510 AN INITIATIVE OF THE NETHERLANDS RED CROSS

CONNECTING STATIONS AND DISTRICTS

in case of extreme discharge levels

Project overview, August 2020

EY VODW

TABLE OF CONTENTS

1. Introduction
2. Approach
3. GloFAS station selection toolProcess
4. Technical deepdive
5. Recommendations

1. INTRODUCTION

CONNECTING (FLOOD PRONE) AREAS TO STATIONS

- Right now, connections between (virtual) stations and areas are based on domain knowledge.
- It was never established, based on data, that the optimal stations were selected.
- A strong connection between (virtual GloFAS) stations and districts is essential because discharge measured at these stations act as an early warning sign of possible floods.
- We created an algorithm that determines the connections between virtual GloFAS stations, areas and districts, and visualised these in the station selection tool.
- This would enhance decision-making by pre-selecting better station-district matches, which in turn, could improve the GloFAS models' predictive performance for Uganda.

A data-driven, standardized, approach to determine the connection between stations, areas and districts

Create a data-driven, standardized, approach to determine connections between GloFAS stations, areas and districts.

Create an interactive tool with actionable insights for the 510 IbF team.

Create a starting point for optimising 'station' locations, determine the districts connected to stations, and specifying the exact flood-prone areas.

The team

Data Scientists

Jacopo Margutti

510, Red Cross

Lisanne van Brussel

EY VODW

Janno Bannink

EY VODW

Hydrologists

Bouke Ottow

510, Red Cross

Aklilu Teklesadik

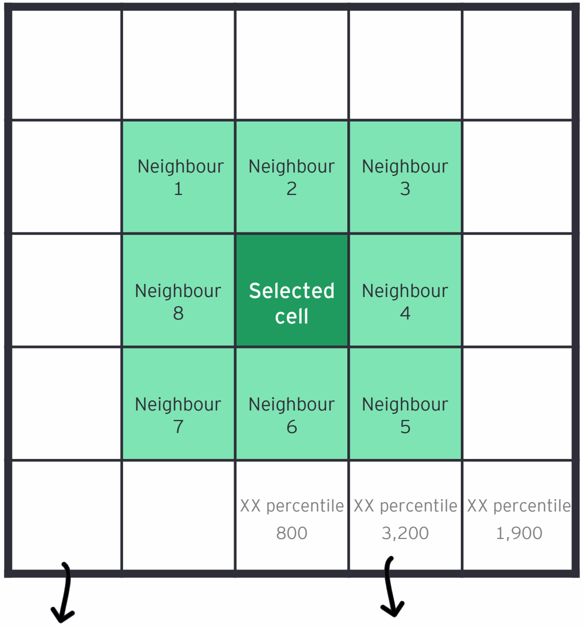
510, Red Cross

Phuoc Phùng

510, Red Cross

2. APPROACH

Identify relations between areas on days of extreme discharge levels



Each cell is a 0.1x0.1 degree grid cell

Each cell has it's own **XX percentile.**

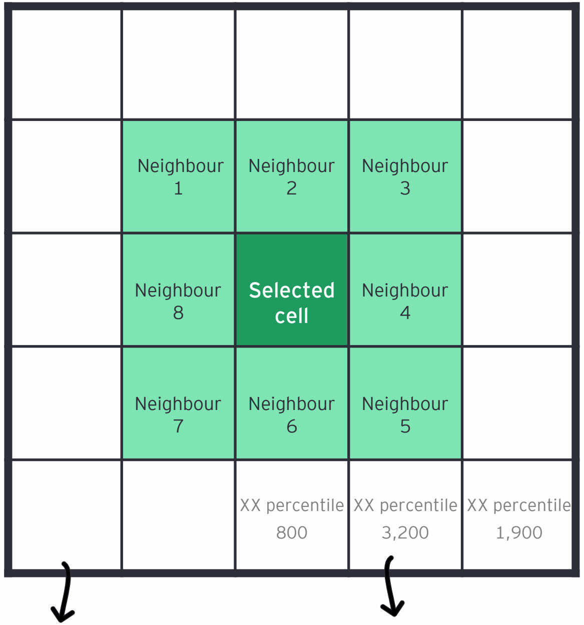
For **each grid cell/area**, we:

1. **derive the XX percentile of a X year return period**,
   Right know, we choose 95 percentile of a 10 year return period. However, this is can easily be adjusted to your preference.

2. **identify days with extreme discharges** for each day and each individual cell, based on the XX percentile of that cell.
   I.e. when the discharge for a specific cell is larger than the XX percentile, we identify extreme discharge that day. By doing so, we create a dummy dataframe where 1 = extreme discharge level:

Date	Cell X	Cell Y	Cell Z	Cell Q	Cell P
17-03-19	1	0	0	1	1
18-04-19	0	0	1	1	1
19-04-19	0	1	1	0	1
20-04-19	1	1	1	0	0

Find the relation between cells in case of extreme discharge



Neighbour 1

Neighbour 2

Neighbour 3

Neighbour 8

Selected cell

Neighbour 4

Neighbour 7

Neighbour 6

Neighbour 5

XX percentile 800

XX percentile 3,200

XX percentile 1,900

Each cell is a 0.1x0.1 degree grid cell

Each cell has it's own **XX percentile.**

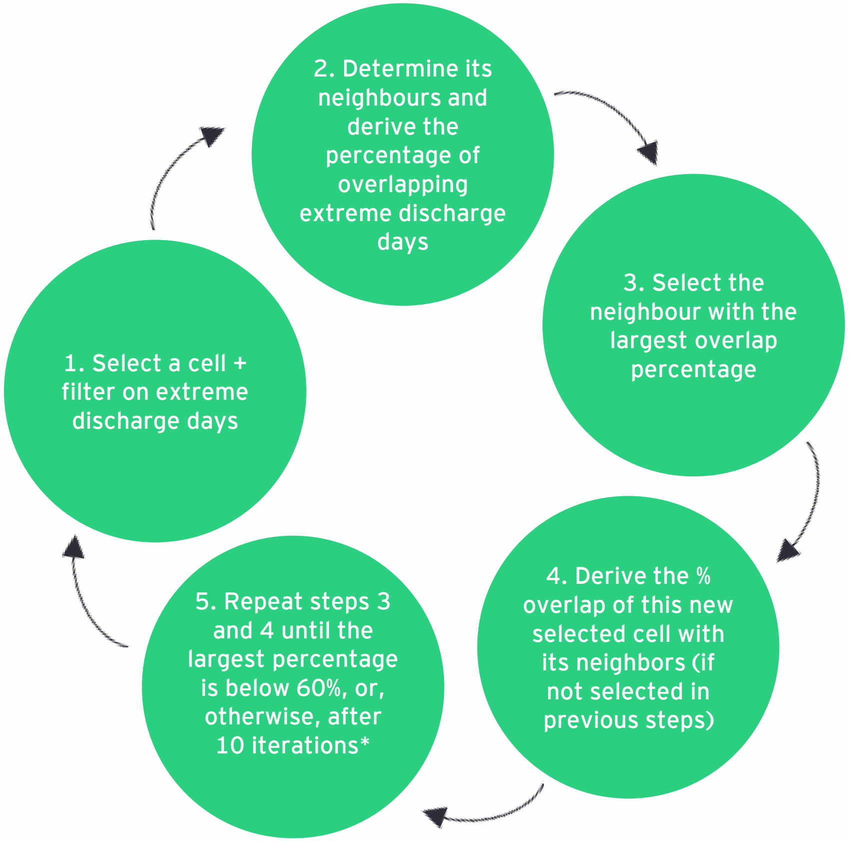
For each **selected cell**, we:

1. **select the neighbour cells.** For example, all the (directly) surrounding areas.
2. **filter on the days with extreme discharges** for the selected cell.
3. **determine the % overlap** of extreme discharge days with the neighbor cells on these filtered days.

Date	Selected cell	Neighbour 1	Neighbour 2	Neighbour 3	Neighbour 4
20-10-17	1	1	1	1	1
9-6-18	1	0	1	0	1
10-6-19	1	0	1	1	1
14-11-19	1	0	1	0	0
% overlap		25%	100%	50%	75%

An iterative process to find the connection between cells

- Executing the iterative process provides a data-driven approach to connect areas whenever extreme discharge levels occur.
- By doing so, one will see which areas often experience extreme discharge levels along with the selected cell.
- The path of connected areas is created after the process is repeated x times, or when the largest percentage is below a specified cutoff percentage. The number of iterations and the cutoff percentage can be customized. The default is 10 iterations and a cutoff percentage of 60%.



1. Select a cell + filter on extreme discharge days

2. Determine its neighbours and derive the percentage of overlapping extreme discharge days

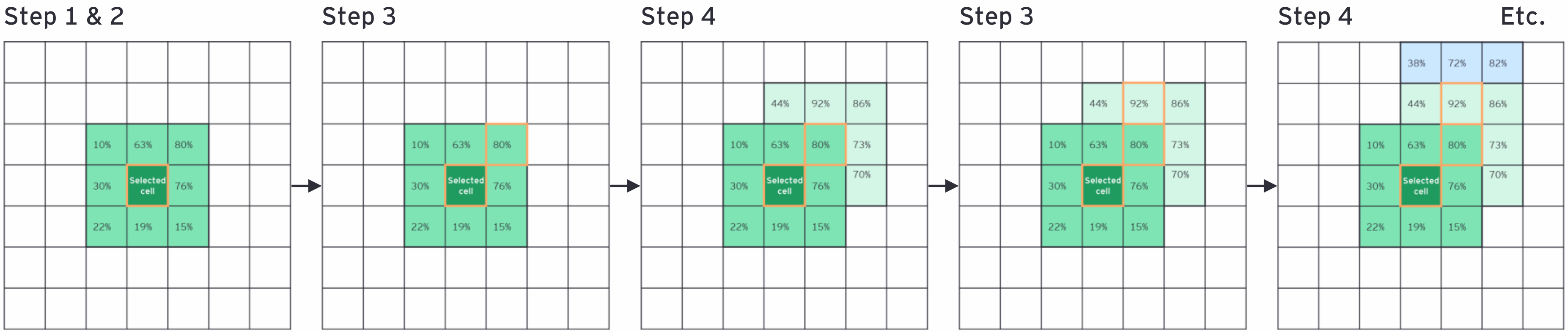
3. Select the neighbour with the largest overlap percentage

4. Derive the % overlap of this new selected cell with its neighbors (if not selected in previous steps)

5. Repeat steps 3 and 4 until the largest percentage is below 60%, or, otherwise, after 10 iterations*

* The selected cutoff percentage of 60% and 10 iterations are arbitrary and one can modify these at any point

Example of the iterative process



Step 1 & 2

Step 3

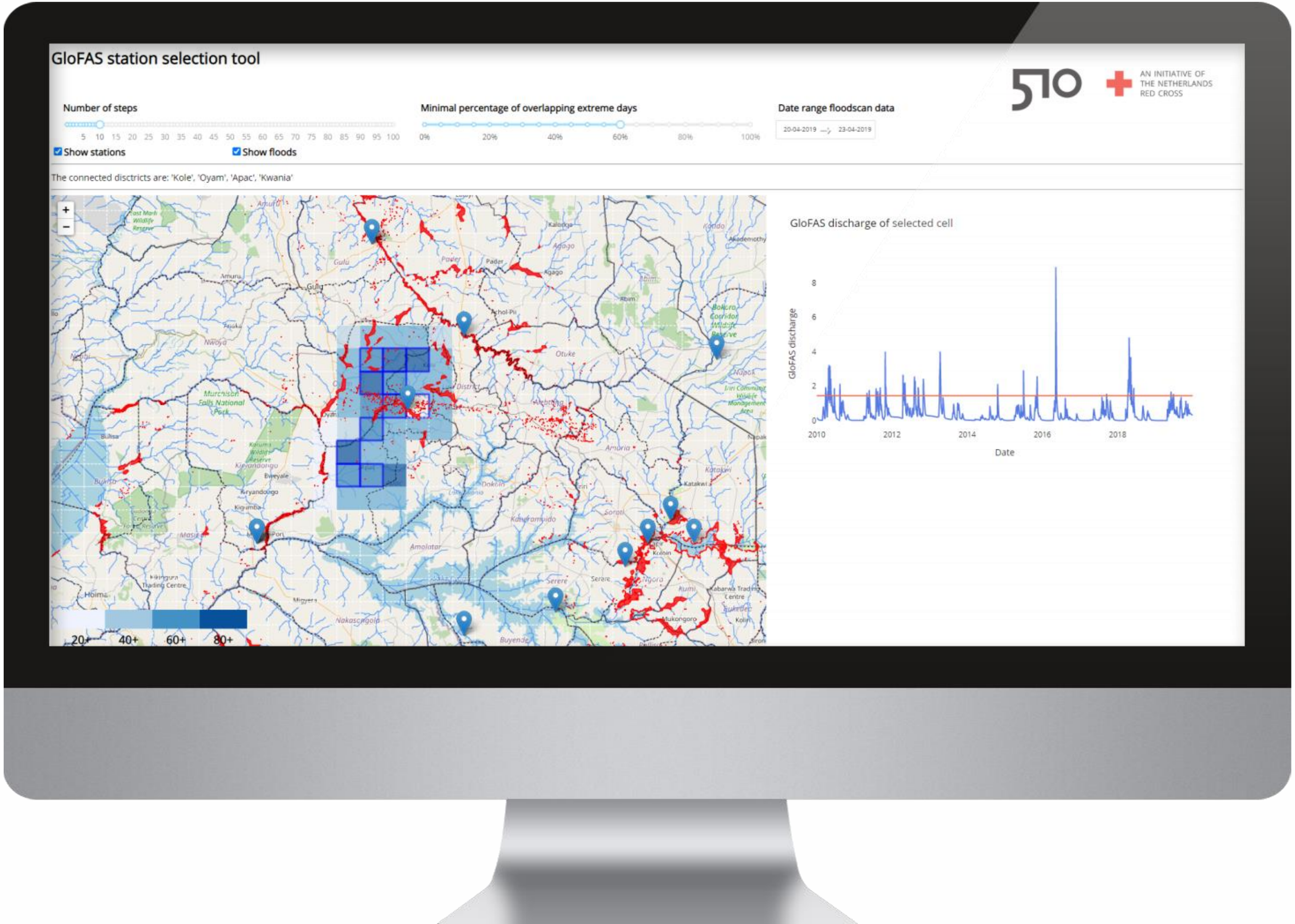
Step 4

Step 3

Step 4

Etc.

3. GloFAS station selection tool



GloFAS station selection tool

AN INITIATIVE OF THE NETHERLANDS RED CROSS

Number of steps

5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

Minimal percentage of overlapping extreme days

0% 20% 40% 60% 80% 100%

Date range floodscan data

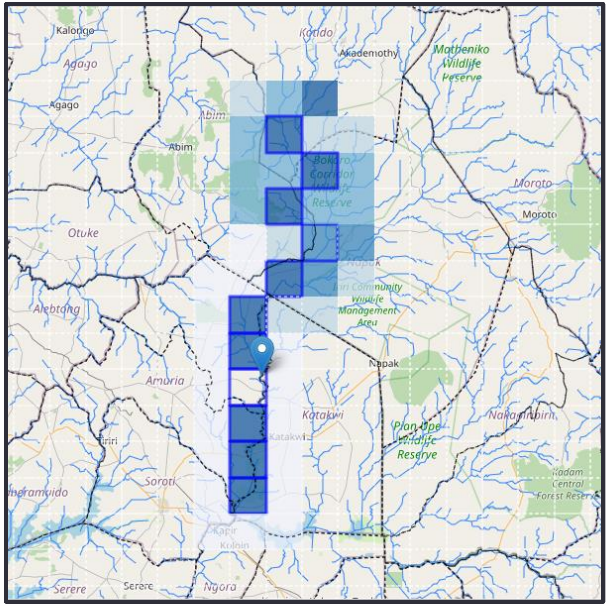
20-04-2019 → 23-04-2019

☑ Show stations

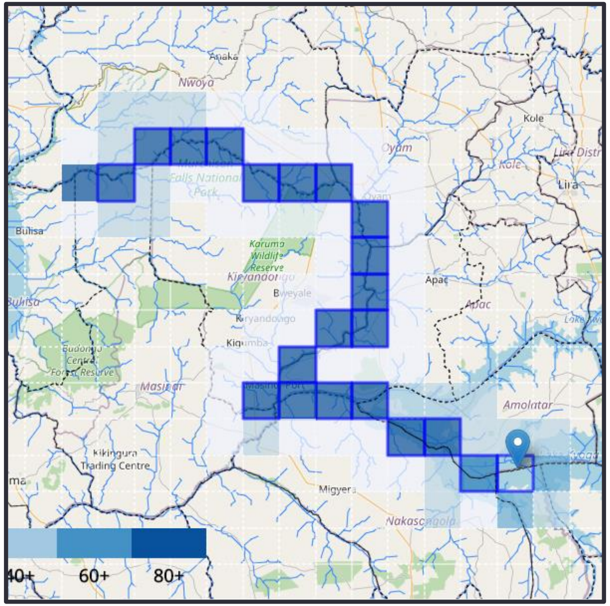
☑ Show floods

The connected discricts are: 'Kole', 'Oyam', 'Apac', 'Kwania'

Our approach results in a path that follows the river

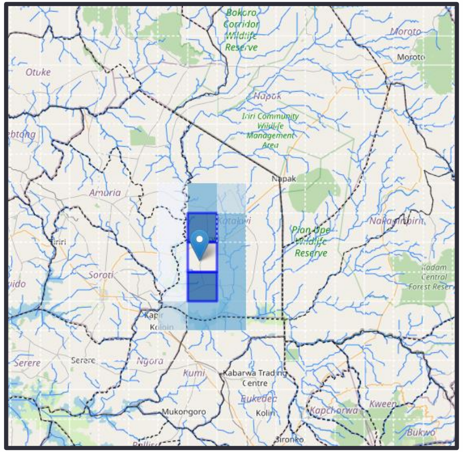

Kalongo
Agago
Kotido
Akademothy
Matheniko Wildlife Reserve
Agago
Abim
Abim
Bokora Corridor Wildlife Reserve
Moroto
Moroto
Otuke
Napak
Iri Community Wildlife Management Area
Alebtong
Napak
Amuria
Katakwi
Katakwi
Nakapiripirit
Pian Upe Wildlife Reserve
Tiriri
Soroti
Kadam Central Forest Reserve
Kapir
Kyere
Serere
Serere
Ngora

Settings: cell [2.05, 33.85] - 10 steps - 60%



Settings: cell [1.45, 32.75] - 20 steps - 70%

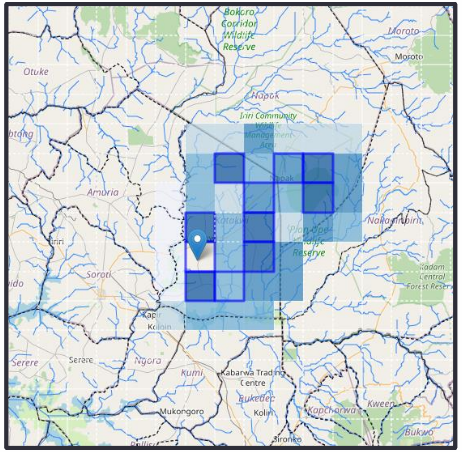
When no/small rivers occur within the start cell, the connected path is relatively small



Cell [1.80, 33.96] - 10 steps - 60%

from

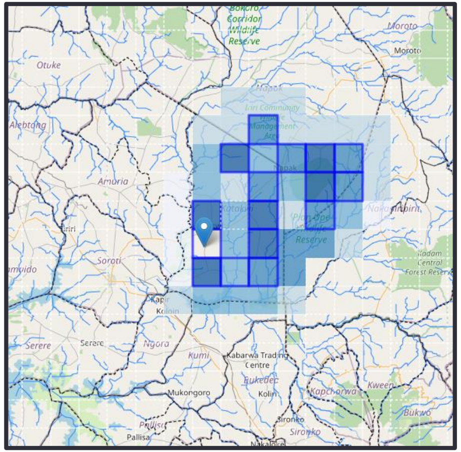
60% to 45%



*Cell [1.80, 33.96] - 10 steps - **45%***

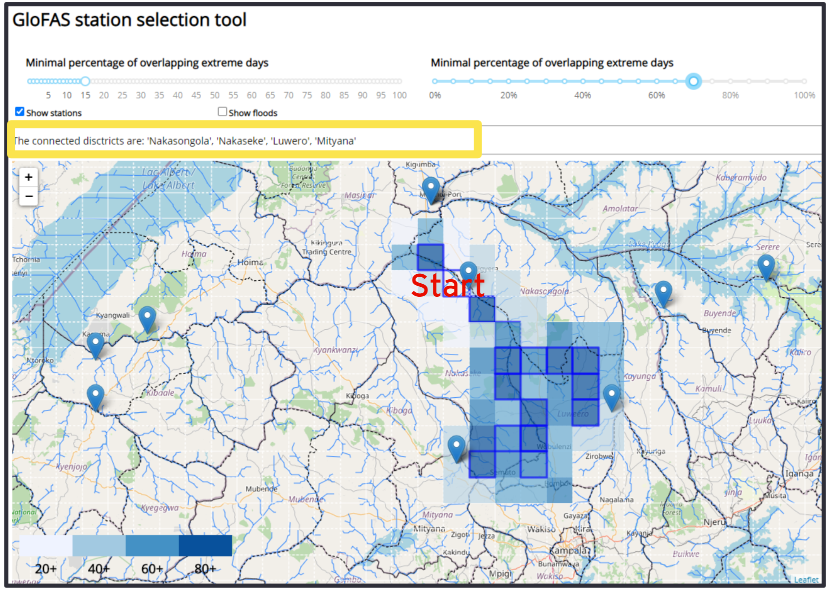
from

10 to 15 steps



*Cell [1.80, 33.96] - **15 steps** - 45%*

TOOL IDENTIFIES THE CONNECTED DISTRICTS



GloFAS station selection tool

Minimal percentage of overlapping extreme days

5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

Minimal percentage of overlapping extreme days

0% 20% 40% 60% 80% 100%

☑ Show stations

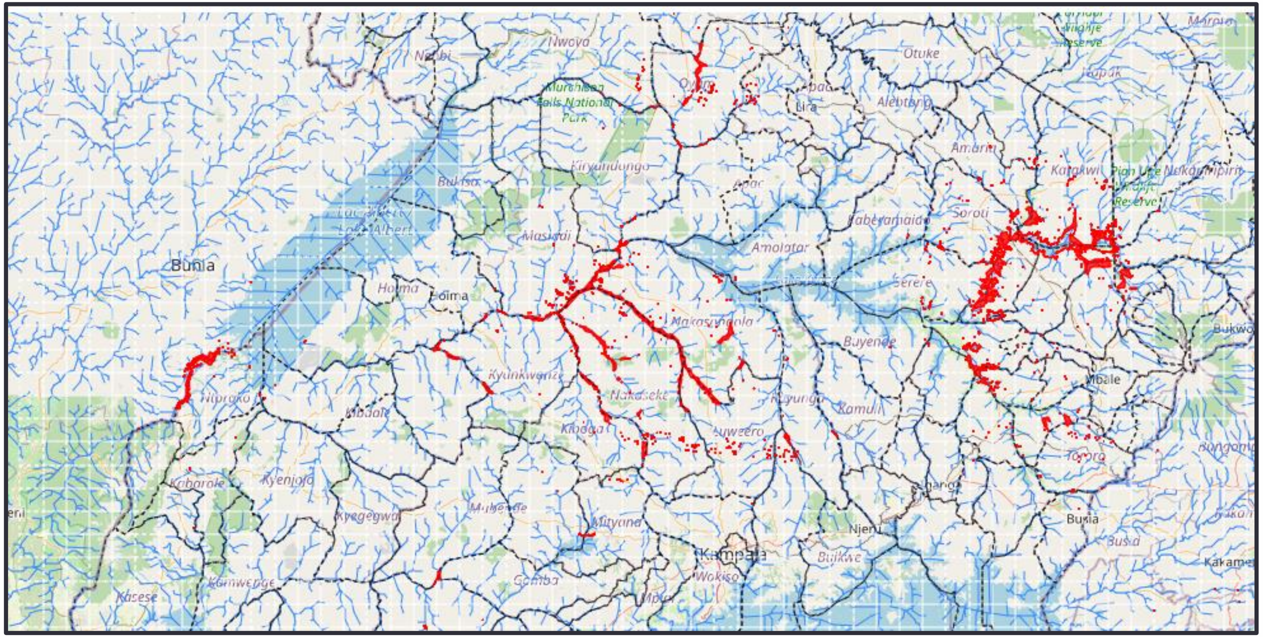
□ Show floods

The connected disctricts are: 'Nakasongola', 'Nakaseke', 'Luwero', 'Mityana'

Settings: F0032 [0.75, 33.75] - 15 steps - 70%

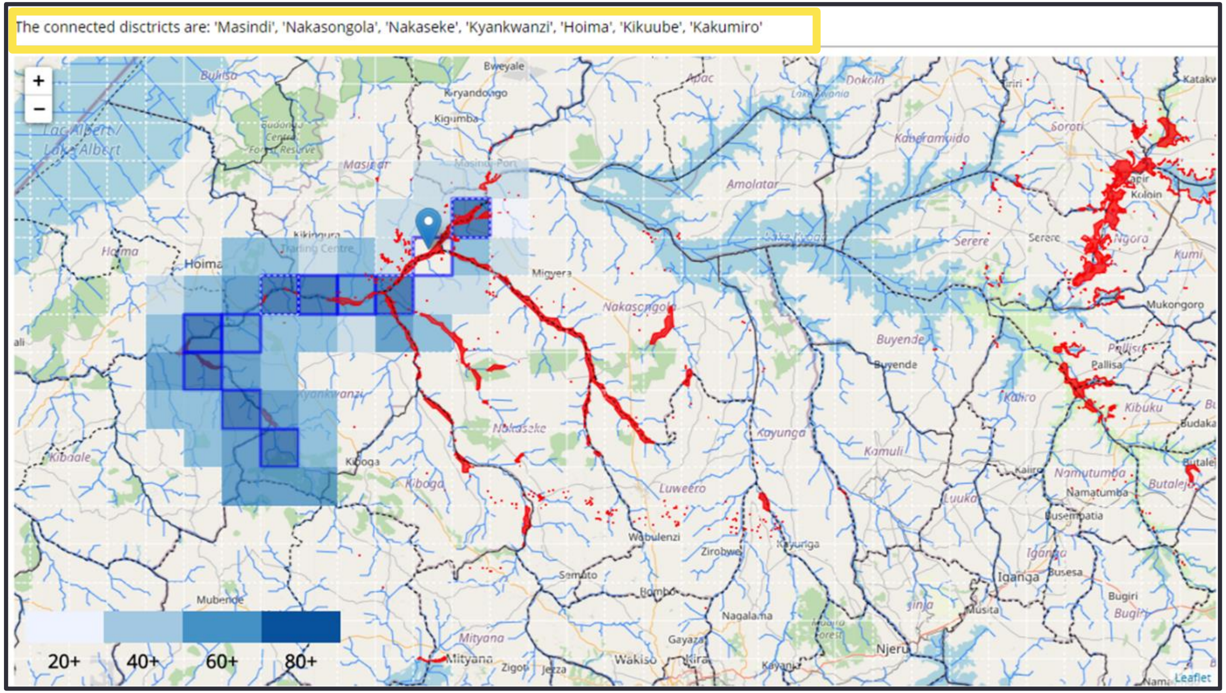
- The tool shows the districts that are connected with the selected cell/station.
- When predicting a flood at a certain station, it can be easily seen which districts might be affected.

We can identify the exact areas that have been flooded in the past (FloodScan data)



Settings: floods 1-8 November 2019

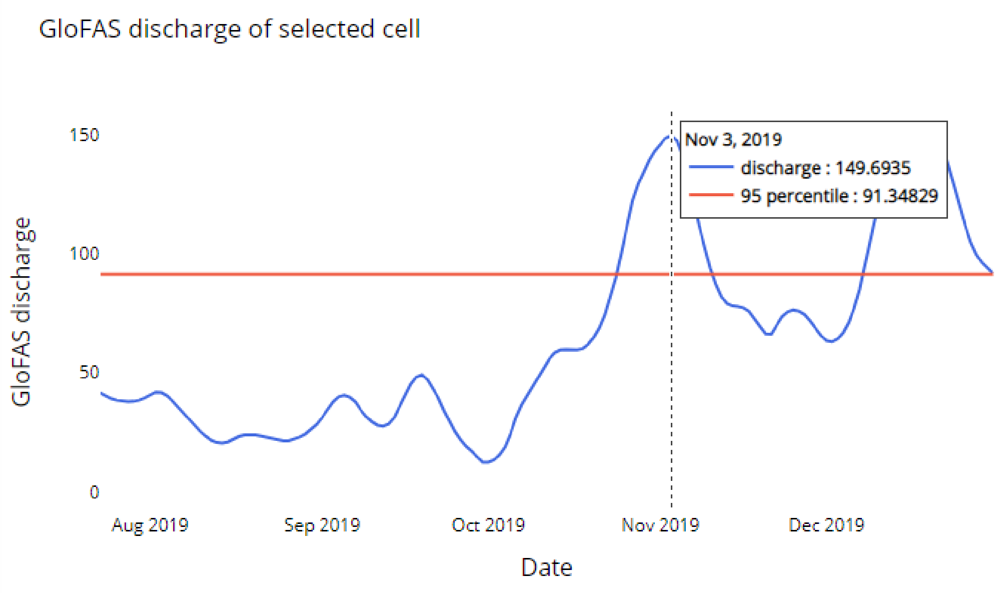
By selecting a flooded cell, we see the strongly connected areas and districts that might be affected



The connected disctricts are: 'Masindi', 'Nakasongola', 'Nakaseke', 'Kyankwanzi', 'Hoima', 'Kikuube', 'Kakumiro'

Settings: [1.45, 31.95] - 10 steps - 60% - floods 1-8 November 2019

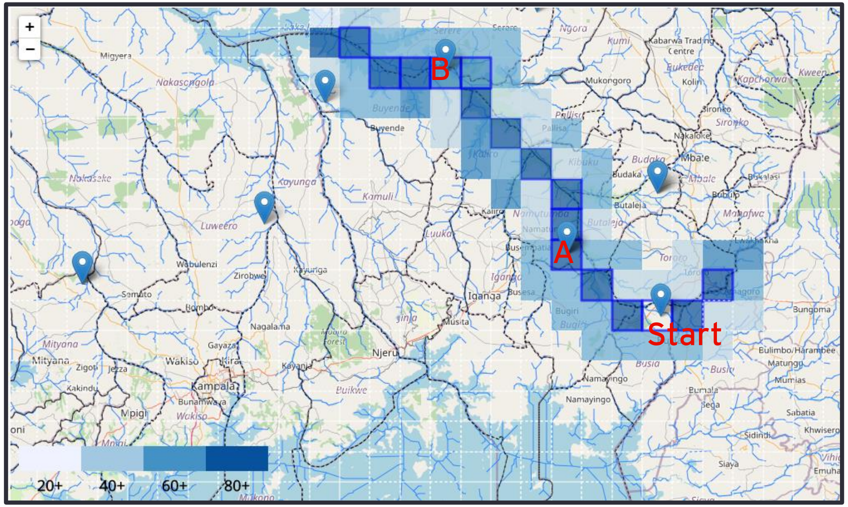
A link between GloFAS discharge data and FloodScan can be made to improve the trigger model



Settings: [1.45, 31.95] - floods 1-8 November 2019

Current locations of stations can be optimised [1/2]

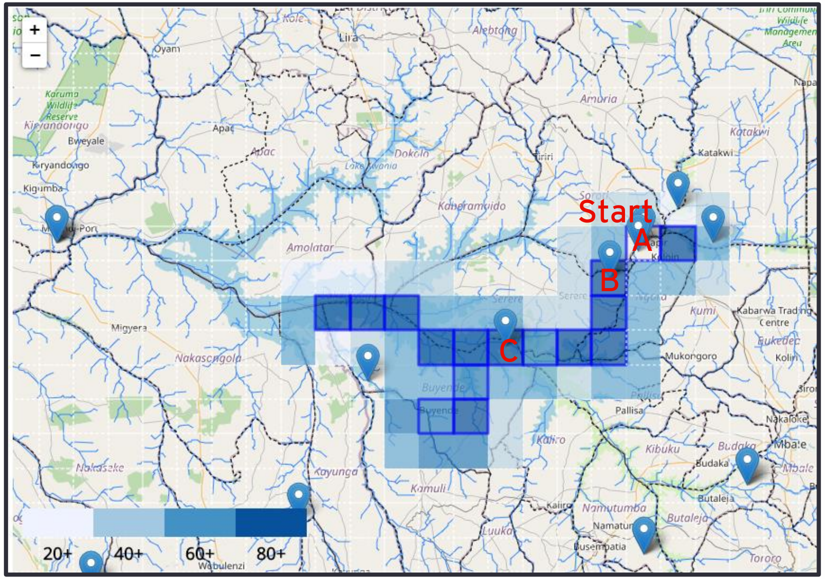
- Selecting in the cell (1.45,32.85), one can see that it is strongly connected with station A (F0032) and station B (UGGVS3).
- This might indicate that a flood at station A and B can already be detected in an earlier stage (i.e. at the start cell).
- Therefore, the start cell might be a more optimal location for a station, than station A and B.



Settings: cell [0.55, 34.05] - 15 steps - 75%

Current locations of stations can be optimised [2/2]

- When selecting station A (G5114), we clearly see that there is a strong connection with station B (F0034) and C (UGGVS3).
- This might indicate that stations B and C experience a flood when station A is flooded, and that floods at stations B and C can already be detected at station A.
- Therefore, the early warning signs of possible floods at station A could be used for possible floods at station B and C.



Settings: cell [1.65, 33.75] - 15 steps - 65%

4. Process

From raw data to an interactive map that visualises the connection between areas

(Down)load data

Download river discharge data via an API. We downloaded data from 2010 up and to 2019. Load shapefiles (e.g. grid, rivers, districts).

Prepare data

Select the right grid cells based on the coordinates of a country. Identify days with extreme discharge levels for each cell.

Analysis

Find the relation (% overlap) between cells in case of extreme discharge levels.

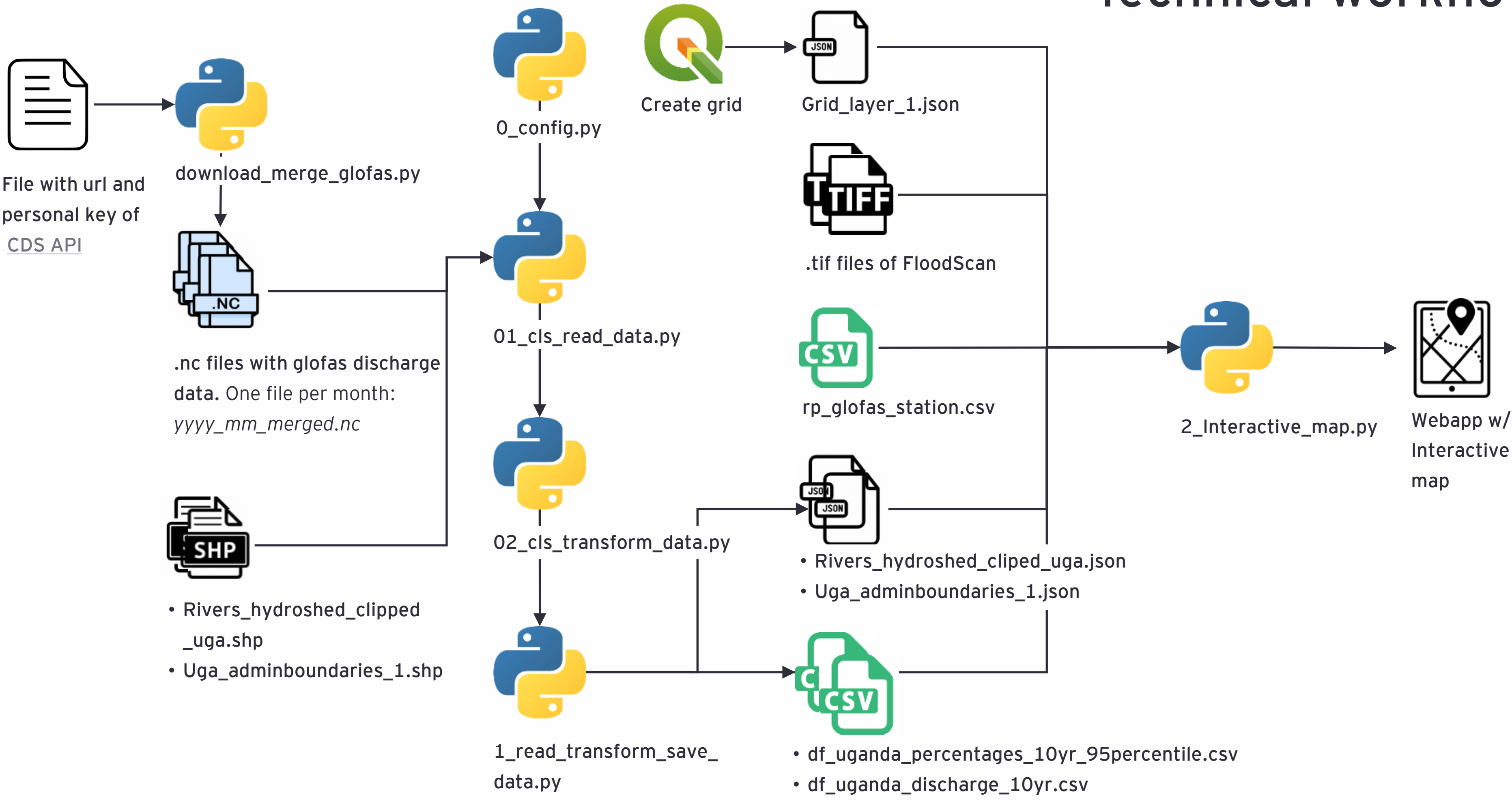
Interactive map

Create an interactive map that visualises the connection between areas.

Decision making

Optimize the locations of the stations. Decide which areas/districts might be impacted when a flood occurs.

Technical workflow



File with url and personal key of CDS API

download_merge_glofas.py

.nc files with glofas discharge data. One file per month: *yyyy_mm_merged.nc*

- Rivers_hydroshed_clipped _uga.shp
- Uga_adminboundaries_1.shp

0_config.py

01_cls_read_data.py

02_cls_transform_data.py

1_read_transform_save_ data.py

Create grid

Grid_layer_1.json

.tif files of FloodScan

rp_glofas_station.csv

- Rivers_hydroshed_cliped_uga.json
- Uga_adminboundaries_1.json

- df_uganda_percentages_10yr_95percentile.csv
- df_uganda_discharge_10yr.csv

2_Interactive_map.py

Webapp w/ Interactive map

5. TECHNICAL DEEPDIVE

Historical river discharge data from GloFAS is used

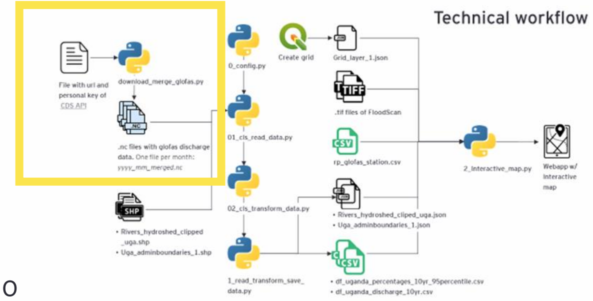
GloFAS data

- Global modelled daily data of river discharge (m^3/s) from the Global Flood Awareness System (GloFAS) is used.
- Daily data is available from 1979 on for each 0.1° x 0.1° grid cell (about 9km x 11km).
- The data and more background information is available at the Copernicus Emergency Management Service website.

Data we used

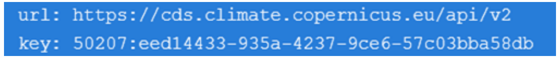
- Daily river discharge data from January 2010 up and to December 2019.
- Uganda, Ethiopia and Kenia are the countries that are in scope of this project.
- Version: GloFAS v2.1.

Get GloFAS data via an API call


Technical workflow
File with url and personal key of CDS API
download_merge_glofas.py
.nc files with glofas discharge data. One file per month: yyyy_mm_merged.nc
• Rivers_hydroshed_clipped _uga.shp
• Uga_adminboundaries_1.shp
0_config.py
01_cls_read_data.py
02_cls_transform_data.py
1_read_transform_save_data.py
Create grid
Grid_layer_1.json
.tif files of FloodScan
rp_glofas_station.csv
• Rivers_hydroshed_cliped_uga.json
• Uga_adminboundaries_1.json
• df_uganda_percentages_10yr_95percentile.csv
• df_uganda_discharge_10yr.csv
2_interactive_map.py
Webapp w/ interactive map

By following these steps, one is able to download the GloFAS data. These steps only have to be completed once. As a result, one is able download to worldwide discharge data for the years/months specified.

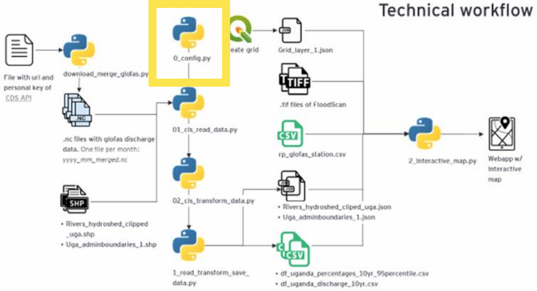
1. Create an Climate Data Store (CDS) or Atmosphere Data Store (ADS) account and log in to get your personal key. Copy the code displayed below in the file $HOME/.cdsapirc , be aware to use your own, personal key (see also here).

   ```
   url: https://cds.climate.copernicus.eu/api/v2
   key: 50207:eed14433-935a-4237-9ce6-57c03bba58db
   ```

2. Install the CDS API client, which is a (Python) package called *cdsapi*.
3. Open *download_merge_glofas.py* and:
   - change start and end year (or specific months) from which you wish to obtain the data;
   - if preferred, change your work directory to download the data directly to the correct location.
   - Run the script
4. Data will be downloaded and stored in a new folder named *data_all_startyear_endyear*.

All settings can be adapted in *0_config.py*



Technical workflow

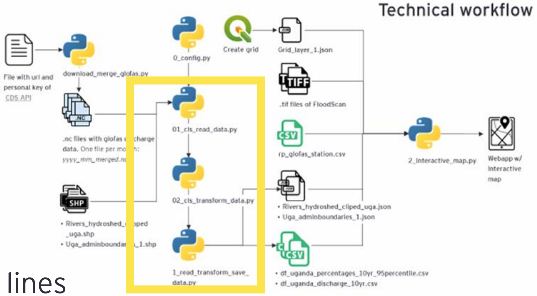
File with url and personal key of CDS API
download_merge_glofas.py
0_config.py
eate grid
Grid_layer_1.json
.tif files of FloodScan
.nc files with glofas discharge data. One file per month: yyyy_mm_merged.nc
01_cls_read_data.py
rp_glofas_station.csv
2_interactive_map.py
Webapp w/ interactive map
02_cls_transform_data.py
• Rivers_hydroshed_cliped_uga.json
• Uga_adminboundaries_1.json
• Rivers_hydroshed_clipped _uga.shp
• Uga_adminboundaries_1.shp
1_read_transform_save_ data.py
• df_uganda_percentages_10yr_95percentile.csv
• df_uganda_discharge_10yr.csv

The script *0_config.py* contains the configuration dictionary with all settings needed to read, prepare and transform the data. The (most important) settings that can be changed are:

- **Path** (i.e.. directory) to the folder with the downloaded GloFAS **data**: '*path_discharge_gridcells*'.
- **Boundaries** of the **country** interested in: '*lat_min*', '*lat_max*', '*lon_min*', '*lon_max*'.
- **Percentile** of the 10 year return period to identify the days with extreme discharge levels: '*percentile*'.
- **Rounds of neighbours** surrounding the selected cell (integer, 1 = 8 neighbours, 2 = 24 neighbours): '*neighbor_rounds*'.
- **Number of files**, can be changed if you don't want to read all the data (time consuming): '*nr_files*'.
- **Indicator** (Boolean) whether to **save** the created **data:** '*save_final_data*'.
- **Path** to the folder where, **newly created**, **data** should be **saved:** '*path_final_data*'.

Adapt these setting when you want to extend to other countries or when changing the criteria such as the percentile of the 10 year return period.

Transform data and do analysis


Technical workflow
File with url and personal key of CDS API
download_merge_glofas.py
.nc files with glofas discharge data. One file per month: yyyy_mm_merged.nc
• Rivers_hydroshed_clipped_uga.shp
• Uga_adminboundaries_1.shp
0_config.py
Create grid
Grid_layer_1.json
01_cls_read_data.py
02_cls_transform_data.py
1_read_transform_save_data.py
.tif files of FloodScan
rp_glofas_station.csv
• Rivers_hydroshed_cliped_uga.json
• Uga_adminboundaries_1.json
• df_uganda_percentages_10yr_95percentile.csv
• df_uganda_discharge_10yr.csv
2_interactive_map.py
Webapp w/ interactive map

Once the settings in *0_config.py* are modified, the classes *01_cls_read_data.py* and *02_cls_transform_data.py* can be used. However, make sure that the path specified in the first lines within both scrips is right. This should be the path of the scripts.

01_cls_read_data.py

- Imports *0_config.py*
- Reads in all .nc files with discharge data and shapefiles.
- Creates one pandas dataframe with all discharge levels of the selected timeframe and coordinates. Here, rows are the dates and the columns are the coordinates (lat_lon) of each grid cell.

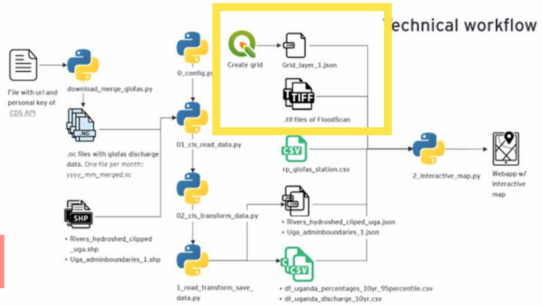
02_cls_transform_save_data.py

- Imports *01_cls_read_data.py*
- Creates a pandas dataframe with dummies to identify extreme discharges per day/grid cell.
- Creates a pandas dataframe with the percentage of overlap of extreme discharge days of a cell and its adjacent (neighboring) cells.
- Saves the three dataframes as csv-files.

1_read_transform_save_data.py

- Imports *02_cls_transform_save_data.py*
- Makes sure that the classes are executed.
- *Running this script will actually load, transform and save the data.*

Replicating GloFAS grid & FloodScan data


echnical workflow
File with url and personal key of CDS API
download_merge_glofas.py
.nc files with glofas discharge data. One file per month: yyyy_mm_merged.nc
• Rivers_hydroshed_clipped _uga.shp
• Uga_adminboundaries_1.shp
0_config.py
01_cis_read_data.py
02_cis_transform_data.py
1_read_transform_save_ data.py
Create grid
Grid_layer_1.json
.tif files of FloodScan
rp_glofas_station.csv
• Rivers_hydroshed_cliped_uga.json
• Uga_adminboundaries_1.json
• df_uganda_percentages_10yr_95percentile.csv
• df_uganda_discharge_10yr.csv
2_interactive_map.py
Webapp w/ interactive map

Create grid using Qgis

GloFas provides a gridded dataset at a 0.1° resolution. For the this tool this grid was replicated using Qgis.

For each grid cell the coordinates of the middle (of the cell) are defined. The boundaries of each extend in all four directions .05° from that middle point.

A research tool from Qgis was used to create the grid. By selecting rectangle (polygon), the bounding box of the desired grid size (in increments of .1°), vertical and horitzontal spacing of .1° one is able to replicate the GloFAS grid.

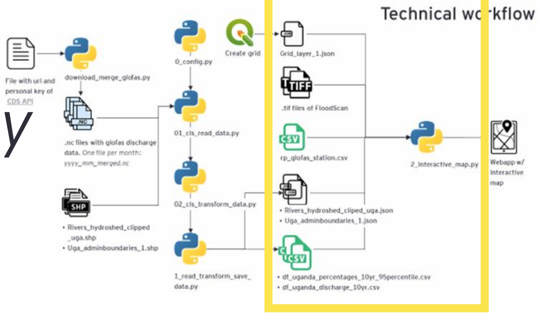
FloodScan geoTIFF files

FloodScan data contains a daily representation of temporarily flooded areas in a region.

The FloodScan data which was used contains data on a ~90m resolution. This relatively small scale representation of the data might not accurately represent changing conditions on the ground.

FloodScan supplies their data using the geoTIFF format. These (Tagged Image Files) often contain one or multiple layers of information, as well as contains geospatial metadata.

Input files of the script 2_*interactive_map.py*

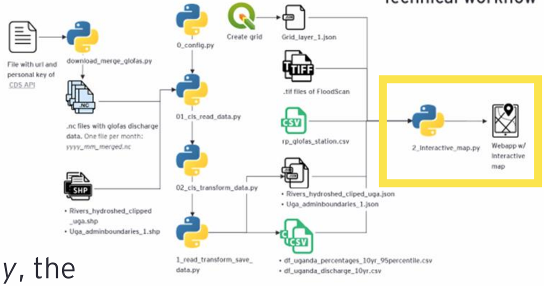

Technical workflow
File with url and personal key of CDS API
download_merge_glofas.py
.nc files with glofas discharge data. One file per month: yyyy_mm_merged.nc
• Rivers_hydroshed_clipped_uga.shp
• Uga_adminboundaries_1.shp
0_config.py
01_cis_read_data.py
02_cis_transform_data.py
1_read_transform_save_data.py
Create grid
Grid_layer_1.json
.tif files of FloodScan
rp_glofas_station.csv
• Rivers_hydroshed_cliped_uga.json
• Uga_adminboundaries_1.json
• df_uganda_percentages_10yr_95percentile.csv
• df_uganda_discharge_10yr.csv
2_interactive_map.py
Webapp w/ interactive map

The interactive_map.py uses several files as input, which will be discussed below.
The order is similar as the order of the files in the technical workflow.

1. **Grid_layer_1.json:** .1°x .1° degree grid cells of Uganda as explained in previous slide.
2. **FloodScan geoTIFF files:** FloodScan supplies their data using the geoTIFF format. These (Tagged Image Files) often contain one or multiple layers of information, as well as contains geospatial metadata.
3. **rp_glofas_station.csv**: csv-file with all 'virtual' GloFAS stations names and locations.
4. **Rivers and admin boundaries jsons:** the to json transformed shapefiles with the river network and admin boundaries (districts) of Uganda. These will be saved when running *1_read_transform_save_data.py*.
5. **GloFAS discharge data and percentage overlap data:** File with the GloFAS discharge data of 2010-2019 and all grid cells of Uganda (*df_uganda_discharge_10yr.csv*). File with the percentage of overlap of extreme discharge days of a cell and its adjacent (neighbouring) cells (*df_uganda_percentage_10yr_95percentile.csv*). These files will be saved when running *1_read_transform_save_data.py*.

The webapp can be activated by running the script *2_interactive_map.py*

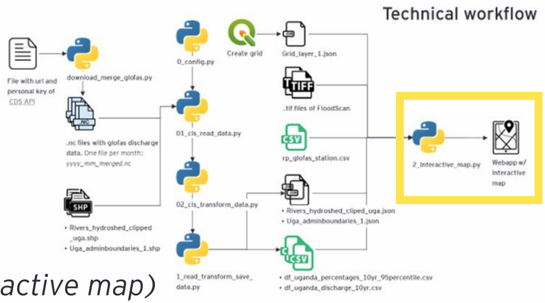
Technical workflow


File with url and personal key of CDS API
download_merge_glofas.py
.nc files with glofas discharge data. One file per month: yyyy_mm_merged.nc
• Rivers_hydroshed_clipped _uga.shp
• Uga_adminboundaries_1.shp
0_config.py
01_cls_read_data.py
02_cls_transform_data.py
1_read_transform_save_ data.py
Create grid
Grid_layer_1.json
.tif files of FloodScan
rp_glofas_station.csv
• Rivers_hydroshed_cliped_uga.json
• Uga_adminboundaries_1.json
• df_uganda_percentages_10yr_95percentile.csv
• df_uganda_discharge_10yr.csv
2_interactive_map.py
Webapp w/ interactive map

Once the files that are created and saved by running the script *1_read_transform_save_data.py*, the interactive map can be created. The interactive map is based on using several functions that will be explained below.

- The iterative process to find the connection between areas is defined by the function '***select_neighbours_percentages***'. The input arguments are: a pandas dataframe with the percentages, a location (lat_lon), the number of steps and the minimal percentage of extreme discharge days that should overlap between areas. Selecting a cell in the map will activate this function.
- The function '***get_districts***' will be activated when clicking on a cell in the map. This will return a list with all the districts that are connected to the path that is created when selecting a cell.
- Another function that is activated when selecting a cell is '***show_discharge_graph'.*** This will provide a graph with the GloFAS discharge data of the past 10 years and a line of the 10 year return period 95% percentile value of the selected cell.
- Areas that flooded in the past (provided by FloodScan) can be displayed in the map. The function that makes sure that the files of the selected days will be merged to one numpy array is defined as '***get_flood_polygons***'.

The GloFAS station selection tool


Technical workflow
File with url and personal key of CDS API
download_merge_glofas.py
.nc files with glofas discharge data. One file per month: yyyy_mm_merged.nc
• Rivers_hydroshed_clipped _uga.shp
• Uga_adminboundaries_1.shp
0_config.py
01_cls_read_data.py
02_cls_transform_data.py
1_read_transform_save_data.py
Create grid
Grid_layer_1.json
.tif files of FloodScan
rp_glofas_station.csv
• Rivers_hydroshed_cliped_uga.json
• Uga_adminboundaries_1.json
• df_uganda_percentages_10yr_95percentile.csv
• df_uganda_discharge_10yr.csv
2_interactive_map.py
Webapp w/ interactive map

When running the script *2_interactive_map.py*, *the GloFAS station selection tool* (*e.g. the interactive map*) will be visible in a webapp. The tool is interactive and has some filters.

- By clicking on the map, the function '*select_neighbours_percentages*' will be activated and the **path of connecting cells** will be shown. The shade of blue indicates the percentage of overlap of extreme discharge days (dark blue = large overlap).
- The **number** of **steps** of the iterative process can be changed.
- The **minimal percentage** of overlapping extreme days of a cell and its adjacent cells can be changed.
- The **districts** connected to selected cell will be displayed in the webapp.
- On the right side of the tool, a graph of the **GloFAS discharge** of the selected cell over time is shown. When interested in a more specific date range, one can zoom in. The orange/red line is the 10 year return period 95% percentile value of the selected cell.
- By clicking on the 'show stations' checklist button, the current **stations** will be shown.
- FloodScan data will be shown when selecting the 'show floods' checklist. The **flooded areas** within the selected date range can be recognized by its red color. This date range can be change to your preference. FloodScan data is available for 1998-2019.

Steps to follow when extendending to new countries or changings settings

1

Download/load all data needed

Download GloFAS data via the API (only needed once). For new countries, get the shapefiles / jsons of the grids, districts and rivers. Besides, geoTIFF files of FloodScan should be collected.

2

Change settings in the configuration (*'0_config.py'*)

For example, change boundaries when adding a new country, or change settings such as the percentile of the 10 yr return period.

3

Make sure that the paths in the scrips are right and run *'1_read_transform_save_data.py'*

This will save the dataframe with the connection percentages, which is the input for the interactive map.

4

Merge the datafiles with the percentages of the different countries/percentiles

5

Update the path to the input files (e.g. the merged dataframes) to update the tool

The necessary packages

For downloading GloFAS data
download_merge_glofas.py

1. cdsapi
2. zipfile
3. shutil
4. xarray
5. os
6. geopandas

Creating csv file with % overlap of extreme discharge days between a cell and his neighbour cells

1. pandas
2. numpy
3. Dataset
4. datetime
5. os

For creating the interactive map

1. pandas
2. numpy
3. os
4. dash
5. dash_leaflet
6. dash.dependencies
7. dash_html_components
8. dash_core_components
9. shapely.geometry
10. json
11. datetime
12. re
13. plotly.express
14. plotly.graph_objects
15. rasterio
16. rasterio.features

6. RECOMMENDATIONS

Next steps that add even more value

- Extend the interactive map with:
    - new countries;
    - other percentiles to identify cell specific discharge levels.
- Automatically optimises the locations of ‘stations’.
- Change the number days with discharges that should be compared. Right know, only single days are compared.
- Extend data with more historical data to make results even more stable.