach to protecting half the terrestrial realm, Apr 2017. URL <https://academic.oup.com/bioscience/article/67/6/534/3102935>.
- Giustarini, L., Hostache, R., Matgen, P., Schumann, G. J.-P., Bates, P. D., and Mason, D. C. A change detection approach to flood mapping in urban areas using TerraSAR-x. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4):2417–2430, April 2013. doi: 10.1109/tgrs.2012.2210901. URL <https://doi.org/10.1109/tgrs.2012.2210901>.
- Hallegatte, S., Vogt-Schilb, A., Bangalore, M., and Rozenberg, J. *Unbreakable: building the resilience of the poor in the face of natural disasters*. World Bank, 2017.
- Ho, J. C., Vu, W., Tellman, B., Dinga, J. B., N’diaye, P. I., Weber, S., Bauer, J.-M., Schwarz, B., Doyle, C., Demuzere, M., Anderson, T., and Glinskis, E. Chapter 6 - from cloud to refugee camp: A satellite-based flood analytics case-study in congo-brazzaville. In Schumann, G. J.-P. (ed.), *Earth Observation for Flood Applications*, Earth Observation, pp. 131–146. Elsevier, 2021. ISBN 978-0-12-819412-6. doi: <https://doi.org/10.1016/B978-0-12-819412-6.00006-7>. URL <https://www.sciencedirect.com/science/article/pii/B9780128194126000067>.
- Lyu, H., Lu, H., and Mou, L. Learning a transferable change rule from a recurrent neural network for land cover change detection. *Remote Sensing*, 8(6), 2016. ISSN 2072-4292. doi: 10.3390/rs8060506. URL <https://www.mdpi.com/2072-4292/8/6/506>.
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B. A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152:166–177, 2019. ISSN 0924-2716. doi: <https://doi.org/10.1016/j.isprsjprs.2019.04.015>. URL <https://www.sciencedirect.com/science/article/pii/S0924271619301108>.
- Martinis, S., Kersten, J., and Twele, A. A fully automated terrasarsar-x based flood service. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104:203–212, 2015. ISSN 0924-2716. doi: <https://doi.org/10.1016/j.isprsjprs.2014.07.014>. URL <https://www.sciencedirect.com/science/article/pii/S0924271614001981>.
- Mateo-Garcia, G., Veitch-Michaelis, J., Smith, L., Oprea, S. V., Schumann, G., Gal, Y., Baydin, A. G., and Backes, D. Towards global flood mapping onboard low cost satellites with machine learning. *Scientific Reports*, 11(1), 2021. doi: 10.1038/s41598-021-86650-z.
- Najibi, N. and Devineni, N. Recent trends in frequency and duration of global floods. *Earth System Dynamics Discussions*, pp. 1–40, 07 2017. doi: 10.5194/esd-2017-59.

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

Potin, P., Rosich, B., Roeder, J., and Bargellini, P. Sentinel-1 mission operations concept. In *2014 IEEE Geoscience and Remote Sensing Symposium*. IEEE, July 2014. doi: 10.1109/igarss.2014.6946713. URL <https://doi.org/10.1109/igarss.2014.6946713>.

Rambour, C., Audebert, N., Koeniguer, E., Saux, B. L., Crucianu, M., and Datcu, M. Flood detection in time series of optical and sar images. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020:1343–1346, 2020. doi: 10.5194/isprs-archives-xliii-b2-2020-1343-2020.

Shen, X., Anagnostou, E. N., Allen, G. H., Robert Brakenridge, G., and Kettner, A. J. Near-real-time non-obstructed flood inundation mapping using synthetic aperture radar. *Remote Sensing of Environment*, 221:302–315, 2019. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2018.11.008>. URL <https://www.sciencedirect.com/science/article/pii/S0034425718305169>.

Tellman, B., Schank, C., Schwarz, B., Howe, P. D., and de Sherbinin, A. Using disaster outcomes to validate components of social vulnerability to floods: flood damage and property damage across the usa. SocArXiv, Jun 2020. doi: 10.31235/osf.io/byrgu. URL osf.io/preprints/socarxiv/byrgu.

Tsyganskaya, V., Martinis, S., Marzahn, P., and Ludwig, R. Detection of temporary flooded vegetation using sentinel-1 time series data. *Remote Sensing*, 10(8):1286, 2018. doi: 10.3390/rs10081286.

UNDRR. The human cost of weather-related disasters 1995–2015. Centre for Research on the Epidemiology of Disasters CRED, 2015.

UNDRR. *Global Assessment Report on Disaster Risk Reduction 2019*. 05 2019. ISBN 978-92-1-004180-5.

Wightman, R. Pytorch image models. <https://github.com/rwightman/pytorch-image-models>, 2019.

Yakubovskiy, P. Segmentation models pytorch. https://github.com/qubvel/segmentation_models.pytorch, 2020.

Zhang, M., Chen, F., Liang, D., Tian, B., and Yang, A. Use of sentinel-1 grd sar images to delineate flood extent in pakistan. *Sustainability*, 12(14), 2020. ISSN 2071-1050. doi: 10.3390/su12145784. URL <https://www.mdpi.com/2071-1050/12/14/5784>.

7. Appendix

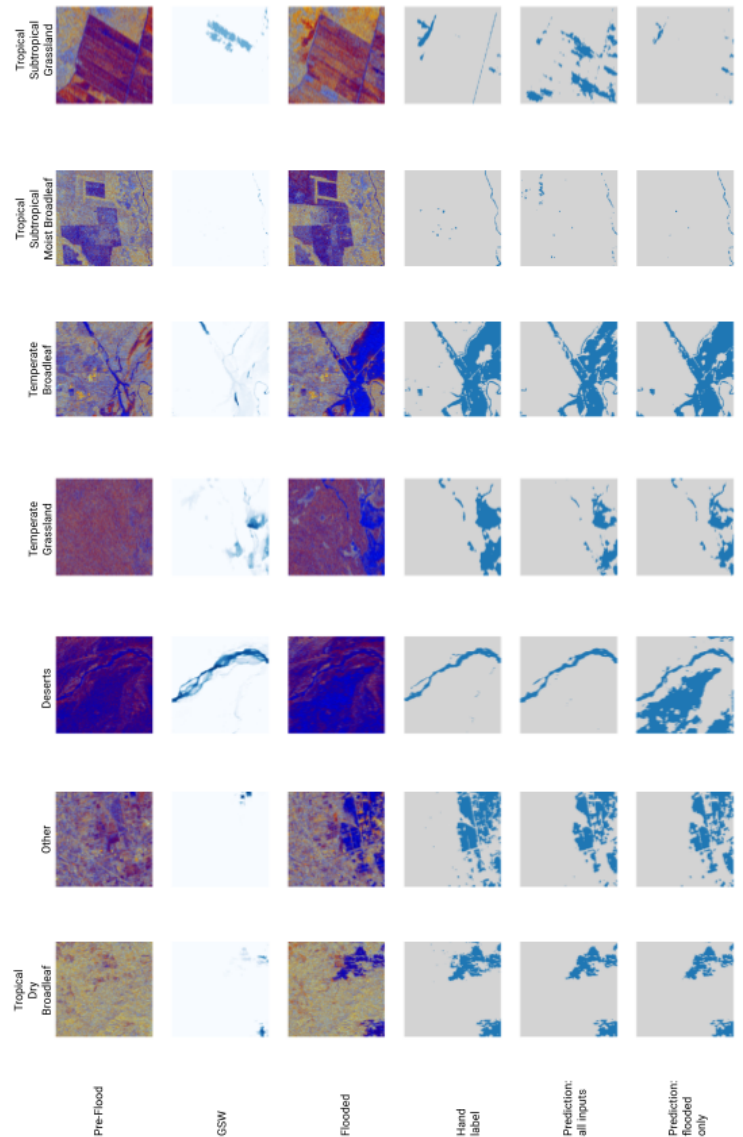


Figure 4. We provide a few visual examples in Figure 4 of the segmentation outputs for various biomes, from the best performing model with inputs of all three flooded, pre-flood, GSW, and the baseline model with flooded images as inputs.