

FLOOD IMPACT-BASED FORECASTING FOR EARLY WARNING AND EARLY ACTION IN TANA RIVER BASIN, KENYA

O.M. Otieno¹, H. S. Abdillahi¹, E. M. Wambui¹, K. S. Kiprono¹

¹International Centre for Humanitarian Affairs, Nairobi, Kenya

¹Corresponding author: osunga.michael@redcross.or.ke or mikeotieno87@gmail.com

KEY WORDS: community risk assessment, vulnerability, flood impacts, forecast-based action

ABSTRACT:

Kenya is mostly affected by floods during the March-April-May (MAM) and October-November-December (OND) rainfall. This often occurs along river basins such as the Tana river basin, leading to disruption of people's livelihoods, loss of lives, infrastructure destruction and interruption of economic activities. This study used openly available data on flood exposure, vulnerability, lack of coping capacity, flood impacts and observed satellite rainfall to analyse and predict forecast-based impacts in Tana river. Earth observation satellites including LANDSAT, sentinel 1 and 2 were acquired based on credible flood event dates to validate flood exposure and flood events. The community risk assessment (CRA) approach was used to delineate communities at high risk of floods using combination of data on vulnerability, flood exposure and lack of coping capacity. Using an ordinary least squares (OLS) predictive model, observed satellite rainfall was used as a covariate in order to predict flood impacts on communities with high flood risk scores in Tana river. Weighted scores from the CRA dimensions were summed up with forecasted hazards from the OLS model in order to derive a flood impact-based forecast. The flood impact information is to be used in forecast-based action through early warning, early action protocols thereby reducing impacts of potential floods in communities living in high flood risk areas based on the flood risk map.

1. INTRODUCTION

1.1 Forecast-based financing

Predictable extreme weather events such as floods lead to disasters that are often intensified by climate change. The impacts of these events can be mitigated if climate forecasts are thoroughly utilised for early action in order to prepare for disasters. Despite the availability of climate forecast, communities, governments and humanitarian agencies always act after a flood occurs. Yet there exists a window of opportunity between when a forecast issued and when the hazardous event occurs, where early actions can be taken to cushion the most vulnerable from the impacts of a flood.

Recognizing this window of opportunity and taking advantage of advances in science, data and technology, humanitarian organisations such as the Red Cross Red Crescent Movement have developed and piloted an approach known as Forecast-based early Action (FbA), in partnership with meteorological and hydrological services and other humanitarian agencies. Coughlan de Perez defines FbA as when a forecast states that an agreed-upon probability threshold will be exceeded for a hazard of a designated magnitude, then an action with an associated cost must be taken that has a desired effect and is carried out by a designated organization (Coughlan de Perez et al., 2015).

The FbA approach seeks to jointly develop standard operating procedures with key stakeholders where each stakeholder commits to undertake certain actions aimed at reducing the impacts of a hazard, when a forecast is issued. For example, in 2013, the Ugandan Red Cross Society with funding from the German Red Cross Society worked with communities in Northern Uganda and national flood management stakeholders

to define the actions that could be taken prior to a flooding event (Stephens & Erin, 2015).

1.2 Flooding in Kenya

Flooding in Kenya has been regularly documented since independence. The most severe flooding occurring in 1962-64 dubbed the "*Uhuru*" floods as it coincided with independence (Opere, 2013). Followed by the 1997 floods occasioned by the El Nino phenomenon, and most recently in 2015 and 2018 where severe flooding was observed across the nation. While these extreme events were felt nationwide, in the intervening years, floods have been observed in Kenya's five river basins particularly in the Western Kenya Lake Victoria basin and the Tana River Basin, which is the focus of this paper.

Kenya is mostly affected by floods during the March-April-May (MAM) and October-November-December (OND) rainfall (Nicholson, 2017), (Gamoyo, Reason, & Obura, 2015). This often occurs along wetland agro-ecological production systems such as Athi and Tana river basins (Leauthaud et al., 2013), leading to loss of lives, disruption of people's livelihoods, infrastructure destruction and interruption of economic activities. The most recent major flood in Kenya occurred during the 2018 MAM rainfall (United Nations Office for the Coordination of Humanitarian Affairs (OCHA), 2018). These led to displacement of over 150,000 people, 72 deaths, 33 injuries in 12 counties, namely; Wajir, Turkana, Garissa, Isiolo, Mandera, Marsabit, Narok, West Pokot, Samburu, Tana River, Kisumu and Taita Taveta (The International Federation of Red Cross and Red Crescent Societies (IFRC), 2018), (United Nations Office for the Coordination of Humanitarian Affairs (OCHA), 2018).

The floods led to livelihood disruptions with over 6000 livestock killed, 8450 acres of farmlands submerged in water, houses and infrastructure such as roads destroyed (Kenya Food Security Steering Group (KFSSG), 2018). 291,171 people who were displaced by floods in the 2018 long rains were at risk of disease outbreaks (UNICEF, 2018). The increase in stagnant water provided conducive conditions for Rift Valley fever (RVF), a mosquito borne viral zoonosis that mostly affected animals and human lives (Kenya Food Security Steering Group (KFSSG), 2018).

The main cause of flood waters in the Tana River catchment is rainfall in the Upper Tana (Opere, 2013) therefore observed weather forecasts and river gauge levels can be used to predict flooding impacts in the lower catchment. This paper aims to obtain and analyze credible reports on flood events and impacts in Tana river, to collate information on temporal river gauge levels and observed rainfall derived from satellites for both Tana river and the upper catchment areas, to investigate linear relationship between flood impacts and observed rainfall and river gauge levels and to predict flood impacts using a predictive model. The study uses openly available datasets for predicting impacts thereby strengthening FbA by enabling unbiased and low-cost targeting of wards at risk of floods for early action based on a flood impact map.

2. DATA AND METHODS

2.1 Study area

The study is conducted in Tana river county which is situated in the coastal part of Kenya. The county has an area coverage of approximately 35,375.8 km² and a population figure of 110,044 inhabitants. The county lies at an elevation ranging between 0 to 200 meters above sea level. Tana river's main economic activities are farming and nomadic pastoralism. The county comprises of 3 sub counties and 15 ward administrative units.

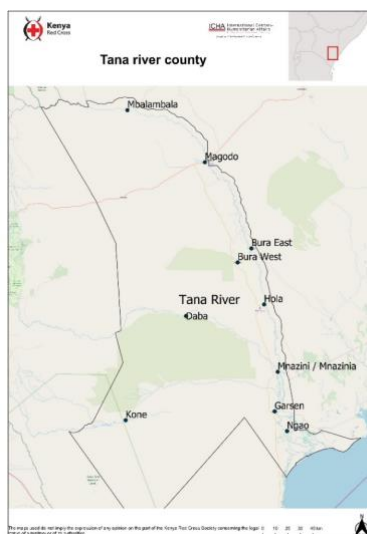


Figure 1. Study area map

2.2 Data acquisition

This research utilized data sources on flood events and impacts primarily from credible reports and earth observation satellites. Acquisition dates for satellite derived flood impacts was

concurrent to flood event dates in Tana river. Table 1 shows a list of datasets used for this study, their sources and characteristics (units).

	Variables		
	Variable name	Units	Source
1	Flood event dates	Date-month-year	IFRC-DREF, Desinventar, KRCS-EoC and WRA
2	Geo-coded flood impacts	Number of houses destroyed per flood events	IFRC-DREF, Desinventar, KRCS-EoC and WRA
3	Flood exposure	Flood extent in acreage	NASA and ESA / global disaster risk platform
4	Observed rainfall	Rainfall in millimeters	CHIRPS
5	Vulnerability, exposure and lack of coping capacity	Percentages	KNBS

Table 1. Data sources

IFRC-DREF, disaster relief emergency fund reports from the International federation of the red cross and red crescent societies; KRCS-EoC, Kenya red cross society - emergency operation center; NASA, National aeronautical space agency's Landsat satellite archives; ESA, European space agency's Sentinel satellite archives; WRA, Water resource authority; CHIRPS, climate hazards group infrared precipitation with station data; KNBS, Kenya national bureau of statistics

2.3 Methods

2.3.1 Tana river flood events

Tana river experiences major floods in the months of March-April-May (MAM) and October-November-December (OND) (Gamoyo et al., 2015) due to rainfall received upstream from neighboring counties, namely Meru and Tharaka. Credible historical flood events were extracted from the United Nations DesInventar disaster information management system, the international federation of the red cross and red crescent societies (IFRC) disaster relief emergency fund, Kenya red cross society - emergency operation center and Water resource authority of Kenya.

2.3.2 Tana river observed rainfall

Due to unavailability of ground weather stations from Kenya meteorological department in Tana river county, mean daily observed satellite rainfall for Tana river and the upper catchment areas in Meru and Tharaka counties were extracted from the climate hazards group infrared precipitation with station data (CHIRPS) (Funk et al., 2015). Observed satellite rainfall was acquired in tandem with reported flood events.

2.3.3 Correlation analysis

Correlation measures the amount of strength to which variables are linearly associated (Rubin, 2012). Pearson's correlation coefficient is used to test linearity between two or more variables with correlation values ranging between 1 to -1. Value 1 is a perfect positive correlation and -1 is perfect negative correlation. Zero denotes no linear association between the variables. The correlation analysis equation formula is as shown below.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}$$

Equation 1 Pearson's correlation analysis

n is number of pairs of scores

$\sum xy$ is sum of products of paired scores

$\sum x$ is sum of x scores

$\sum y$ is sum of y scores

$\sum x^2$ is sum of squared x scores

$\sum y^2$ is sum of squared y scores

2.3.4 Ordinary least squares analysis

An ordinary least squares (OLS) is a predictive modelling technique that strives to predict the value of an outcome variable based on one or more input predictor variables (Bruce & Bruce, 2017). The aim of this model is to establish a linear relationship between the response and predictor variable(s) in order to estimate the value of the response when predictor values are well-known. The response variable is denoted as y . The set of predictor variables would be denoted as $x_1, \dots, x_3$. The OLS of y on $x_1, \dots, x_2$ describes how y is related to $x_1, \dots, x_2$ and the error term using the equation;

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

Equation 2 OLS model

where y is number of houses destroyed by floods

x_1 is observed rainfall from Tana river

x_2 is observed rainfall from upper catchment areas

β are unknown parameters to be estimated

β_0 is the intercept

β_1 and β_2 are the slopes

u is the error term

The OLS model performance is to be determined by the coefficient of determination (R^2) commonly referred to as R-squared. R^2 is a measure of goodness of fit for an estimated OLS equation. Values of R^2 that are close to 1 indicate perfect fit, while values close to zero indicate poor fit. The R^2 implies the fraction of variance for the response variable that is described by predictor variables in the OLS model (Myers & Myers, 1990).

2.3.5 Flood extent maps

Flood extent maps are derived from earth observation satellites from the National aeronautical space agency (NASA) and European space agency (ESA) archives based on flood event dates in order to validate reported flood events and also to compute the spatial distribution of floods in Tana river. The environment used for satellite imagery processing and information extraction is google earth engine (GEE) due to its high computation capabilities and accessibility of historical satellite archives in one cloud platform (Gorelick et al., 2017). The normalized difference water index (NDWI) (Gao, 1996) is used as a spectral index of choice in order to extract flood extents from Landsat 7 (Jain, Singh, Jain, & Lohani, 2005) archives between 2008 to 2012, Landsat 8 (Nandi, Srivastava, & Shah, 2017) archives between 2013 and 2014 and sentinel 2 (Du et al., 2016) archives between 2015 to 2016. Image differencing of sentinel 1 radar archives (Huang et al., 2018) before and during a flooding event is used to extract flood information for flood impact years of 2016 to 2018.

$$NDWI = (NIR - SWIR) / (NIR + SWIR)$$

Equation 3 Normalized difference water index

Where:

$NDWI$ is the normalized difference water index

NIR is the near infrared

$SWIR$ is the short-wave infrared

2.3.6 Flood community risk assessment

Using the index for risk management (INFORM) approach, a community risk assessment (CRA) approach seeks to highlight the most vulnerable communities, the underlying conditions that make these communities vulnerable to flood hazard, their coping capacity and if these communities are exposed to flood hazards or not (De Groeve, Poljansek, & Vernaccini, 2015). In order to delineate communities at high risk of floods a combination of data on vulnerability, flood exposure and lack of coping capacity is used to highlight communities at risk of flooding. Integrating analysis from rainfall forecasts with information generated from the flood community risk assessment enables the population at high risk to act ahead of impending floods. Components for the flood CRA are obtained from the Kenya national bureau of statistics (KNBS). These components are grouped and weighted within the 3 INFORM dimensions in order to give a flood risk score.

$$\text{Flood risk score} = \text{Vulnerability} \times \text{Flood exposure} \times \text{Lack of coping capacity}$$

Equation 4 Index for risk management formula

2.3.7 Flood impact-based forecasting using statistical modelling

Weighted scores from vulnerability, lack of coping capacity and hazard exposure from the CRA dimensions are summed up with hazard forecasted from the OLS model in order to derive a flood impact-based forecast of number of houses likely to be destroyed by floods. This is as illustrated in equation 5.

$$\text{Flood impact-based forecast} = \text{Vulnerability} \times \text{Flood exposure} \times \text{Lack of coping capacity} + \text{Flood hazard forecast}$$

Equation 5 Flood impact-based forecast

3. RESULTS

3.1.1 Tana river flood events

From credible reports, flood events and impacts were obtained in areas along the Tana river. These events were spatially referenced with geographical coordinates in order to point out locations of where flooding has occurred with significant impacts over the years. This is as shown in figure 2.

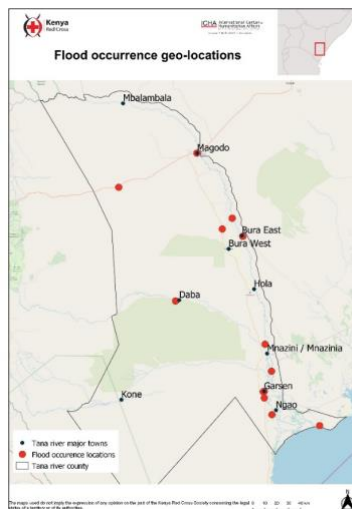


Figure 1 Flood event geo-locations

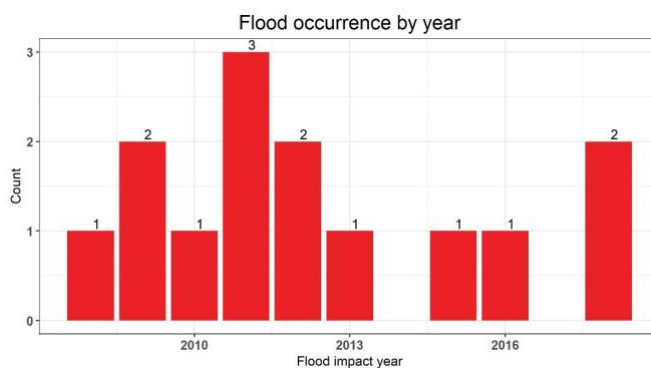


Figure 2 Flood event bar graph

From the credible sources listed above, 25 flood events were obtained from 2008 to 2018. Most of the flood events were reported in 2011 as shown in figure 3.

3.1.2 Tana river flood impacts

The flood impact of interest for this study was houses destroyed by floods which according to United Nations DesInventar's data dictionary is defined as number of homes that are either buried, levelled, collapsed or damaged to the extent that they are no longer habitable. Figure 5 shows the spatial distribution of houses destroyed by floods over the years with most impacts reported in Tarassa, Garsen and Bura areas.



Figure 3 Flood impacts spatial distribution

The year 2018 recorded the highest number of houses destroyed by floods. This is shown in figure 7.

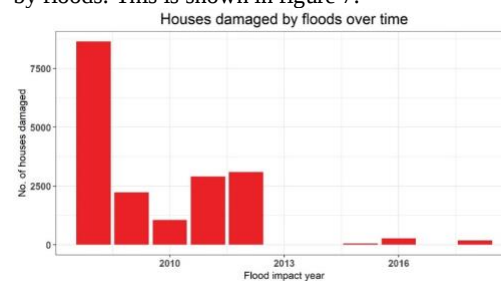


Figure 4 Houses destroyed by floods over time

3.2 Tana river flood extents

Flood extent maps were derived from earth observation satellites including sentinel 1, Landsat 7, sentinel 2 and Landsat 8 based on flood event dates in order to validate reported flood events and flood exposure map in Tana river. This is as illustrated in figures 9 to 12, where flood events derived from NDWI and image differencing gave concrete evidence of flood occurrence in the various flood event dates.

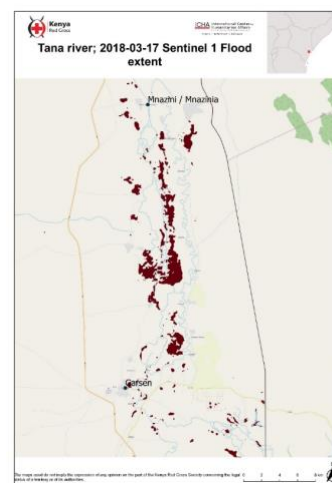


Figure 5 2018-03-17 Sentinel 1 flood extent

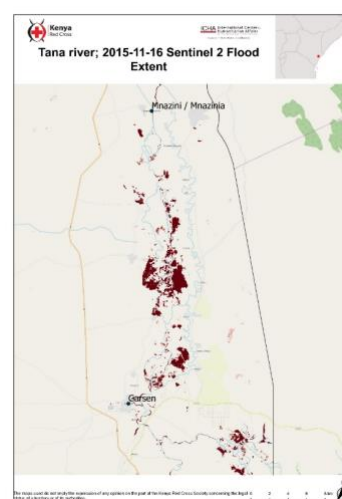


Figure 6 2015-11-16 Sentinel 2 flood extent

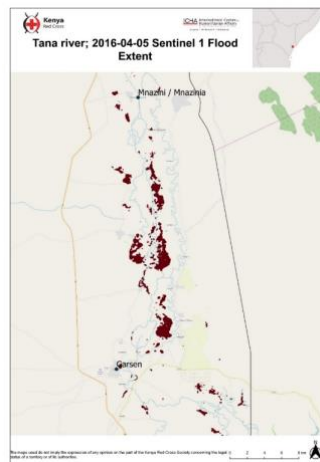


Figure 7 2016-04-05 Sentinel 1 flood extent



Figure 8 2014-11-16 Landsat 8 flood extent

3.4 Tana river observed rainfall

Due to unavailability of ground weather stations in the Tana river county, mean daily observed satellite rainfall for Tana river and the upper catchment areas in Meru and Tharaka counties were extracted from the climate hazards group infrared precipitation with station data (CHIRPS) (Funk et al., 2015). Observed satellite rainfall was acquired in tandem with reported flood events from 2008 to 2018. The highest observed rainfall in Tana river was in the 2015 flood event.

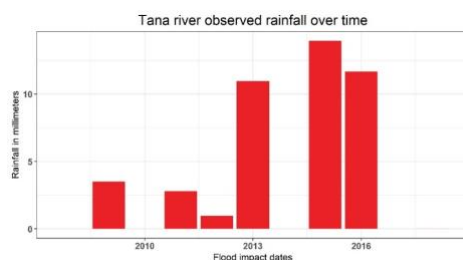


Figure 9 Tana river observed rainfall over time

3.5 Flood community risk assessment

From the flood community risk assessment most households are located in Madogo, Chewani and Kipini East.

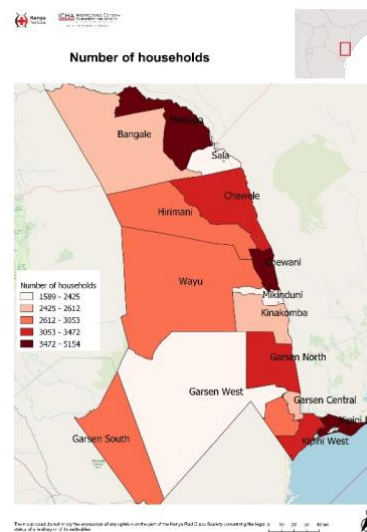


Figure 10 Households in Tana river

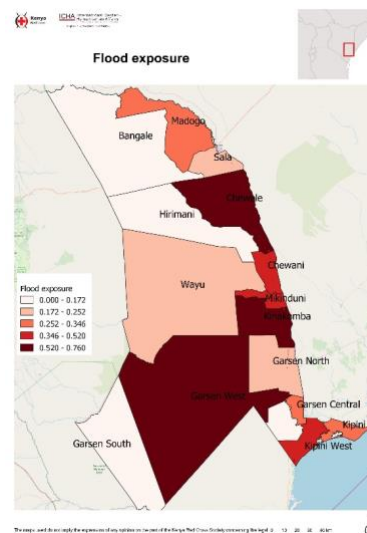


Figure 11 Tana river flood exposure

From the flood exposure layer, Garsen west, Kinakomba and Chewele are the most flood exposed wards in Tana river.

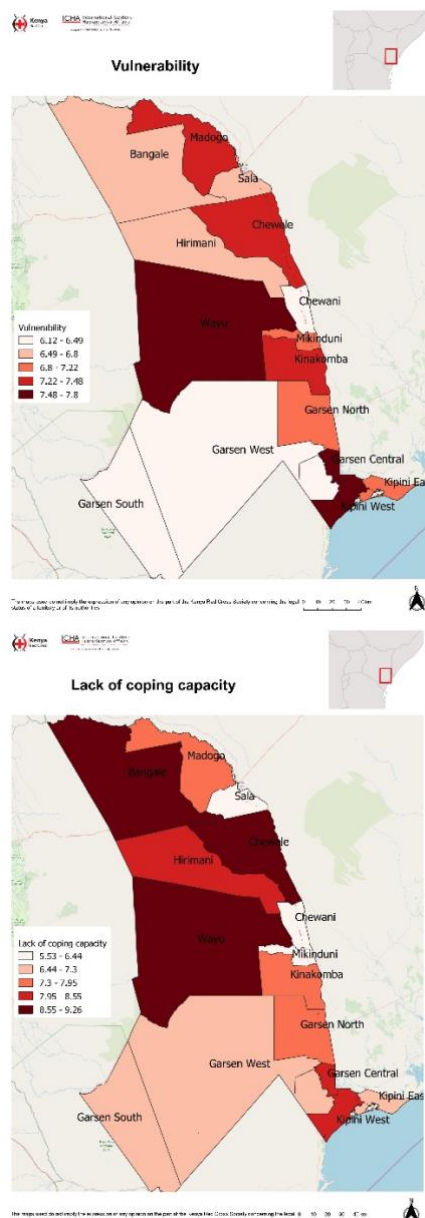


Figure 12 Vulnerability and lack of coping capacity

Kipini west and Wayu are the most vulnerable wards. Wayu, Chewele and Bangale wards lack capacity to cope with floods.

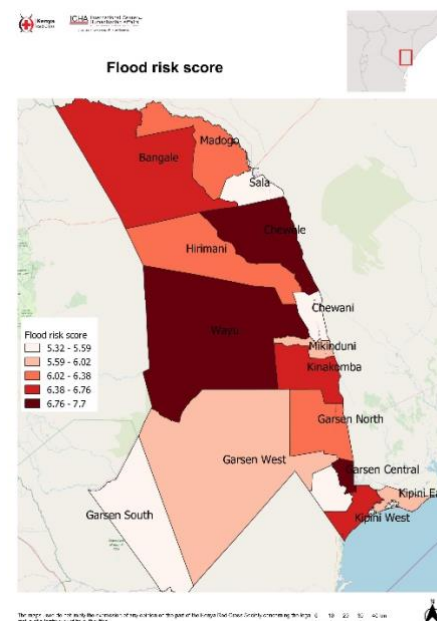


Figure 13 Flood risk score

Flood INFORM scores from vulnerability, lack of coping capacity and hazard exposure from the CRA dimensions were calculated with a multiplicative equation where each dimension was treated equally. Wayu, Garsen central and Chewele are at high risk of floods.

4. DISCUSSIONS

4.1 Discussions of results from correlation analysis

Pearson's correlation analysis was utilized in this study in order to test for linear association between number of houses destroyed and the predictor variables. Table 2 shows the linear correlation coefficients results from the analysis.

	Houses destroyed
Houses destroyed	1
Tana river precipitation	-0.212
Meru and tharaka precipitation	0.233

Table 1 Correlation matrix

A negative association of 0.212 exists between destroyed houses and observed rainfall in Tana river revealing that the lower the houses destroyed the higher the observed rainfall in Tana river. A positive association of 0.233 exists between destroyed houses and observed rainfall in upper catchment areas revealing that the higher the houses destroyed the higher the observed rainfall in the upper catchment areas.

4.2 Discussions from OLS model

Observed satellite rainfall was used as predictor variables while number of houses destroyed by floods was used as the response variable. An OLS model was used to link the predictor variables with number of destroyed houses. Significance levels (as p-values) were computed for each of these predictor variables where the probability level of each variable was set at $p < 0.05$. The predictor variable was used to predict number of destroyed houses using an OLS model in order to ascertain whether

observed rainfall positively or negatively influence houses destroyed.

	variable	Estimate	Pr(> t)
(Intercept)	(Intercept)	1195.338	0.14558
Tana river precipitation	precipitation	-106.8	0.26677
Meru and Tharaka precipitation	meru_tharaka_precipitation	46.06881	0.22852

Table 2 OLS outputs

From the OLS model estimates in table 3, observed rainfall in Tana river has a negative influence on the number of houses destroyed while observed rainfall in the upper catchment areas of Meru and Tharaka counties having a positive influence on number of houses destroyed by floods. The model has an R^2 of 0.25 meaning 25% of the variation is explained by the OLS model and the rest is due to error, therefore the model has a good fit.

4.3 Predicting flood impacts at high risk areas

Given a credible rainfall forecast in Tana river and the upper catchment areas, the OLS model would then be used to predict flood impacts in Wayu, Garsen central and Chewele which are at high risk of floods based on the flood community risk assessment.

5. CONCLUSIONS AND RECOMMENDATIONS

The results based on correlation coefficients suggest a positive correlation between number of houses destroyed by floods and observed rainfall in the upper catchment areas. Model estimates from the OLS analysis give indications that observed rainfall from the upper catchment areas positively influence the number of households destroyed by floods. This suggests that flood impacts in Tana river are mostly due to heavy rainfall received in the upper catchment areas. However, these results are not statistically significant. This means that no inferences could be derived from the model as p-values from the covariates are greater than the predefined threshold of 0.05. These results could be attributed to a low sample size achieved (25 flood events) in collection of historical flood events from the listed credible sources. Worth noting is that flood impacts can be as good as the predictor weather variables. This study recommends sourcing for other predictor variables that measure rainfall in Tana river and the upper catchment areas. The study also recommends sourcing for other flood impacts from other credible sources such as media reports to improve sample power. This will aim at improving model performance for a better prediction of flood impacts in Wayu, Garsen central and Chewele wards which are at high risk of floods based on the flood community risk assessment.

Acknowledgements

We acknowledge the support from the Kenya Red cross society - International centre for humanitarian affairs - and the Netherlands Red cross – 510 global for the support. All authors in this manuscript have no conflicting interests.

REFERENCES

- Bruce, P., & Bruce, A. (2017). Practical statistics for data scientists: 50 essential concepts. O'Reilly Media, Inc.
- Coughlan de Perez, E., van den Hurk, B., Van Aalst, M. K., Jongman, B., Klose, T., & Suarez, P. (2015). Forecast-based financing: An approach for catalyzing humanitarian action based on extreme weather and climate forecasts. *Natural Hazards and Earth System Sciences*, 15(4), 895–904.
- De Groeve, T., Poljansek, K., & Vernaccini, L. (2015). Index for risk management-INFORM. Concepts and Methodology (Version 2015); Joint Research Center of the European Commission: Luxembourg.
- Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., & Li, X. (2016). Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band. *Remote Sensing*, 8(4), 354.
- Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... Hoell, A. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2, 150066.
- Gamoyo, M., Reason, C., & Obura, D. (2015). Rainfall variability over the East African coast. *Theoretical and Applied Climatology*, 120(1–2), 311–322.
- Gao, B.-C. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3), 257–266.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27.
- Huang, W., DeVries, B., Huang, C., Lang, M., Jones, J., Creed, I., & Carroll, M. (2018). Automated extraction of surface water extent from Sentinel-1 data. *Remote Sensing*, 10(5), 797.
- Jain, S. K., Singh, R. D., Jain, M. K., & Lohani, A. K. (2005). Delineation of flood-prone areas using remote sensing techniques. *Water Resources Management*, 19(4), 333–347.
- Kenya Food Security Steering Group (KFSSG). (2018). The 2018 long rains season assessment report (p. 69). Retrieved from Kenya Food Security Steering Group (KFSSG) website: <https://reliefweb.int/sites/reliefweb.int/files/resources/2018%20LRA%20National%20Report.pdf>
- Leauthaud, C., Duvail, S., Hamerlynck, O., Paul, J.-L., Cochet, H., Nyunja, J., ... Grünberger, O. (2013). Floods and livelihoods: The impact of changing water resources on wetland agro-ecological production systems in the Tana River Delta, Kenya. *Global Environmental Change*, 23(1), 252–263.
- Myers, R. H., & Myers, R. H. (1990). Classical and modern regression with applications (Vol. 2). Duxbury press Belmont, CA.
- Nandi, I., Srivastava, P. K., & Shah, K. (2017). Floodplain mapping through support vector machine and optical/infrared images from landsat 8 OLI/TIRS sensors: Case study from Varanasi. *Water Resources Management*, 31(4), 1157–1171.
- Nicholson, S. E. (2017). Climate and climatic variability of rainfall over eastern Africa. *Reviews of Geophysics*, 55(3), 590–635.

- Opere, A. (2013). Floods in Kenya. In *Developments in Earth Surface Processes* (Vol. 16, pp. 315–330). Elsevier.
- Rubin, A. (2012). *Statistics for evidence-based practice and evaluation*. Cengage Learning.
- Stephens, E., & Erin, C. (2015). *Forecast-based action*. University of reading.
- The International Federation of Red Cross and Red Crescent Societies (IFRC). (2018). Emergency appeal, Kenya: Floods (p. 12). Retrieved from The International Federation of Red Cross and Red Crescent Societies (IFRC) website:
<https://reliefweb.int/sites/reliefweb.int/files/resources/MDRKE043ea.pdf>
- UNICEF. (2018). Kenya Floods Response, Update – 15 June 2018 (p. 2). Retrieved from UNICEF website:
<https://reliefweb.int/sites/reliefweb.int/files/resources/UNICEF%20Kenya%20-%20Flash%20Update%20-%2015%20June%202018.pdf>
- United Nations Office for the Coordination of Humanitarian Affairs (OCHA). (2018). Flash Update #6; Floods in Kenya | 7 June 2018 (p. 2). Retrieved from United Nations Office for the Coordination of Humanitarian Affairs website:
https://reliefweb.int/sites/reliefweb.int/files/resources/ROSEA_180606_Kenya%20Flash%20Update%20%2036_final.pdf