STEVEN GONG

A Novel Machine Learning Approach to Automating Global Flood Forecasting

There are several types of flood

The fundamental flood forecasting techniques can be broken into two parts: measuring or computing the discharge of the river, and modelling the water’s hydraulic behavior in order to generate an inundation map. The discharge of the river is an extremely important measure since it is one of the true indicators of flooding and directly affects the water’s hydraulic behavior. To obtain this data, stream gauges are used, but in many parts of the world, these instruments are either inexistent or out of use, forcing scientists to use other methods to compute the discharge, what is called a hydrological model.

This has traditionally been done by building an extremely complex representation of the physical processes in the environment and see how that affects the flow of the river (factors such as precipitation and soil type take a great importance (insert image). This would be ideal given we have enough data, as we have a fairly good understanding of these physical processes now. However, in many regions across the world, the scarcity of data makes this an extremely difficult task. For this reason, the implementation machine learning techniques shows great promise in resolving this issue, as it is extremely good at approximating functions without requiring the complex and exact parameters that the traditional models require in order to compute the output.

The integration of a machine learning pipeline can also bring extraordinary value to the second part, now that the discharge data has been computed, scientists need to represent how that water will be affecting the area. Traditionally, satellite imagery has been used in order to generate Digital Elevation Maps (DEMs), which provide an overview of the topography of the Earth. A hydraulic model can then be created to model the way water navigates the ground and map out which areas are at risk of flooding. However, traditional DEMs are known to be outdated and low in resolution, since they require specialized missions which means a great cost and cannot be funded to be updated every year. Part of this paper shows how the implementation of multi-asset spatial alignment (MASA) using machine learning in order to create a DEM of a much higher resolution.

Finally, machine learning can start bringing automation to the process. As the amount of data doubles every year, climate scientists struggle to keep up with traditional methods which are very time consuming. Big Data analytics has emerged in the last decade. This paper shows the integration of parallel computing and cloud technologies to facilitate the calculations of flood forecasting.

The paper will also demonstrate the machine learning used for modelling coastal flooding is also a very similar process, except that instead of using discharge data, which is often useless when it gets to the scale of the ocean, scientists use sea level elevation maps.

Part 1: Data Download

All three datasets that have been used to create the model are open-sourced. The first dataset I used is ERA5, which provides hourly data on many atmospheric, land-surface and sea-state parameters together with estimates of uncertainty. This dataset is used in the intent for the model to see the meteorological conditions such as precipitation which has shown to have a great impact on flooding.

The second dataset is the Global Flood Awareness System (GloFAS), which provides global modelled daily data of river discharge. These two datasets have been downloaded from the Climate Data store. <http://datastore.copernicus-climate.eu/c3s/published-forms/c3sprod/cems-glofas-historical/GLOFAS_CDS_PAPER_2020.pdf>

Part 2: Data Preprocessing

As soon as we started this project, we knew that it was imperative that we used libraries that could support parallel computing. Xarray was the perfect dataset for that. It was created for . Alongside Xarray, we used Dask.

The Data Preprocessing was fairly simple. All we needed to do was reshape the datasets to be aligned in the same time and coordinate system. Then, we split the dataset into different time periods. To verify the accuracy of our model later on, it was important during this step to isolate specific cases before creating a global model. For this reason, we selected Elbe and Danube, which have seen flooding precedently in Europe in *2016*.

Part 3: Training and choosing the appropriate hydrological model

We experimented which many models in order to create the model. The first instinct to use a long-short term-memory (LSTM), which are known to be extremely good in forecasting trends.

Part 4 : Building a hydraulic model

Building a hydraulic model was one of the hardest parts, since it forced us to link our discharge and link it directly to the map. There was little information out there, as most forecasts which were open source focused on events and not the spatial-temporal characteristics of the flooding itself.

Part 4.5: Scaling the models globally

Scaling up required the use

Part 5: Implementing an online interactive website for ease of information

The Google Earth Engine was used for its wide adoption, long documentation and supportive community. Link the pieces together was a bit hard tho

Part 6: Implementing a global alert system

Finally, as the final part of the pipeline, it was important that a live feature would also be implemented.