Sommaire

Global Flood Extent Mapping from Optical Satellite Imagery with Cloud Handling

Floods are one of the most devastating natural hazards, affecting millions of people each year, especially in low- and middle-income countries. Accurate mapping of their extent from satellite imagery is crucial for planning emergency response and quantifying damage. However, current approaches have limitations in terms of spatial resolution, detection accuracy, and handling of omnipresent clouds on post-event scenes. This project aims to develop a complete deep learning solution for flood mapping that addresses these major challenges.

1 Introduction

Floods are among the most devastating and frequent natural disasters. Between 1995 and 2015, 2.6 billion people were affected by floods, representing 56% of those exposed to weather-related hazards. A large portion (89%) live in low- or middle-income countries where the consequences often lead to food crises. Climate projections indicate that the frequency and intensity of flood events will continue to increase.

After a major disaster, mapping flooded areas from satellite imagery is crucial for two main reasons :

1. Planning field emergency operations by identifying priority areas for rescue..
2. Assessing damage by cross-referencing flood extent with other geospatial data (population, infrastructure, agricultural areas, etc.).

However, current global operational flood mapping approaches have several limitations :

1.1 Spatial Resolution

Most products currently used to cover large areas like major floods come from low or medium spatial resolution sensors (100m) such as MODIS or VIIRS. At these resolutions, many details are missed, especially in dense urban areas where estimates of exposed population are highly inaccurate.

1.2 Detection Accuracy

Most automated methods used operationally are based on spectral indices like NDWI that exploit the spectral properties of water surfaces. Although used for a long time, these indices suffer from numerous sources of error : dark surfaces, shadows, debris/pollutants in the water, flood traces, etc. More accurate deep learning methods lack generalization to different geographic areas.

1.3 Cloud Cover

About 50% of Sentinel-2 satellite acquisitions after a flood event are partially or totally cloud-covered, preventing any analysis of the land surface. Most existing methods fail to handle these cloudy scenes, which are critical in crisis management.

2 Objectives

To address these limitations, this project aims to develop an end-to-end system to produce accurate and usable flood extent maps from optical satellite images, without human intervention. The main objectives are :

1. Develop a deep learning semantic segmentation model capable of accurately detecting water surfaces, including under semi-transparent clouds.
2. Train this model on a vast global dataset covering multiple eco-regions to maximize its generalization.
3. Set up a complete production pipeline integrating data download, inference, and post-processing of results to generate high-resolution (10-30m) vectorized flood maps.
4. Deploy this system on recent major flood events to demonstrate its large-scale operational applicability.
5. Couple flood maps with other geospatial data to estimate damage to infrastructure, agricultural areas, and affected populations.

3 Training Data

3.1 Extended WorldFloods

To train a deep learning semantic segmentation model, large amounts of annotated training data are required. This project builds on the public WorldFloods dataset, composed of approximately 180,000 256x256 pixel tiles from Sentinel-2 images, along with reference masks of flood extent.

This initial dataset has been significantly expanded as part of this project :

— Addition of new flood events occurring between 2019 and 2023.

— Removal of low-quality (very cloudy or poorly annotated) maps.

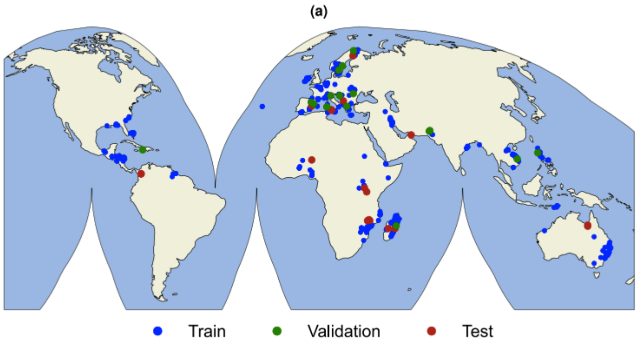
— Manual review of over 200 maps to correct errors.

— Creation of new validation and test sets with strong spatial variability for robust evaluation.

In the end, the extended WorldFloods dataset comprises approximately 75,000 256x256 pixel training tiles covering more than 180 flood events distributed worldwide.

Dataset	Flood events	Flood maps	256x256 patches	Permanent water (%)	Flood water (%)	Land (%)	Thick cloud (%)	Thin cloud (%)
Test	11	18	4703	13.82	15.81	56.01	11.04	3.32
Val	14	16	3524	18.62	2.02	68.01	8.25	3.10
Train	119	475	65582	3.56	2.28	66.25	23.90	4.01
All	144	509	73809	7.95	3.36	59.25	24.91	4.54

(a)



(b)

FIGURE 1 – (a) Statistics of the extended WorldFloods dataset. (b) Locations of flood events with color codes for training and validation.

3.2 Cloud Handling

Unlike the initial version, the new version of the dataset specifically handles the presence of clouds, which can severely obstruct visibility on post-event scenes. During data annotation, two independent binary masks were produced :

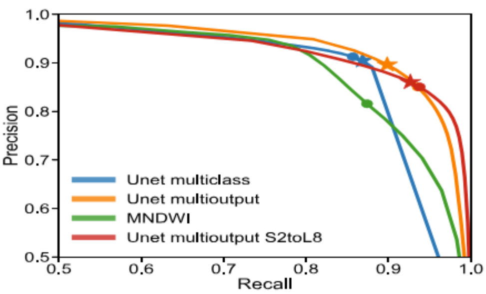
— A cloud/clear-surface mask.

— A water/emerged-surface mask.

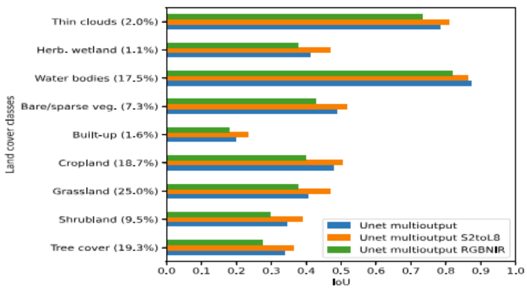
This "multi-kernel" strategy allows for separate handling of the two segmentation problems, while preserving maximum information on the scene when clouds are semi-transparent and the surface remains partially visible.

Architecture	Average Recall flood water	Average Precision flood water	Average IoU flood water	Average Recall permanent	Average Precision permanent	Average IoU permanent
MNDWI	55.36	66.71	40.95	54.11	98.76	53.68
Unet multiclass	**79.80**	54.02	48.27	74.61	90.67	70.20
Unet multioutput	75.71 ±2.16	63.45±9.57	52.67±6.34	70.81 ±2.74	98.43±1.01	69.95±2.95
Unet multioutput S2-to-L8	75.48±1.49	**70.81** ±3.07	**57.89**±1.43	68.46±2.86	99.35±0.28	68.19±2.70
Unet multioutput RGBNIR	72.42 ±3.14	60.18±3.12	48.57 ±1.27	69.07±3.15	94.27±1.85	65.80±1.96

(a) Performance metrics obtained in the evaluation set.



(b) ● Default threshold (0.5) ★ Highest IoU threshold.



(c) IoU water metrics stratified by land cover type.

4 Model Development :

A pre-trained deep learning U-Net++ model was developed to perform semantic segmentation of satellite images into three classes : water, clouds, and emerged surfaces. This model was trained on the extended WorldFloods dataset using the two independent binary masks as targets.

Different model variants were produced by varying the input bands. The best model, called UNet S2-to-L8, uses 6 bands : the visible and near-infrared bands, as well as the two shortwave infrared (SWIR) bands, which are particularly useful for water detection. This variant has the advantage of being applicable to both Sentinel-2 and Landsat 8/9 satellite images.

As output, this model produces two probability maps : one for clouds and one for water. Post-processing then combines these two maps with the scene brightness to generate a final segmentation mask containing the three desired classes. An adjustable brightness threshold allows controlling whether semi-transparent clouds should be considered as opaque clouds (unobservable surface) or as visible surfaces.

The input data is loaded, including the Sentinel-2 image, the ground truth mask, and the JRC permanent water data. The spectral bands used are defined by the configuration.

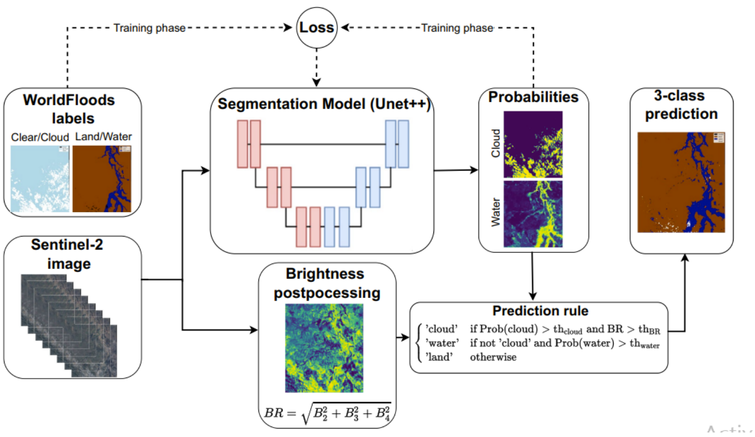


FIGURE 2 – The Multioutput binary model training and inference fowchart

5 Case Study : Floods in Pakistan and Australia

Provide a detailed analysis of flood events in Pakistan and Australia.

The trained UNet++ model is loaded and set to evaluation mode for inference. If a GPU is available, the model is transferred to the GPU.

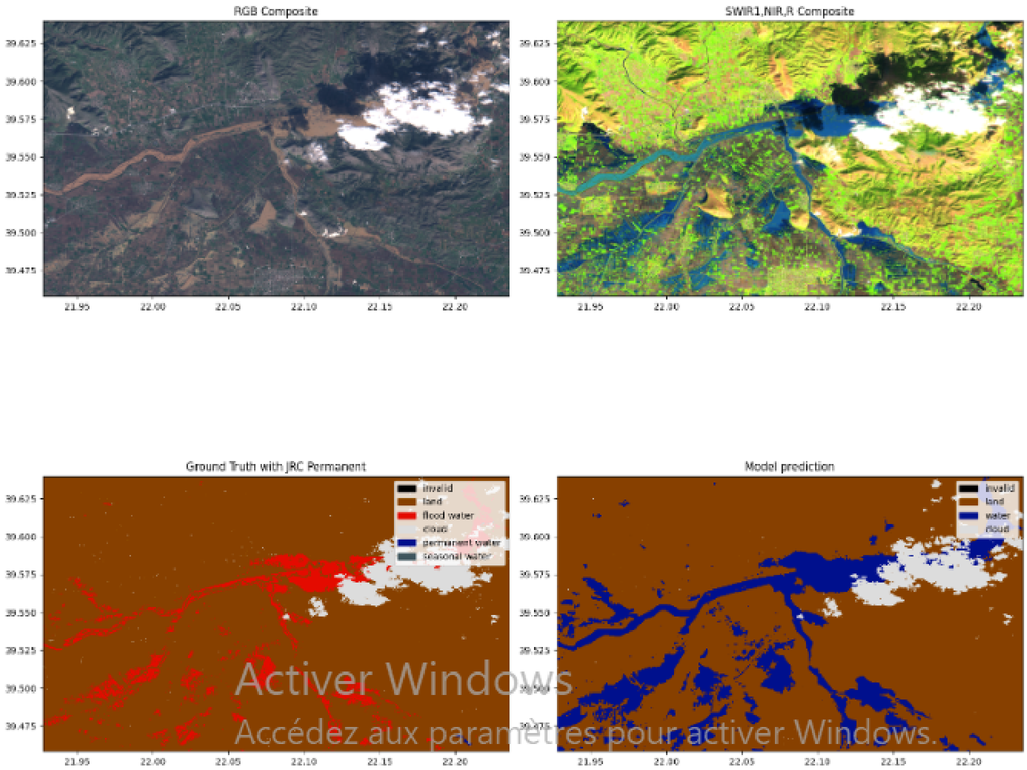


FIGURE 3 – Multioutput binary model predictions in evaluation images.

The model performs inference on the input image to produce the water and cloud probability maps. Post-processing is applied to obtain a binary water mask by thresholding the water probability map. Then, water polygons are extracted from this mask.

After that, the results are visualized : the input image, the ground truth mask, the permanent water mask, and the model's prediction are displayed side by side. The predicted water polygons are also plotted on a map.

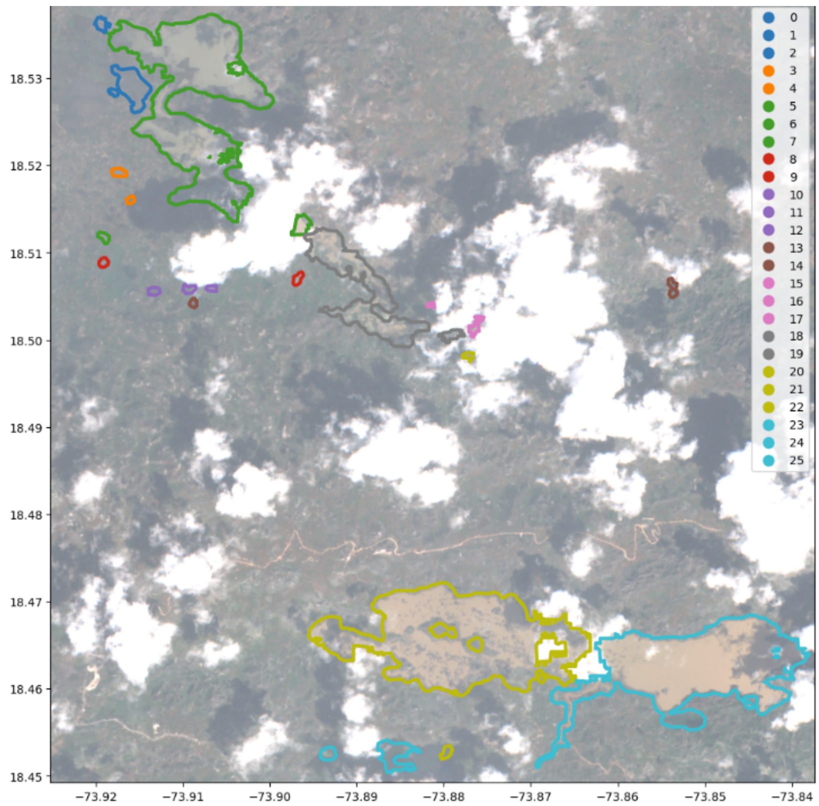


FIGURE 4 – Vectorise the water masks

6 Evaluating Model Performance :

code	Recall_land	Recall_water	Recall_cloud	IoU_land	IoU_water	IoU_cloud
EMSR286 (Colombia)	95.775230	96.068488	98.995678	94.447091	85.008647	96.262742
EMSR333 (Italy)	90.770697	86.519387	99.119795	89.944669	59.832617	73.401532
EMSR342 (Australia)	69.547385	97.030021	91.994864	67.459669	72.339948	89.279442
EMSR347 (Malawi)	98.847444	81.699537	95.940271	96.859947	76.400101	91.132286
EMSR284 (Finland)	90.794550	97.332282	100.000000	89.959569	76.971944	0.000002

FIGURE 5 – Mean values across flood events

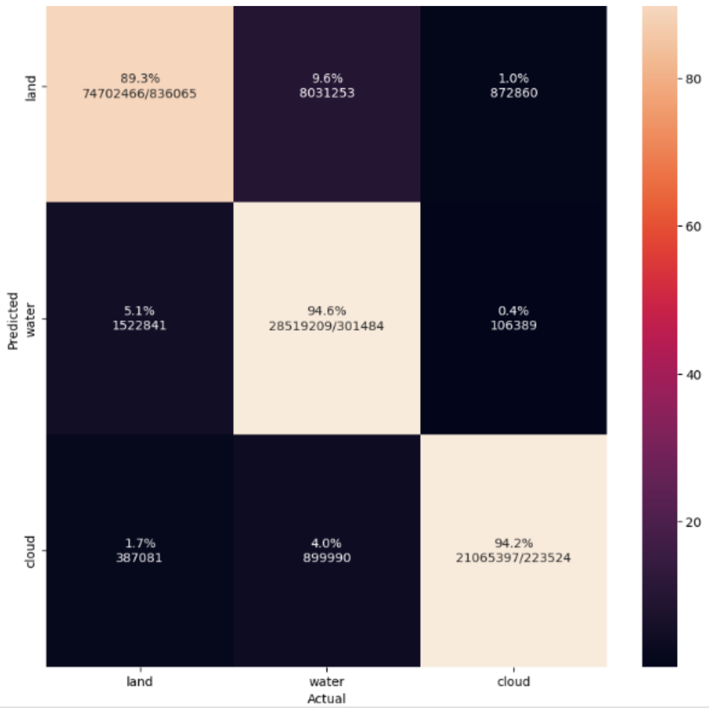


FIGURE 6 –