Trigger methodology for floods in Zambia

Assessing the forecast skill of GLOFAS and validating historical impact data for their use in the trigger methodology of the EAP for river floods in Zambia

# Flood Risk in Zambia

Flood risks in Zambia consist of:

* Fluvial/river flood: these are floods cause by large discharge of water within river basins. These floods impact large areas and happen on a timescale odmostly large space and time scale riverine floods, characterized by usually high direct damage and displacement of many people due to their wide flood risk extent, and generally densely populated flood plain areas.
* Flash flood, that occur at smaller time and spatial scales. A flood that rises and falls quite rapidly with little or no advance warning, usually the result of intense rainfall over a relatively small area (Glossary of the American Meteorological Society, 2000 edition). Key aspect of the definition is the time scale: sudden hydrological response to the causative event.
* Pluvial flooding: these are the floods caused by heavy rains in areas where there is no water course and no mountain area. Floods happening in Lusaka are of this type. These floods are caused by lack of proper drainage in the impacted area (e.g. sewage or vegetation) (see http://ec.europa.eu/environment/water/flood\_risk/index.htm)

One line on why riverine flood is the most impactful in humanitarian terms in Zambia

## Hydrological profile of Zambia

Zambia has extensive surface water resources, with a number of large perennial rivers. The major rivers in Zambia are the Zambesi and the Kafue river. Flood plains are the areas adjacent to rivers that are most prone to seasonal flooding and where the most extreme riverine flood events are expected to happen. Zambia has several flood plains as identified by the light blue areas in the ecoregions map in Figure 1.

The regions of interest are:

**24 —**[Barotse floodplain](https://en.wikipedia.org/wiki/Barotse_floodplain" \o "Barotse floodplain), [Luanginga River](https://en.wikipedia.org/wiki/Luanginga_River" \o "Luanginga River) floodplain and [Luena Flats](https://en.wikipedia.org/wiki/Luena_River,_Western_Zambia" \o "Luena River, Western Zambia), in Western Province

**25 —**[Bangweulu Swamps](https://en.wikipedia.org/wiki/Bangweulu_Swamps) and floodplain, Northern and Luapula Provinces

**26 —**Kafue Flats, Central and Southern Provinces

**27 —**[Lukanga Swamp](https://en.wikipedia.org/wiki/Lukanga_Swamp" \o "Lukanga Swamp) and floodplains of the Kafue and its tributaries in Central Province and south west Copperbelt Province

**28 —**Lake Mweru Wantipa/Mweru Marsh floodplain, Northern Province

**29 —**Busanga Swamps and plain, Kafue National Park, North Western Province

**30 —**Upper [Chambeshi River](https://en.wikipedia.org/wiki/Chambeshi_River" \o "Chambeshi River) floodplain, Northern Province

**31 —**Luapula Swamps south of [Lake Mweru](https://en.wikipedia.org/wiki/Lake_Mweru), Luapula Province

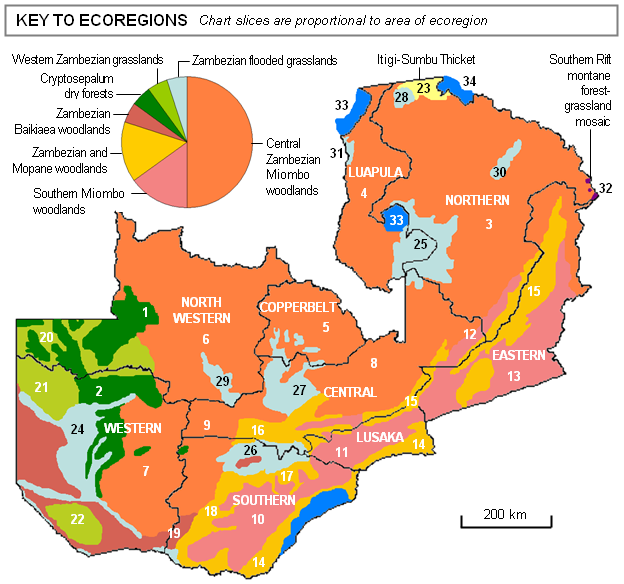


Figure 1 By User:Rexparry\_sydney - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=3077995

# Historical flood events

We have records of floods events from XX reports by DMMU over a time span between 2003 and 2014.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Description** | | | **Data Quality** | | | | **Costs of data** | | |
| **Data set** | **Source of data** | **Date of source/disasters covered** | | **Source reliability** | **Content accuracy** | **Granularity** | | **Type** | **Structure level** |
| Flood report | IFRC relief workers | Historical floods | | 3 | 4 | District | | Report | unstructured |
| Disaster Database | EM-DAT | 1967-2017 | | 4 | 4 | National | | Database | structured |
| Internally displaced persons | OCHA | 2010-2016 | | 4 | 3 | District | | Dataset | structured |
| Damage and loss assessment | Government | 2015 | | 3 | 4 | District | | Report | Semi-structured |

We apply a text mining algorithm to export the locations (districts and wards) where the reported floods took place (Figure 2 shows the number of reports analyzed for each year). The result is a table with a flag if the corresponding location has been reported flooded in that year.

[INSERT TABLE]

Of course these data include any reported flood event in these years. However not all these floods are riverine floods. For example floods in the Lusaka and Eastern province are likely not of this sort because of the lack of large water bodies. Since our flood forecasting system limits its scope to riverine floods, we clean the historical data of the events reported in these provinces.

In the reported periods, there is a total of 375 wards affected by floods in Zambia, of which 292 in provinces where major water bodies are present. Figure 2 shows the occurrence of flood over the reporting period. The most affected provinces in the reporting period (i.e. the ones where most wards have been reported as flooded) are, in this order: Southern, Western, Central and North Western. This corresponds with the areas where the major rivers are located, e.g. the Zambesi Basin (Northwestern, Southern and Western) and the Kafue river basin (Southern and Central).

Figure 2 histogram showing the number of flood reports in Zambia for each year that have been analyzed by the text mining algorithm

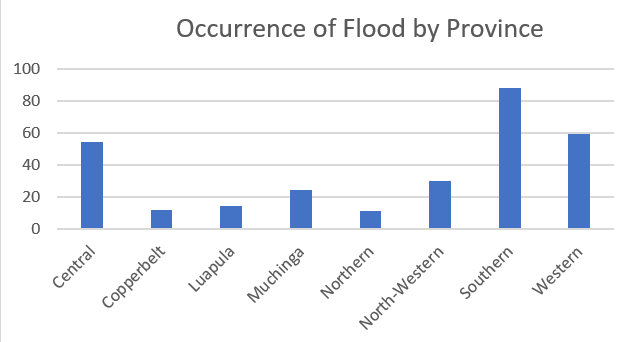


Figure 3 Occurrence of floods by province as from reports by DMMU. The bar represents the number of wards in each province that have reported flood in the reporting period (not all these events were DREF activations)

These areas of major concern are also validated by the historical archive of IFRC DREF activations over the reporting period. Table 1 shows a list of IFRC DREF activations in the reporting period 2003-2014. With the exception of 2010 urban flooding in Lusaka (not caused by river discharge), these events are all linked to riverine floods in one of the main flood plains. DREF activation can also be used as an indicators for selecting the type of events that we want to consider in the reporting period (the rationale is : we would like to be able to forecast those events). Indeed the threshold for DREF activation is XXX.

Table 1 DREF ACTIVATIONS IN THE REPORTED PERIOD (NOTE 2010 is not a riverine flood but likely pluvial)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **month** | **Year** | **Province** | **Affected provinces/districts** | **IFRC appeal** |
| January-march | 2004 | Western | Sesheke, Senanga, Kalabo Lukulu, | M04ME008 |
| Dec-March | 2006 | Southern | Kazungula | MDRZM002 |
| Dec-March | 2007 | North Western | Solwezi, Mpulungu | MDRZM004 |
| January | 2008 | Southern | Mazabuka | MDR63001 |
| March | 2009 | Western, North Western | Western , Northwestern | MDRZM006 |
| March | 2010 | Lusaka | Lusaka | MDRZM007 |
| February | 2013 | Central | Mumbwa | MDRZM008 |
|  |  |  |  |  |

# Characterization of extreme events

The flood forecasting system we are designing is thought to support Forecast Based Financing by Disaster Relief Emergency Funds (DREF). These funds are only released by eligible events that are selected via a set of impact criteria such as CRITERIA FOR DREF.

We analyze the historical events by extracting the associated flood extent area for each district using Floodcast. Then we perform a statistical analysis to independently find the extreme events in the reported sample and see if the DREF activations were falling in this category.

For each flood event we calculate its t-score[[1]](#footnote-1), which quantifies how far (in terms of number of standard deviations) is the given event is from the average event. We then highlight the reported events that fall outside the 99% percentile according to the one tiled student distribution. This is a standard method to spot extreme events in a sample.

[TABLE]

We find that the DREF activations all fall within extreme flood extents with respect to the average event at least for some of the reported districts therefore they can be used as an indication of the type of disasters we would like to forecast with the system.

# Prioritization of impacts & impact level forecasting

In collaboration with ZRCS and DMMU we selected impact indicators that are relevant for the local context and we would like to forecast for. Once that the FbF system triggers for flood it is useful to give an indication of the following parameters:

|  |
| --- |
| Exposed Shelter |
| Exposed Crops type |
| Exposed Livestock type |

We note that values of these indicators are sometimes mentioned in the historical data: the number of displaced is linked to the shelter exposed to the hazard. Often loss of lives is also mentioned.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **month** | **Year** | **Province** | **Affected provinces/districts** | **Displaced** | **Loss of life** | **loss of livestock** | **loss of Crops** |
| March | 2003 | southern | Gwembe | 1,679 | NA | NA | 90% \* |
| January-march | 2004 | Western | Sesheke, Senanga, Kalabo Lukulu, | 20000 | 2 | 90-100% | 90-100% |
| Dec-March | 2006 | southern | Kazungula | 2500 | 0 | NA | NA |
| Dec-March | 2007 | North Western | SOLWEZI AND MPULUNGU | 250000 | 11 | 0-25% | 25-75% |
| January | 2008 | southern | Mazabuka | 800 | 21 | 28% | 20-60% |
| March | 2009 | Western & NW | Western & NW | 1000 | 31 | NA | 80% |
| March | 2010 | Central | Serenje | 788 | NA | NA | NA |
| March | 2010 | Lusaka | Lusaka & NW | 1200 | 8 | 60% | 60% |
| January | 2013 | southern | Mazabuka | 336 | NA | NA | NA |
| February | 2013 | central | Mumbwa | 1800 | 2 | NA | NA |
| February | 2014 | southern | Southern | 20000 | NA | NA | NA |

A method to forecast for this parameters is to overlay a forecasted impact area to a raster of population, crops extent and livestock distribution and then calculate the numbers or percentages of population, crops and livestock in the overlapping area.

We obtain raster data for these indicators from the following sources:

|  |  |  |  |
| --- | --- | --- | --- |
| Exposed Shelter | Ward | Facebook HRSL | 30x30m |
| Exposed Crops type | Ward | IISA-IFPRI | 8x8km |
| Exposed Livestock type | Ward | ILRI - FAO | 8x8km |

These data can be used to show impact on forecasted events, but also to study the impact of past events in the historical data. To link the historical events to their impact we extract the flood extent from the GLOFAS hidcast simulations and calculate back the exposed population, crops and livestock.

TBD

# GLOFAS forecasting system

In order to determine the danger levels for a flood in Zambia we use historical simulated data from GLOFAS. The Global Flood Awareness System (GloFAS), jointly developed by the European Commission and the European Centre for Medium-Range Weather Forecasts, is a global hydrological forecast and monitoring system that delivers global ensemble discharge forecast for major rivers.

We acquired historical simulations from GLOFAS for 24 stations in Zambia. The simulations span over a time range of 22 years (from 1997 to 2018) and represent the hindcasted discharge at a given point in time for the next 30 days calculated with 10 ensemble members. Figure 1 shows the position of the GLOFAS data extraction points in Zambia.

Figure 4 Map showing the position of the monitoring station in Zambia represented within the GLOFAS model. The borders represent the administrative borders of the districts (source OCHA). [SOME OF THE NAMES NOT SHOWING: CHECK ALSO NOT ALL THE STATIONS ARE SHOWN]



### Comparison with observations

We use observed discharge data from the WARMA monitoring station to validate GLOFAS hindcast data. Out of the 24 GLOFAS stations we have overlapping data for 9 stations. The monitoring data is unfortunately not continuous in time however we compare overlapping time ranges.

We find that the quality of the comparison of hindcast data and the observed discharge varies for each station (see Figure 3 ). For some station even if the general pattern of higher discharge peaks is the same in observation and simulations, there is large difference in the absolute number.

However the absolute value of the discharge is not the relevant parameter in this case: also in the Uganda context it has been seen tha GLOFAS does not perform well in forecasting absolute discharge and the forecast is generally on the high side[[2]](#footnote-2). Relative comparisons are more relevant i.e. to understand if the largest peaks in GLOFAS are also the largest peak in the observations.

TEST TO DO : we should rank the GLOFAS yearly maximum peak for each station and see if there is a correlation with the ranked yearly max peaks in the observations

# Trigger Levels and skill of GLOFAS

Figure 2 shows the hindcasted riverine discharge at Katima Mulilo station (extracted from GLOFAS) for multiple years. Since these are simulated historical data and a forecast refers to the future, we refer to these simulations as “hindcast data”. Figure 3 shows the hindcast timeseries data (blue line) for a monitoring station in Katima Mulilo. is located in the Southern district at the border with Namibia and often experience floods during the rainy season (December - March). If we look at the historical data we can easily identify the peaks of discharge within a given year. Within the scope of this study we want to test the hindcast skill of this model for extreme events only.

Caughlan et al. (2016) describes a methodology to identify extreme events in a riverine discharge time series, by qualitatively selecting the 95% percentile as a proxy for extreme event

# 

Figure 5 Hindcast data from GLOFAS for the station of Katima Mulilo in the Lukulu district. The green, purple and red dashed line mark the 90%,95% and 98% percentiles respectively. The yellow dotted vertical lines mark the years when significant flood has been reported. Blue line indicate the median of the ensembles for time step 1

In general it is possible to select as a threshold values which are defined based on percentile and flow return period, for each GLOFAS station in ZAMBIA. This thresholds are defined based on the ensemble median values of glofas simulation with a time step=0 and used as a proxy for the amount of discharge that causes a flood. The list of the selected threshold are:

* 98 percentile,
* 95 percentile,
* 90 percentile,
* 5 Year return Period,
* 10 Year return Period,
* 15 Year return Period and
* 20 Year return period.

The figures below show these threshold on the simulated data.

A close up of a map

Description generated with high confidence

A picture containing text

Description generated with high confidence

# A screenshot of a cell phone Description generated with high confidence

For FBF activation our interest is to test the forecasting skill of GLOFAS with a lead time that gives as a window of opportunity to take actions, thus we did skill assessment of GLOFAS simulation with time step of 4 and 7. For each Threshold we calculate probability (based on frequency) for false alarm and hit, this was done by

* first comparing the maximum simulated flows against threshold (glofas signal to detect flood) for each year
* then we compared glofas signal with impact data. If impact data showed a flood was reported for that specific period a positive glofas signal is translated as hit. If glofas missed the flood event it will be a miss. On the other hand if Glofas indicated a positive signal (flooding) whereas no impact was reported for that specific year glofas signal is interpreted as false alarm
* For each GLOFAs Signal probability was calculated by comparing the % of ensemble members indicating the identified signal.

We can evaluate the skill of GLOFAS by evaluating variables calculated from both the observation and simulation sample:

– hit: event observed and simulated;

– miss: event observed and not simulated;

– false alarm: event simulated and not observed;

For humanitarian actors the most interesting indicator for forecast verification is the false alarm ratio (FAR) which is defined as the number of forecast-based actions that were not followed by a flooding, divided by the total number of actions that were triggered by the system( False alarm + Hit Rate). False alarm represents the proportion of actions that are taken “in vain”. Following the methodology in Weeink (2010), we can compute a number of scores to assess the skill of a forecast method. the False Alarm Ratio (FAR), the Probability of detection (POD) and the critical success index (CSI) (WMO, 2007). Which are defined as follows:

Where hits are events identified as above the threshold in the model and corresponding to a flood report for the same time and district. Misses are reported floods event that are not showing as above threshold in the model and false alarms are events above threshold in the model but not corresponding to a reported flood event for that time or place.

TABLE

# A system to detect an upcoming flood based on GLOFAS (trigger model)

* GLOFAS will release the forecast every 7 days. Forecast is for the next 30 days.
* The forecast is compared to the trigger level and the probability of exceeding the trigger level in the following 5 days is calculated.
* If prob\_exceed> XX than :
  + Calculate return period of incoming flood (already in glofas data)
  + Produce forecasted flood extent map based on return period info
  + Produce intervention map
  + Initiate communication for early action implementation

# 

1. We use a t-statistics instead than a z-statistic because the number of observed events is too low to know for sure the parameters of the distribution. [↑](#footnote-ref-1)
2. Erin Coughland private communication [↑](#footnote-ref-2)