**A Novel Deep Learning Approach to Automated Global Flood Forecasting**

**Jun Yu Gong, Marianopolis College, Montreal, Quebec, Canada**

ABSTRACT

Between 1995 and 2015, over 2.3 billion people have been affected by flooding events across the world (Wallamacq et al. 2015). To mitigate the effects of these types of natural catastrophes on the human population, scientists employ two different forecasting techniques: they build statistical systems that utilize the return period as the identification method of flash flood disasters, or create complex hydrological models that use a list of physical processes to model the behavior of a given river (Cloke and Pappenberger 2009). However, while each of these techniques have their advantages, they are also flawed. Firstly, while probabilistic models are fairly easy to set up, they are unable to incorporate exogenous variables into the equation. On the other hand, while hydrological models offer the ability to generate detailed flooding simulations, they require enormous amounts of measurements and careful tuning from hydrologists in order to offer an accurate riverine simulation. The implementation of a deep learning technique can bring the best of both worlds because of their ability to incorporate external variables for forecasting without the costly setup requirements that are necessary to simulate accurate flood maps. The project involves the use of Big Data technology to offer worldwide flood prediction in the goal of providing many parts of the world with timely flooding information to local populations who are notified through a live email-alert system and can be visualized through an online website made with the Google Maps API. The design has demonstrated improvements to current ensemble prediction systems (EPS) made by the Global Flood Awareness System (GloFAS) 30-day forecasting models. This research provides new insight into the ways Deep learning and Big Data techniques can be used to provide the human population with flood forecasts and inundation models.

Key Words: Flood forecasting, Deep Learning, Big Data, Alert System, Long Short-Term Memory, Hydrological Model,

1. INTRODUCTION

Floods are the most common and devastating type of natural disasters in the world. A recent report showed that in the last decade, over 2.3 billion people have affected by flooding events (Wallamacq et al. 2015), and floods’ occurences represent 43% of all natural disasters[[1]](#footnote-1). For this reason, it is paramount that we develop innovative technologies to mitigate the effects of this devastating natural disaster. Early warnings can prevent up to 43% of fatalities, and 35% of economic damages due to flooding[[2]](#footnote-2), as anticipated floods allow for precautions to be taken and people to be warned so that they can be prepared in advance for flooding conditions, avoiding potential fatalities. As part of this process, accurate models needs to be created in order to properly warn individuals of the danger in their area.

The traditional flood forecasting techniques can be broken down 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. In the first part, a hydrograph can be built from the obtained discharge data which allow scientists to visualize how an area (ex: basin, river, etc.) reacts to events such as a period of rainfall. To obtain these values, 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, mainly by building a hydrological model.

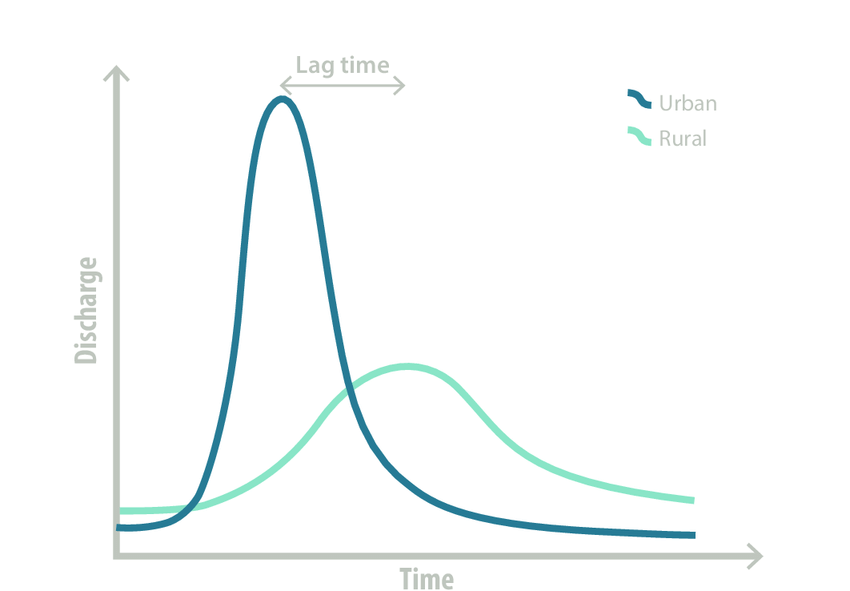


Figure 1. Hydrograph

This extremely complex representation of the physical processes in the environment gives us a detailed overview of how factors such as precipitation and soil type affects the flow of the river (insert image). This would be ideal given we have enough data, as scientific advances have given us a fairly good understanding of these physical processes. However, in many regions across the world, the scarcity of instruments and stations to directly measure the different values[[3]](#footnote-3) makes this an extremely difficult task. For this reason, the implementation machine learning techniques shows great promise in not only resolving the problem of computing complex discharge values, but also delivering accurate future forecasts, 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. This paper will demonstrate the implementation of a Long Short-Term Memory Model (LSTM) which has been proven to work well on forecasting time-series data which has been compared to the Global Flood Awareness System (GloFAS).

The second part of flood forecasting is generating a flood map which delineates the areas which will be affected. 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. This paper shows a simple yet novel methodology which removes the need to build a hydraulic model by building a pixel-precise deep learning forecasting model of the areas which are going to be flooded.

Finally, this paper shows how machine learning and cloud technologies can start bringing automation to the flood forecasting 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.

1. METHODS

*2.1. Data extraction*

All three datasets that have been used in this project are open-sourced and can be accessed through the Climate Data Store after the creation of an account. The first dataset we 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, and thus will act as the predictand for the model..

The second dataset is the Global Flood Awareness System (GloFAS), which provides global modelled daily data of river discharge, simulated by forcing the hydrological “river routing model with modelled gridded runoff data from global analysis” .

Uploading the dataset used

<http://datastore.copernicus-climate.eu/c3s/published-forms/c3sprod/cems-glofas-historical/GLOFAS_CDS_PAPER_2020.pdf>

The data was downloaded through the Python Climate data store API which

As later on a server will be used to load this data, an Amazon S3 bucket has been created where all the data has been uploaded. However,

* 1. *Data Preprocessing*

As we are using cloud technologies

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 previously seen flooding in Europe in *2016*.

*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.

1. UNISDR, C., 2015. The human cost of natural disasters: A global perspective. [↑](#footnote-ref-1)
2. Sella Nevo. [↑](#footnote-ref-2)
3. Needs reference [↑](#footnote-ref-3)