Methodological Note: 2022 REACH - CCCM Flood Risk Analysis of IDP sites in Yemen

Background & objectives

Next to airstrikes and eviction, flooding is one of the major threats to Internally Displaced Person (IDP) sites in Yemen. Based on assessments of 1,563 IDP sites, estimates suggest that around 20 percent of IDP hosting sites in Yemen are at risk of flooding.[1] Moreover, between April 2021 and January 2022, 117 flood events were reported in 89 sites.[2] Since there currently exists no data source that shows the flood risk of all IDP sites in Yemen, REACH, in collaboration with the Camp Coordination and Camp Management Cluster (CCCM) Cluster, aims to better understand the extent of this issue. Overall, this analysis aims to provide a complementary data source for the 2022 flood contingency planning of the CCCM Cluster and OCHA. This methodological note explains how the 2022 CCCM Flood Risk Analysis will be conducted, following three main phases

1. REACH HEC-RAS modelling
2. Data Triangulation
3. CCCM Sub-National Cluster Coordinator (SNCC) field validation

Scope & Sources

This analysis aims to provide a Flood Risk Score for all IDP sites in Yemen. Overall, REACH analyzed a total of 2,428 IDP sites. The data sources to be used for this analysis will include the following:

- **REACH HEC-RAS modelling**: 12 basins covering 589 IDP sites with about 37% of site population, across 53 districts and 8 governorates (status: March 2022)
- **CCCM IDP Hosting Site Master List**: Official CCCM Master List presenting the total number of IDPs in IDP hosting sites per district (n = 2,331 sites), covering 212 districts across 22 governorates (status: January 2022)
- **CCCM Flood Report**: Official CCCM Flood Tracking system, established in 2021. 117 number of flood events reported in 89 sites in 34 districts across 12 governorates (status: March 2021 – January 2022)
- **CCCM Site Report:** Site information reporting whether flooding is considered a site threat based on Key Informant interview. 1,563 IDP Site reports available, covering 153 districts across 20 governorates (status: November 2019 – January 2022)
- **Site Verification Exercise**: Provides an overview of available coordinates both verified and not verified (n = 1,510 coordinates, status: March 2022)

[1] CCCM. Site Report. November 2019 – January 2022.
[2] CCCM. Flood Report. April 2021 – January 2022.

Expected outputs

- **Dataset** containing suggested Flood Risk Scores
- **National and governorate maps**
- **Presentations** of findings to CCCM Cluster and partners

Limitations

As part of this analysis, REACH identified the below limitations:

- **Limited data sources:** All main data sources from which flood risk score scan be derived, have diverging data gaps (i.e., HEC-RAS modelling, CCCM Site Report and Flood Report). Based on these datasets, REACH was able to provide a Flood Risk Score for 1,275 sites out of 2,428 sites (53%).
- **Site boundaries**: The exact site extent/site boundaries of IDP sites in Yemen are not available to this date. Thus, REACH/CCCM had to develop an estimate of the extent of each IDP site used in the HEC-RAS analysis. This was done by establishing estimated buffer radiuses based on population size, which may not be perfectly accurate since IDP population density per site is also not known (see section: Phase 1. HEC-RAS Analysis – Buffer radius).
- **GPS coordinates:** GPS coordinates are only available for 1,510 out of 2,453 sites. In addition, not all GPS coordinates have been verified, so they might contain errors. To estimate to what extent we can trust GPS coordinates, REACH will provide a simple scoring system for each coordinate.
- **REACH HEC-RAS modelling**
    - **HEC-RAS terrain data:** A 30-meter DEM resolution was used in the modelling, which likely underestimates the depth of smaller water flows and overestimates the extent.
    - **Hydraulic structures**: Bridges and culverts, piped drainage networks, irrigation canals and open channels have not been included in the hydraulic model.
    - **Precipitation**: The precipitation data used in different governorates are from different sources, and although they have been processed using scientifically established methods, they may differ significantly and there is currently no way to validate the data.
    - **Storm-events**: The storm events designed to be used as inputs for the models varied in duration (and intensity) due to the different methods used to disaggregate the data. Storms of 6 to 12-hour durations were used in our analysis, however flooding events can sometimes be caused by several consecutive storm events that can last for several days. This phenomenon was not modelled in this analysis.
    - **Big basins**: Large basins are difficult to model with HEC-RAS (i.e. Marib) and were not included in the current modelling. For these areas new methodological approaches (i.e. HEC-HMS) will need to be explored in the future.

Phase 1. HEC-RAS Analysis

In Phase 1, REACH will aim to develop Flood Risk Scores for IDP Hosting sites by modeling separate watershed basins through HEC-RAS. A two-dimensional (2D) unsteady flow hydraulic model was built using HEC-RAS to derive flood hazard and depth products which were then translated to a flood risk score. The results from these types of modeling outputs can provide a high-level understanding of flood hazards on a catchment-wide scale and help to identify flood susceptible areas, especially areas at risk of flash flooding.[3] Catchment areas with a higher overall number of IDP population and IDP population density were prioritized for this initial phase of the exercise, which included primarily basins from the Hajjah, Al Hodeida, and Taiz governorates.

Precipitation data

The model output products inform the extent, depth, and hazard of areas where flooding may occur based on extreme precipitation events of a 25-year return period. The model applies a direct precipitation method, where precipitation is applied to all cells generated in a computational mesh. Due to the lack of trustworthy precipitation data from weather stations in Yemen, data from Saudi Arabia was used in basins in the Hajjah governorate, near the northern border, and for all other basins modeled in this analysis, precipitation data derived from satellite imagery was used.

The precipitation data from Saudi Arabia is from the Abha city region, which is a coastal wadi area with similar characteristics as the catchments found in Hajjah, located approximately 100 km north of the Yemeni border in Hajjah.[4] The Saudi precipitation time-series data was processed into intensity-duration-frequencies (IDF) to generate rainfall temporal distribution data. The IDF data was then used to design a 12-hour storm with 10-minute steps for the 25-year return period using the alternating block method.[5] The storm data was validated using Google Earth Engine (GEE) by analyzing a real-life storms that occurred in the same coastal region in Yemen, bordering Saudi Arabia. The storms displayed very similar durations and total depths as the designed storm.

The satellite precipitation data used in the modeling is from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) dataset, which builds on 'smart' interpolation techniques and record precipitation estimates based on infrared Cold Cloud Duration (CCD) observations, incorporating station data and the spatial correlation structure of CCD-estimates in interpolation methods.[6] The CHIRPS data is processed through a series of statistical techniques to estimate extreme precipitation based on the

[3] Flooding that begins within 6 hours, and often within 3 hours, of the heavy rainfall (or other cause). Flash Floods can be caused by a number of things, but is most often due to extremely heavy rainfall from thunderstorms. Flash Floods can occur due to Dam or Levee Breaks, and/or Mudslides.
[4] Al-anazi, K. and El-sebaie, I. (2013) Development of Intensity-Duration-Frequency Relationships for Abha City in Saudi Arabia. International Journal of Computational Engineering Research, 3, 58-65.
[5] A designed storm is an artificial hyetograph that takes the precipitation depths for time intervals over a specified total storm duration.
[6] Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shraddhanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell & Joel Michaelsen. "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes". Scientific Data 2, 150066. doi:10.1038/sdata.2015.66 2015.

Generalized Extreme Value (GEV), a commonly used statistical distribution in precipitation analysis. The 25-year return precipitation values were disaggregated into 12-hour storm events based on sub-daily ratios of precipitation, also taken from the Jeddah region in Saudi Arabia, with similar characteristics as the Yemeni Coastal area.[7] Ratios for mountainous and transitional areas will also be used in future modelling analyses.

Terrain and infiltration data

The terrain data used to generate the computational mesh of the 2D flow areas for this analysis was a 30-meter resolution digital elevation model (DEM) product called AW3D30.[8] This product is available freely online thanks to an open license conceded by the Japan Aerospace Exploration Agency (JAXA). The 2D flow area extents were defined for each basin in the HEC-RAS processing environment, and computation points were generated also at 30-meter intervals. The datasets used to complement the analysis include landcover data[9] from the European Space Agency (ESA) and hydrological soil group (HSG) data.[10] An infiltration layer is created based on these two products, to account for the precipitation losses by infiltration and help determine the excess or net rainfall. For this purpose, the Soil Conservation Service (SCS) Curve Number (CN) approach is adopted, where estimates of the CN are a function of the different combinations of hydrologic soil groups and land cover types present in the terrain.[11] The CNs are derived from reference tables suggested in the HEC-RAS documentation. The landcover product has 11 different categories which are all assigned a roughness coefficient (Manning's value), also based on HEC-RAS documentation.

[7] Awadallah, A. (2015), Regional I-D-F for Jeddah, Saudi Arabia. J. Flood Risk Manage, 8: 195-207. https://doi.org/10.1111/jfr3.12085
[8] Japan Aerospace Exploration Agency (2021). ALOS World 3D 30 meter DEM. V3.2, Jan 2021. Distributed by OpenTopography. https://doi.org/10.5069/G94M92HB. Accessed: 2022-03-03.
[9] Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, Linlin, Tsendbazar, N.E., Ramoino, F., Arino, O., 2021. ESA WorldCover 10 m 2020 v100. https://doi.org/10.5281/zenodo.5571936.
[10] Ross, C.W., L. Prihodko, J.Y. Anchang, S.S. Kumar, W. Ji, and N.P. Hanan. 2018. Global Hydrologic Soil Groups (HYSOGs250m) for Curve Number-Based Runoff Modeling. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1566.
[11] SCS. 1956. Hydrology. National Engineering Handbook, Supplement A,Section 4. Soil Conservation Service, US Department of Agriculture: Washington, DC; Chapter 10.

Buffer radius

Information on the extent or boundaries of IDP hosting sites is not available in Yemen. Accordingly, REACH had to establish estimated buffer radiuses for each IDP hosting site. Based on the below two example CCCM Site Plans, REACH created **different estimated buffer radiuses based on IDP population size**. See Table 1 below for more details.

Table 1. Buffer radius of IDP hosting sites

	Very small sites	Small sites	Medium sites	Large sites	Very large sites
Population size	1-200	201-500	501-1,000	1,001-5,000	>5,000
Buffer radius	100m	200 m	300 m	500m	1km

The below two example CCCM Site Plans were used to estimate the site boundaries in Table 1.

- Site A: 80 families (around 480 people): 47 M radius
- Site B: 512 families (around 3,072 people): 303 M radius

HEC-RAS Flood Risk Severity Scoring

For the **HEC-RAS model Flood Risk Severity Scoring**, REACH will use two main indicators, which are based on flo**od hazard and flood depth** (see Table 2 below). Based on these two indicators, REACH will implement the below **steps** to draw up a HEC-RAS Flood Risk Severity Score per IDP site:

- **Step 1: Classify** the flood hazard percentage coverage (%) and flood depth coverage (%) per **severity class** (No risk – Extreme risk, see Table 2) through ArcGIS.
- **Step 2**: **Exclude** sites that do not cover the **minimum threshold** level of flood hazard (5% of total buffered site area) OR flood depth (10% of total buffered site area) to be included in Flood Risk calculations. For these sites **no score** will be determined due to a negligible flood coverage.
- **Step 3**: **Aggregate** the flood hazard coverage (%) and flood depth (%) per buffered IDP site to derive the **severity class** per indicator (see Table 3).
- **Sep 4:** Determine the **final severity scores per IDP** sites based on the **highest value** of the two indicators.

Table 2. HEC-RAS Flood Risk Severity Classes per Indicator

The below table shows the severity classes for flood hazard and flood depth.

Nr	Indicator	0 – No value (no risk)	1 – Low risk	2 – Medium risk	3 – High risk	4 – Very High risk	5 – Extreme risk
1	flood hazard[12] (m^2/s)	No value	<0.2	0.2 – 0.5	0.5 – 1.5	1.5 – 2.5	>2.5
2	flood depth (m)	No value	<0.5	0.5 – 1	1.0 – 2.0	2.0 – 5.0	>5.0

Table 3. Determination of HEC-RAS Flood Risk Score per IDP site

The below table will be applied to both flood hazard and flood depth indicators.

No risk	Low risk	Medium Risk	High Risk		
0 – No value	**1 – Low**	**2 – Medium**	**3 – High**	**4 – Very High**	**5 – Extreme**
No value	All remaining categories	If Severity classes 2, 3, 4 and 5 are >=5% of total buffer site area. **OR** If Severity classes 2, 3, 4 are are >=20% of total buffer site area. **OR** If Severity classes 2 and 3 are >=30% of total buffer site area. **OR** If Severity class 2 is >=40% of total buffer site area.	If Severity classes 3, 4 and 5 are >=10% of total buffered site area. **OR** If Severity classes 3 and 4 are >=20% of total buffered site area. **OR** If Severity class 3 is >=30% of total buffered site area.	If Severity classes 4 and 5 are >=10% of total buffered site area. **OR** If Severity class 4 is >=20% of total buffered site area.	If Severity class 5 is >=20% of total buffered site area.

[12] Flood Hazard is defined as = flood velocity x flood depth

<u>**Example: HEC-RAS Flood Severity Score Calculation per IDP site**</u>

Step 1: Classification of flood hazard (%) and flood depth (%) indicator per severity class

This classification is done automatically in ArcGIS.

Nr	Indicator	0 – No value (no risk)	1 – Low	2 – Medium	3 – High	4 – Very High	5 – Extreme
1	% of covered IDP site area by flood hazard	35%	0%	0%	10%	50%	5%
2	% of covered IDP site area by flood depth	35%	45%	15%	5%	0%	0%

Step 2: Exclude that do not cover the minimum threshold level of flood hazard (5% of total buffered site area) OR flood depth (10% of total buffered site area)

IDP Site	Indicator	0 – No value (no risk)	1 – Low	2 – Medium	3 – High	4 – Very High	5 – Extreme	Included/Excluded from analysis
IDP Site A	% of covered IDP site area by flood **hazard**	35%	0%	0%	10%	50%	5%	Included
	% of covered IDP site area by flood **depth**	35%	45%	15%	5%	0%	0%	Included
IDP Site B	% of covered IDP site area by flood **hazard**	99.99%	0.01%	0%	0%	0%	0%	Excluded from analysis, negligent flood hazard coverage (>0.05%)
	% of covered IDP site area by flood **depth**	99.95%	0.05%	0%	0%	0%	0%	Excluded from analysis, negligent flood depth coverage (>0.1%)

Step 3: Aggregation of flood hazard (%) and flood depth (%) per severity class

This aggregation is calculated in Excel, based on the ArcGIS values.

Nr	IDP Site	Indicator	0 – No value (no risk)	1 – Low	2 – Medium	3 – High	4 – Very High	5 – Extreme	Aggregation
1	IDP Site A	% of covered IDP site area by flood hazard	35%	0%	0%	10%	50%	5%	**55%**
2	IDP Site A	% of covered IDP site area by flood depth	35%	45%	15%	5%	0%	0%	**20%**

Step 4. Calculate Total Severity Score per IDP site

IDP Site A

Indicator 1: Severity Classes 4 and 5 = 55% = **Severity Score 4**

Indicator 2: Severity Classes 2, 3, 4 and 5 = 20% = **Severity Score 2**

→ Total severity Score = MAX of Indicator 1 and Indicator 2 = 4

IDP Site B

Indicator 1: Severity Classes 1 = <0.05%

Indicator 2: Severity Classes 1 = <0.1%

Total severity Score = No score, negligible flood hazard/depth coverage

Phase 2. Data Triangulation

Following the HEC-RAS modelling, REACH will **triangulate** the HEC-RAS Flood Risk Severity Scores with the other available data sources (i.e., CCCM Site Report, CCCM Flood Report, CCCM 2022 SNCCs Flood Estimates) to **estimate which IDP hosting sites are at risk of flooding**.

While the **HEC-RAS model** determines the extent to which an IDP site is at risk of flooding based on a designed storm, the **CCCM Flood Report** highlights sites where flooding actually occurred in 2021/2022. The **CCCM Site Report** reports whether a site is at risk of flooding, based on Key Informants perception. Overall, the CCCM Flood Report is considered the most **authoritative** dataset in this analysis, since it reports actual events. The CCCM Flood Report also allows REACH to validate its HEC-RAS model findings over time.

Table 4. Determination of Flood Risk of IDP sites

Scenarios	IF	THEN, ESTIMATED Flood Risk of IDP site				NOTE
		High risk	Medium risk	No/low risk	Not known	
Scenario A	Flood report recorded for IDP site	x				
Scenario B	2 out of 3 data sources suggest IDP site at risk of flooding	x				If possible, REACH/CCCM to investigate why third data source is at conflict.
Scenario C	Flood Report recorded for IDP site BUT CCCM Site Report does <u>not</u> report flood as site threat	x				CCCM SNCCs/partners will be asked for feedback. If possible, REACH/CCCM to determine where the conflict might stem from.
Scenario D	Flood Report recorded for IDP site BUT REACH HEC-RAS model does <u>not</u> suggest site at risk of flooding	x				If possible, CCCM/REACH to determine where the conflict might stem from.
Scenario E	HEC-RAS records Medium risk for sites and no other data source available		x			
Scenario F	Data source suggests IDP site is at no/low risk of flooding (incl. Flood Report, Site Report, HEC-RAS model)			x		

Scenario G	CCCM Site Report AND REACH HEC-RAS model at conflict				x	CCCM SNCCs/partners will be asked for feedback. If possible, REACH/CCCM to determine where the conflict might stem from.
Scenario H	No HEC-RAS modelling, CCCM Site Report OR Flood Report available				x	CCCM SNCCs/partners will be asked for feedback.
Scenario I	If low trust in IDP site coordinates				x	CCCM SNCCs/partners will be asked for feedback

Phase 3. Field validation by CCCM Sub-National Cluster Coordinators

Following the Data Triangulation and REACH's estimate on which IDP sites are at risk of flooding, CCCM Sub-National Cluster Coordinators (SNCCs) will be asked to review the findings. Especially for sites for which the different data sources provide different results, SNCCs will be needed to provide a field perspective. SNCCs can reach out to CCCM Area Coordinators or CCCM partners in the field to verify the results.

Ultimately, the CCCM Cluster will determine the total number of IDP sites at risk of flooding per governorate based on REACH's analysis and SNCC feedback.