International Journal of Progressive Sciences and Technologies (IJPSAT)
ISSN: 2509-0119.

http://ijpsat.ijsht-journals.org
Vol. 28 No. 2 September 2021, pp.584-602

Geomorphological Appraisal of the Madogo-Mororo Floods using Neuro-Fuzzy Inference and the Multinomial Logistic Regression Models

Meshack Owira Amimo[1] and Jibril Ahmed Shune[2]

[1]Research Scholar, IIC University of Technology,
Cambodia

[2]Hydrology Research Intern, Garissa County
Kenya

Abstract*—The Tana River County has had serious issues relating to management of its riparian corridors and gazetted catchments. Floods have been an annual episode in Madogo-Mororo areas, with disastrous consequences every time they occur. Lives of both humans and livestock have been lost in every such episode, alongside the destruction of infrastructures and damages to properties, valued in the range of millions of Dollars. The causes do not directly relate to the rainfall volume registered whenever it rains, but rather to myriads of anthropogenic, geomorphological and hydrological variables. A combined statistical - and Artificial Intelligence-aided mapping of the study area was deemed a necessity as a DESK-TOP planning tool for flood risk assessment in the Study Area, as well as for detailed flood-risk mapping. The objective of this study was to identify all these three components, which involved mapping the longitudes, latitudes, elevations, soil types prevalent, vegetation density, grainsize of rocks or soil particles deemed prevalent in the surficial geology, culvert types constructed in the areas visited, hosing types, nearness of settlement to rivers, erosional intensity status of an area, as the input variables, to predict the output variable, flooding probability of an area. Visits were made to the Madogo -Mororo areas and the coded weightings of the aforementioned variables assigned, as deemed appropriate by researcher. For instance, soil erosion intensity was coded as 1, 2, and 3, with the numbers respectively coding for low intensity, moderate intensity and high intensity. The R software modeling libraries were used in this exercise. The neuro-fuzzy algorithms were then used to develop prediction models for floods, based on these coded parameters to assess then accuracy levels of the algorithm test datasets were used, generating over 98.4 percent levels of accuracies. The logistic regression models were then used, for comparison, also recording accuracy levels of 98.4 percent. This affirmed the supremacy of fuzzy logic and logistic regression as useful assessment tools for floods probability mapping in the study area. A code value of 1, represented a high probability of an area being flooded (flood risk areas), and a code of 2 was used to represent high-probability of moderate risk to flooding, whereas the value of 3 signaled a risk-free high-probability in an area. These probability codes were then plotted in a GIS map to help display the geospatial models of flood risk in Madogo and Mororo areas. The study notes the role of human activities and geomorphic factors in exacerbating the propensity to flooding episodes in the study area. Summarily, this study validates the use of neuro-fuzzy inference and logistic regression as dependable mapping tools for flood episodes in Madogo Mororo areas, used alongside the GIS techniques.

Keywords—Floods, Riparian Corridor, Neuro-fuzzy inference, R Software, Logistic Regression, Madogo, Mororo.

I. INTRODUCTION

The entire livelihoods and economy of Madogo residents has been repeatedly ravaged by floods nearly every year, at times, even twice a year. Efforts on flood control, design and construction of structural measures and taking proper measures for flood mitigation including non-structural measures like flood plain zoning, flood forecasting and warning shall come in handy in order for the town to realize its development potential in the coming years. Forecasting and prediction models on which areas shall be flood prone over the years require accurate and reliable data for flood estimation using statistical and or deterministic methods (Kumar et al, 2015).

The term river flooding may be defined as the high flow of water or run off, which exceeds or surpasses the capacity, either the natural or the artificial banks of a flowing stream. The flooding resulting from excessive rainfall on the ground surface, land, streams overflowing channels or the un-characteristic high tides or waves in coastal plain lands. Though, the following factors/variables may singularly or jointly influence flood frequencies (Acreman, 2013):

i) Landscape location

ii) Configuration of the landscape

iii) Soil characteristics

iv) Soil moisture status

v) Environmental management issues relating to catchments and floods

It is noted that some of the most critical variables that determine the occurrences of floods are rainfall event characteristics, depth of the flood, the speed of the flow, and duration of the rainfall event being studied in the project area. Also, there are many different types of flooding, the most common types of which are the river floods and flash floods. It instructive that much of the time, floods will occur in Madogo even when it is not necessarily raining at all, as long as it is raining in the Mt. Kenya highlands. The flooding episodes are heavily controlled by soil type and the gradient of land as river Tana is located downstream whereas Madogo Township is located on a slanted piece of land upstream. The vegetation cover progressively destroyed by charcoal burning in Mororo area has also exacerbated the progression of flood damages in the project locality.

A. Geomorphology of Project Area

The project areas geomorphology plays a critical role towards enhancing or diminishing floods probability. To that effect, it is imperative that one studies the specific contributions to the geomorphic factors active in the area. Madogo stands on a raised platform of high ground, relative to Mororo located on its Eastern conclaves. Owing to this, all the runoff inflow from Madogo seasonal streams flow in the direction of Mororo, and on the way empties its input waters to the riparian corridor and catchment systems thereby increasing volume of moisture into an already saturated ground. The terminal end-point of this flow is the river Tana channel.

The soils predominant in the area are mostly of the loamy and sandy species which facilitate the rapid infiltration of the runoff flow into the shallow sub-surface. Where the surface material is predominantly clay, flooding occurs almost instantaneously since the clay cannot admit/ transfer the water into its sub-surface systems. The agriculture being practiced in both Madogo and Mororo areas have the negative effect of exacerbating the rate of erosion into the flow systems such that much of the water arriving into river Tana, flowing from Mororo and Madogo contain high volumes of sands, clay and silts, which implies diminished reservoir/ storage capacity of the river bed at the end of the runoff flow episode. The implication here is that any slight episode of rainfall will cause flooding, since the flowing runoff does not have a space to accommodate its input within the river flow channel.

The vegetation cover is somehow sparsely distributed in the project area and comprises mainly of the xerophytes and allied short undergrowth which are unable to check the velocity of the flowing waters from Madogo enroute Mororo and the R. Tana flow systems.

The soil grain size is lowest in Mororo and highest in Madogo areas, making the flooding impact more in Mororo than in Madogo.

The dendritic drainage of the Lower Tana Hydrology makes it possible for so many streamflow inputs to converge in Mororo and the river channels, thereby exacerbating the damage of floods amongst the population residing within or in the proximity of the riparian corridor. The flow velocity is relatively higher in Madogo than in Mororo due to the differential slope coefficient, between the two settlements. The Mororo area is relatively flatter whereas Madogo is undulating in terms of its sloping characteristics.

II. Literature Review

Man-made structures in the Madogo tributaries have changed the hydro-morphological characteristics of the ephemeral flow regimes, since the seasonal streams have lost the high morphological suitability and high variation in floodwater and sediment-load that used to help the tributaries generate high transport velocity (Reisenbuchler,et al, 2019) for transporting water into River Tana, thanks mainly to human activities like buildings, farming, and sand-mining

Floods usually occur as result of heavy rainfalls in the Northeastern parts of Kenya. The run-offs rate of infiltration is usually lower than the rate of over-land flow, making the water accumulate on the ground surface due to the predominantly clayey nature of the land. Whenever it rains heavily, seasonal tributaries from as far away as Bangale and Tula-Abakiik areas bring in flood-water into the Madogo area, as their south-west north-easterly flowing streams direct their input into the Madogo area, flowing towards River Tana.

Madogo is actually an undeveloped wetland. According to Nyunja et al (2013), In Eastern Africa, the wetlands zones provide resources at a multiplicity of both geospatial and temporal levels. These are through the practice of farming, the art of fishing, the culture of livestock rearing and a host of other ecosystem services, deemed vital enough to sustain the local economy and individual livelihoods. The flooding episodes usually destroy the people's livelihoods, rendering them a begging social bracket at the mercy of relief feeding programmed by both the national government and donor agencies like the Kenya Red Cross Society and WASDA-Wajir South Development Agency. This is a tragic social dimension given that in the pre-flood moments, these affected communities usually supply green groceries and fruits for self-sustenance. As the flow event is on, the seasonal streams traverse the Bula-Kunaso, Madogo, Adele, Kona Punda and the Hatata-Mororo-Maramtu areas. In the upper Madogo areas, the floods occur due to the destroyed river channels arising from multiple of reasons. One, the river banks have been eroded to the point whereby the flow meant to be confined on the riverbed now moves outside the river channels. The velocity of the flow is subsequently checked, and the stagnant or semi-stagnated flow ends up splashing onto people's residences. Before the riverbanks were eroded and river channel depths tampered with due to sand harvesting and allied activities, the velocity of flow would be so fast, to the point that very little water would be left standing, after the rainfall event along the ephemeral channels.

As the water now flows towards the Adele-Mororo stretch, the velocity is quite low, and it floods the surrounding farms and neighborhoods in the study area. As it nears the Kona Punda area, the back flow from the river Tana tributaries flowing towards Madogo meets with these streams, and the water now forms a shallow lake in the area. Due to the absence of sufficient culverts and poor road design on the Garissa-Mwingi Highway, the water takes too long to drain across the road, owing to few culverts. This causes the flooding to persist for too long a time. This is a north –southerly flow.

The west-easterly flowing floodwater flows perpendicularly towards the River Tana course at Mororo, and it goes through small scale farms used by the peasant settlers in the area. Due to the unregulated invasion of the riparian corridor of the river Tana catchments, the Mororo-Kona Punda and Maramtu areas flood, displacing thousands. The farming activities are unregulated, some of which lie on the flowing channel-ways of the seasonal tributary systems.

Houses have been built near or within the tributary flow systems of the ephemeral streams, blocking speedy inflow into the larger River Tana. This exacerbates the problems of displacements during flooding.

Whenever it is a rainfall season in the Mount Kenya zone, comprising Muranga, Nyeri, parts of Isiolo, and Meru areas, the river Tana floods the downstream areas, and bursts its banks in the upper Tana zones. The remaining flowing waters in the River arrives in Garissa at very high flow velocity. Where the river encounters some embayment, flow velocity is checked and in the process, the ensuing back flow is directed Madogo-wards via existing tributaries. This point is of note in that floods usually occur in Madogo when it may not even be raining at all in Garissa or Madogo-Abakiik-Bangale areas. The displaced populace builds /erects tents by the roadside in Mororo along the highway all the way to Madogo secondary school, located over seven kilometers away. During this period, pneumonia, malaria and malnutrition takes toll on the displaced.

It is with this in mind that a study at possible intervention tools has been suggested in the present study to help predict and mitigate as appropriate. Soil types, slope angles, river morphology (widths and depths of seasonal streams transporting flood waters), longitudes, latitudes, elevations, proximity of housing to seasonal streams, vegetation depletion levels, coarseness of soil grains, amongst other factors, were studied using a coded scale, and the resulting data frames were analyzed using machine learning algorithms, which clustered the variables into factors. Supervised classification was then used to make flood prediction models (Mousavi et al, 2018) for areas with known geospatial coordinates, where development is imminent-whether the areas are appropriate or not for building, farming or settling.

III. RESEARCH METHODS USED

For the collection of field data, field work and visits to relevant offices to acquire data were undertaken between June 2020 and January 2021. Also, some trips were made in months of March 2021, to help avail data on Madogo and Mororo areas, relating to the soils and flooding.

Field visits were undertaken to the relevant regional and water offices, namely, the sub-county water offices and the WRMA offices. Here, the secondary hydrological and soils data-bank in the relevant offices shall avail the required data. Area chiefs and the local sub county water officer were visited and fully briefed on our study for the purposes of mapping the flood frequencies of the Madogo area.

Roads and schools were visited and assessed for info deemed to be of relevance to the study. A team of four, comprising the Hydrologist and three technical crew, then visited the township and assessed flood damage levels. The team went into individual residential plots where flooding had chased people in the past events and took in the technical info relating to geospatial coordinates, soils, vegetation, land use, soil types, rock types dominant, slope angles and other variables. Relevant maps displaying the drainage hydrology and soil characteristics of the study area were then generated and labeled to aid the field work. The tabulated dataframes were then analyzed using the R and python open source softwares to aid understanding into the flooding history of the area and contributions of the various variables mapped. The socioeconomic variables and physical variables were then entered in to an excel table dataframe to aid statistical analysis. Models will be prepared that can be used to predict the variables in question-there will be both contour kriging and the Machine learning models. Areas that have not been flooded and whose data exist shall then be test-run using the models generated to verify the efficacy of the models as diagnostic tools for flood prediction in arid Madogo area.

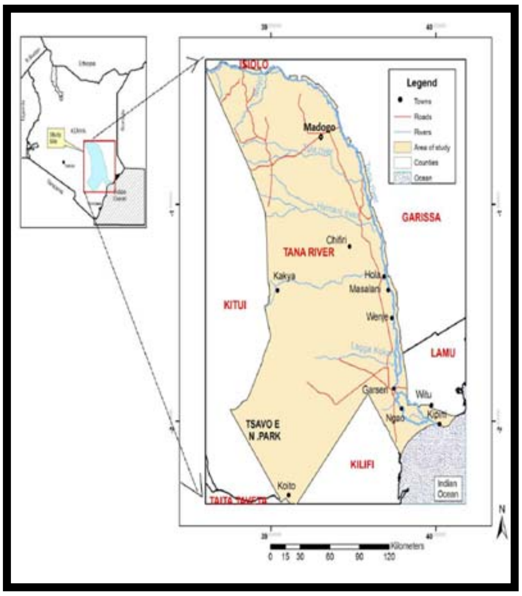


Figure 1: Map of The Study Area Adopted from a Case Study of Lower Tana Basin in Kenya, 2017.

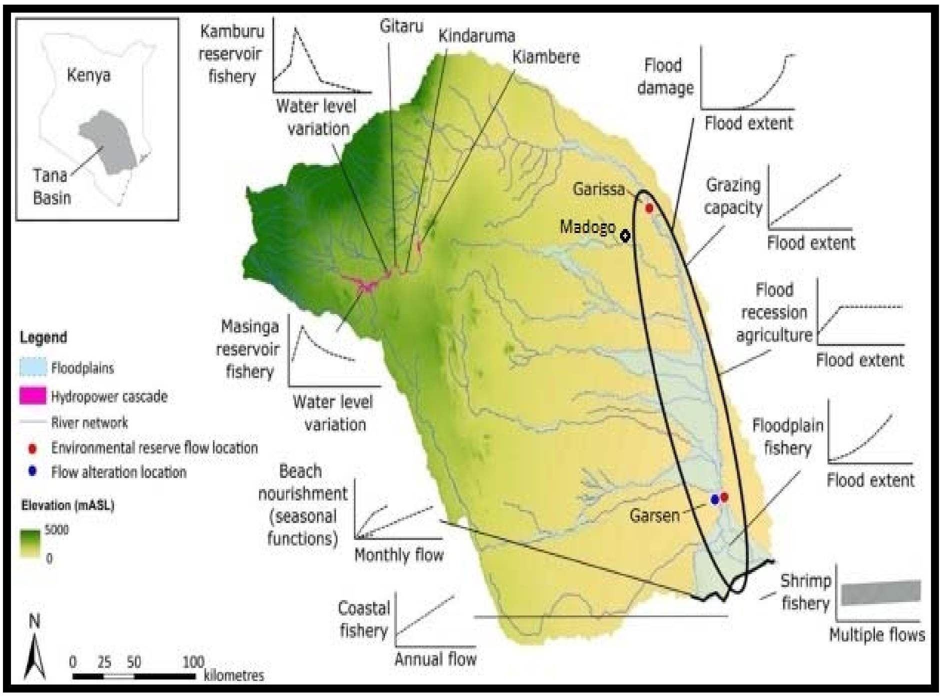


Figure 2: Map Showing Madogo Area in Relation to Floods and Floodplains-as adopted from River Basin Trade Off Analysis, 2020.

A. DESCRIBING THE VARIABLES USED TO PREDICT FLOODING IN MADOGO

1) Longitudes

The variable represents the coordinates of the surface of the earth being studied in respect of flooding occurrences. It could thus represent the geology, hydrology or hydrometeorology of the terrain which makes it vulnerable to flooding. The study and analysis shall reveal the actual contribution of longitudes to the understanding of floods in Madogo area.

2) Latitudes

The variable latitude goes hand in hand with longitudes as a co-coordinate variable, but may either speak the same story or a different one from the longitudes in relation to flooding occurrences' and the study shall determine the extent to which this variable plus others may positively bear contribution to floods occurrence in Madogo

3) Elevation

The variable also represents the coordinates of the terrain being mapped with respect to flooding and may be a real factor in that the raised ground are relatively free of floods whereas the relatively undulating or flat-lying grounds are floods-prone.

4) Soil types

The sandy soils soak up all the run-off pretty faster owing to the high porosity, and it is thus rare that a sandy-dominated terrain impounds water long enough to cause flooding to occur. The highly plastic and impermeable clays impound water, hence are more flood prone. The porous and fractured limestones and gypsite species may also be safe grounds that are floods free.

5) Vegetation density

The vegetation density determines the ease with which an area floods. In the present study the intensity of vegetation growth was used as a guide to help predict flooding. The areas with high density of vegetation are observed to be relatively flood risk-free. The coding tabulated for the modeling is shown in the table for vegetation.

6) Grainsize

The grain-size of the material dominant on the surface such as sands, loams, clays, fractured gypsites, et cetera, will determine the ease with which the ground admits runoff flow into the sub surface. If the material is highly porous, the water will get admitted fast and therefore be of limited vulnerability to flooding unless the rainfall event proceeds for hours.

If the grain-size is small, such as in the case of clays and allied clayey material, the ground harbors and impounds water, as it cannot speedily percolate and infiltrate into the ground. The flooding vulnerability risk gets higher, thus.

7) Culverts

Here, the poorly drained portions of the roads that stand high risks of flooding usually get fitted with culverts. Highway engineers and hydrologists usually co-inspect the roads and come up with recommendations as appropriate-on where to put the culverts. The portions so fitted are relatively risk-free. The other portions are not. However, there are culverts that area poorly designed and cannot transmit flood waters as fast as desirable, rendering the road portions along which they are erected to flood risk. There are higher risks in the areas that have no culverts at all

8) Drainage

The streams, tributaries, and sub tributaries that flow into and out of an area determine its flood risk vulnerability. The coding is shown in the tables indicating whether an area is well drained, moderately drained or poorly drained altogether

9) Housing

This is the variable used to indicate the type of housing, whether permanent or temporary, and also its closeness to flood plains. The variable has been coded as appropriate.

10) Nearness to the streams

It is noted that the further away a settlement is, the relatively safer it is from effects of flooding. Coding has been administered as appropriate to aid the present quantitative study

11) Degree of erosion in the terrain

The areas with advanced erosion levels like rill and sheet erosions appear to be more at risk of flooding than areas with minimal erosion levels.

It is imperative that the coding is stated here to represent the intensity of various phenomena mapped in the study.

TABLE 1. THE TABLE OF VARIABLES USED WITH THEIR RESPECTIVE NUMERICAL CODING

Soil type	Grainsize	Culvert type	Drainage Density	Housing Type	Nearness To Stream Courses	Erosion Intensity	pFloods
1-pure sands 2-loams 3-clays 4-coralites 5-gypsites 6-hard rocks /pebbles	1-fine grained 2-medium grained 3-coarse grained	1-no culvert 2-destroyed culvert 3-operational but small 4-good culvert	1-poorly drained 2-average drainage 3-well drained	1-no housing 2-semi permanent 3-permanent housing 4-houding on flood plains	1-very close to river –less than 50m away from stream 2-over 50m but less than 100m away 3-over 100m away from stream bed	1-very high 2-moderate 3- low 4-very low	1-high risk flood 2-average risk area 3-low risk area

IV. Theory Of The Algorithm

ACTUAL STUDY, METHODS AND MODELS FROM FUZZY LOGIC AND LOGISTIC REGRESSION

A. FUZZY LOGIC AND NEURO FUZZY INFERENCE

The fuzzy logic method may be safely combined with the traditional neural networks to form a hybrid method of flood mapping. This is known as the neuro-fuzzy inference.

1. Background

Science is never perfect. A serious researcher has to find a tradeoff that enables him to model using both précis and imprecise information to generate models and allied outputs.

Lotfi Zadeh pioneered the science of mathematical fuzzy logic in 1965, when it became apparent in the scientific community that human reasoning was imperfect, and yet the theories like the Newtonian Laws were working just fine in spite of not being able to address precisions at the molecular levels. Even classical physics had begun working with averages, since the exact number of molecules could not be established in statistical mechanics, a branch of theoretical physics. The paper by pioneer researcher in the field, Lotfi Zadeh, did have an everlasting impact on the thinking about uncertainty REASONING, since it challenged not only probability theory as the sole representation for uncertainty but also addressed the very critical mathematical foundations, upon which the theory of probability was presumed to be based. Much of the scientific info presently in our possession about natural processes is to a large extent a product of the analysis of imprecise data, on the account of the imperfect or imprecise human reasoning. The aspect of imprecise information may be quite useful to human thinkers/ researchers. The ability of scientist to find a way or ways of accommodating such imprecise reasoning and imprecise data into unimaginable situations and difficult scientific challenges of nature happens to be the sole criteria by which the success of fuzzy inference is judged today, globally. Without any doubt, such levels of reasoning may not practically aid the precision required for shooting beams of laser over hundreds of kilometers into the space, during space research programs, or impact positively on the levels of accuracy required in nanometrics. However, most human challenges in a natural world do not really need these levels of precisions, to surmount. One may consider driving or navigating motor cars or lorry on the highway, judging who is the most beautiful woman on the planet, computing the EXACT amount of maize flour one uses to cook maize-meal dish using two liters of water daily, and how much fuel one ACTUALLY utilizes in his car to travel a twenty kilometer journey daily. Implying that precision must be attained in technological models and products is akin to insisting on high-cost and long lead- times in the process of industrial or factory production and development. Moreover, in the natural setting, some projects are such that expense involved is proportional to precision, so that if more precision is a priority, then greater costs will be involved as well. In the consideration of using of fuzzy logic metrics for a certain task, the researcher or scientist ought to critically think in terms of the need for exploiting the tolerance for imprecision.

In the present context, fuzzy logic was employed to map floods, by using the hybrid algorithm combining neural networks and fuzzy logic. The hybrid algorithm is the neuro-fuzzy rule-based system. If for example there's is flooding at longitude 200-400 Metric Units, moderate floods between 401 to 601Metric Units, then the imprecise reasoning comes up to be something like this:

a) There's flooding in the longitude range 180 to 420 metric Units.

b) There's moderate flooding in the longitude range 381 to 621 metric Units.

c) One may now see the overlap in the two sets of events, high floods risks and the moderate floods risks. Remember 180 to 420 metric units belonged to the flood risk longitudes. But now moderate floods risk begins from 381 to 621. This means that even the value of 420m considered ideal for floods free effect now has some degrees of moderate flooding in the foregoing example

Membership functions may be sketched which will highlight the reasoning involved here and will also show the degrees to which the floods-free zone is a member of the moderate-floods zone, and also the extent or degrees to which the moderate-floods zone of longitudes also belong to the floods free longitudes. This is a more powerful modeling tool, unlike the conventional Boolean algebraic probability which is a 1 or 0 for flood –risks being mapped, in otherwise, implying that an area is being declared as being wholesome floods-risks free or wholesome floods –risks moderate. This is not true in real world setting. An area may be 95 percent floods free but still be 5 percent floods prone. Also, an area which is 90 percent floods moderate may be floods-free by a degree of ten percent, in real life, as per the foregoing example. Floods mapping have taken place on this Rivers catchment in the past, but NONE was performed using the ML algorithms.

2. Tana Floods Previous Studies

Andrews et al (1979) undertook a general study on the ecological variables along the Tana catchments that had promoted flooding propensities.

Leauthaud, et al (2013) researched on floods and livelihoods along the Tana Delta, with respect to how the flood outputs may be used to promote food security, akin to the proverbial "turning a curse into a blessing" kind of situation.

Jia et al (2016) mapped coastal flooding hazards using Fuzzy logic. In the paper, he opines that the major challenge in evacuating the subjects happens to be the imprecise, incomplete and spatially varying nature of the crisis information availed and has to be used during the critical phases of decision making. He argued in that paper, that fuzzy-logic based algorithm may be combined with GIS methods to effectively help facilitate an evacuation decision-making relating to helping save people's lives and property in emergencies. The method was found to accommodate both qualitative and quantitative info simultaneously, thereby avoiding the sudden changes of decisions impacted by uncertainties. The approach helped in speedily evaluating the spatial necessity for evacuations, in the coastal city of Bordeaux, in France.

Zhong et al (2019) also performed a study using fuzzy logic to map and predict floods in China. The project locality was a mountainous watershed in Southern China known as the Yangshan County.

B. HOW THE METHOD WORKS

1. Case of scanty data

Data may be scanty to a researcher so that he has no choice but employ this approximate reasoning to get to analyze a phenomenon such as the floods probability risk in an area. Consider the example hereunder:

Suppose that the Longitudes of an area mapped are such that the area has Longitude range of between 2000 metric units (minimum) and 5000 metric units (maximum). Latitudes may be of the range 6000 to 10000 metric units. Floods risk may also range between the longitudes and latitudes in a manner suggesting 20 to 74 percent risks for the minimum longitudes to the maximum longitudes, and this is the same for the latitudes as well. This means that an area in between the minimum and maximum longitudes will have average risk, which is the mean of 20 and 74, and this is 47 percent risk. We need to map the flood risk of a new area given to have longitudes 4400m etic units, 8000 latitudes, and let the floods risk be X. to solve the problem, one need to write the linguistic modeling ANTECEDSENTS and CONSEQUENTS . For example, IF X is x1, the Y is x2. ONE MAY WRITE THE SAME AS FOLLOWS:

i) IF LONGITUDES ARE LOW(2000-3500),and the LATITUDES ARE LOW(6000-7000), then the flood risk is LOW(20 to 35 percent)

ii) IF LONGITUDES ARE average(2500-4500),and the LATITUDES ARE AVERAGE(7500-8000), then the flood risk IS MODERATE (30 to 55 percent)

iii) IF LONGITUDES ARE HIGH(3500-6500),and the LATITUDES ARE HIGH(8500-10000), then the flood risk IS HIGH (50 to 74 percent)

iv) IF LONGITUDES ARE LOW, then the flood risk is LOW

v) IF Latitudes ARE moderate, then the flood risk is moderate

The ensuing process of defuzzifiction may then be used to predict the floods risk value for the problem assigned using modeling software such as MATLAB or R. one notices that fuzzy logic perfects the art of modeling using words. This reasoning has to be based on the fact that the hypothetical area being studied has no much data and that the researcher desperately wishes to use his expert prior knowledge of the area's floods history to model the uncertainty involved. In the vent that the area has sufficient data, then normal statistical modeling/ analytical tools may be employed to map the area's floods.

2. *Membership functions*

The membership may be displayed in terms of the following types of functions

a) Trapezoidal MF

b) Triangular MF

c) Gaussian

d) Sigmoid

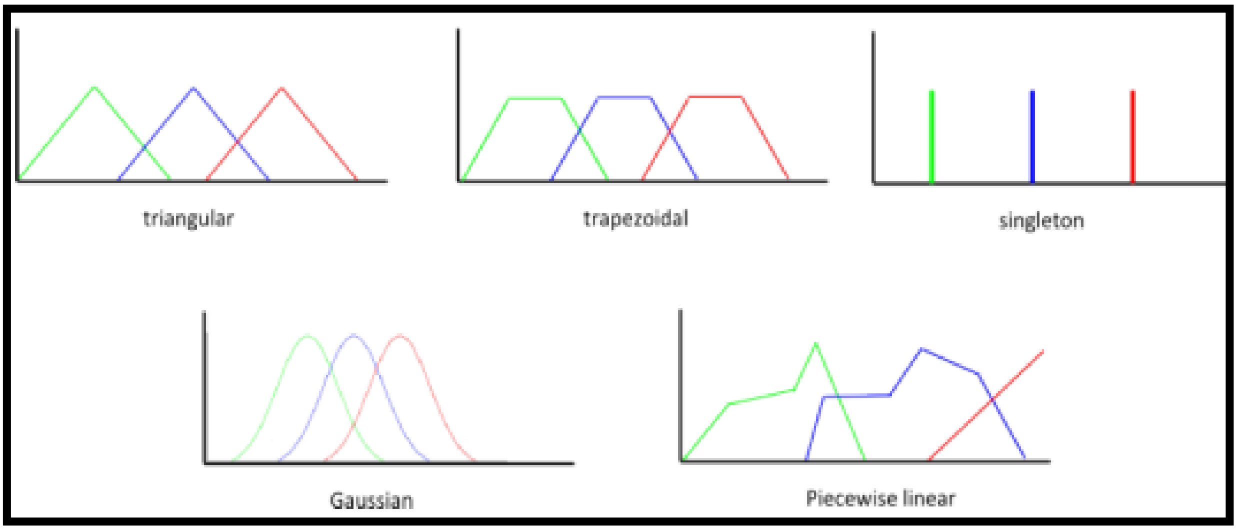


Figure 3: Illustrating Various Membership Functions in Fuzzy Logic

The use of membership functions implies that if a row of data represents , say a class of "safe from floods", then this class may be 80 percent in membership, Moderate floods, may be 15 percent and "risky floods area" be of 5 percent membership.

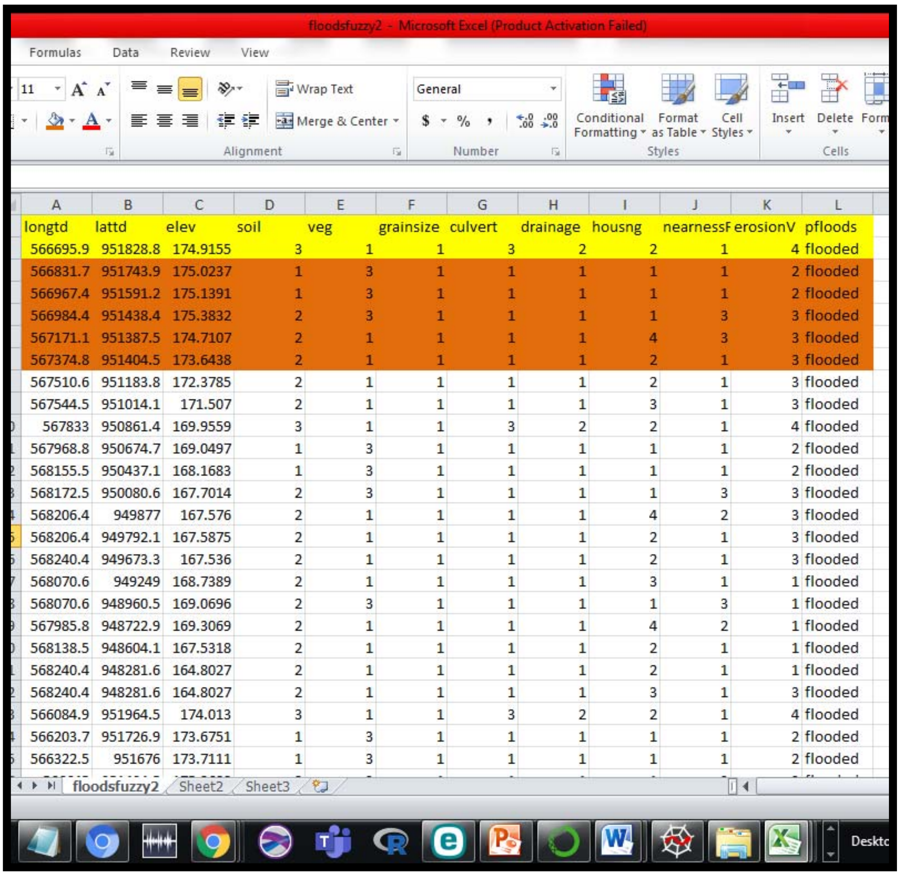
3. *Case of where there is sufficient data*

This is the scenario where fuzzy reasoning is again employed, combined with neural networks, so that the imprecisions that may be active in the datasets being used may be cancelled out during the process of defuzzification, and hence has the predictor model accuracy being much more improved. It is this scenario that was presented in the course of this research as the study team went out there and generated data, subsequently subjecting it to modeling and achieving the results deemed viable for both present Planning for and future mitigations of Floods Monitoring in the Madogo-Mororo areas of Tana River County.

4. *The Dataframe Used In the Study of Madogo Floods.*

The present study was facilitated by the dataset comprising 198 rows and 12 columns of datasets generated via the subjective coding of geomorphological, land use, soil mechanics, hydrogeological and geological variables prevalent in the area of study.

The following is an extract of the data in csv excel format as used in the study:



longtd	lattd	elev	soil	veg	grainsize	culvert	drainage	housng	nearnessF	erosionV	pfloods
566695.9	951828.8	174.9155	3	1	1	3	2	2	1	4	flooded
566831.7	951743.9	175.0237	1	3	1	1	1	1	1	2	flooded
566967.4	951591.2	175.1391	1	3	1	1	1	1	1	2	flooded
566984.4	951438.4	175.3832	2	3	1	1	1	1	3	3	flooded
567171.1	951387.5	174.7107	2	1	1	1	1	4	3	3	flooded
567374.8	951404.5	173.6438	2	1	1	1	1	2	1	3	flooded
567510.6	951183.8	172.3785	2	1	1	1	1	2	1	3	flooded
567544.5	951014.1	171.507	2	1	1	1	1	3	1	3	flooded
567833	950861.4	169.9559	3	1	1	3	2	2	1	4	flooded
567968.8	950674.7	169.0497	1	3	1	1	1	1	1	2	flooded
568155.5	950437.1	168.1683	1	3	1	1	1	1	1	2	flooded
568172.5	950080.6	167.7014	2	3	1	1	1	1	3	3	flooded
568206.4	949877	167.576	2	1	1	1	1	4	2	3	flooded
568206.4	949792.1	167.5875	2	1	1	1	1	2	1	3	flooded
568240.4	949673.3	167.536	2	1	1	1	1	2	1	3	flooded
568070.6	949249	168.7389	2	1	1	1	1	3	1	1	flooded
568070.6	948960.5	169.0696	2	3	1	1	1	1	3	1	flooded
567985.8	948722.9	169.3069	2	1	1	1	1	4	2	1	flooded
568138.5	948604.1	167.5318	2	1	1	1	1	2	1	1	flooded
568240.4	948281.6	164.8027	2	1	1	1	1	2	1	1	flooded
568240.4	948281.6	164.8027	2	1	1	1	1	3	1	3	flooded
566084.9	951964.5	174.013	3	1	1	3	2	2	1	4	flooded
566203.7	951726.9	173.6751	1	3	1	1	1	1	1	2	flooded
566322.5	951676	173.7111	1	3	1	1	1	1	1	2	flooded

Figure 4: The dataframe used in developing the predictive Model for floods in the Madogo-Mororo areas. The response column is pfloods in words like “flooded”.

One may note that the probability part which is in the column of pfloods is stated in words, as ‘flooded”. A different version was also employed, so simplicity, where the coding was as thus:

a) FLOODED=1

b) MODERATE FLOODS=2

c) FLOODS-FREE=3

The altered dataframe appears hereunder:

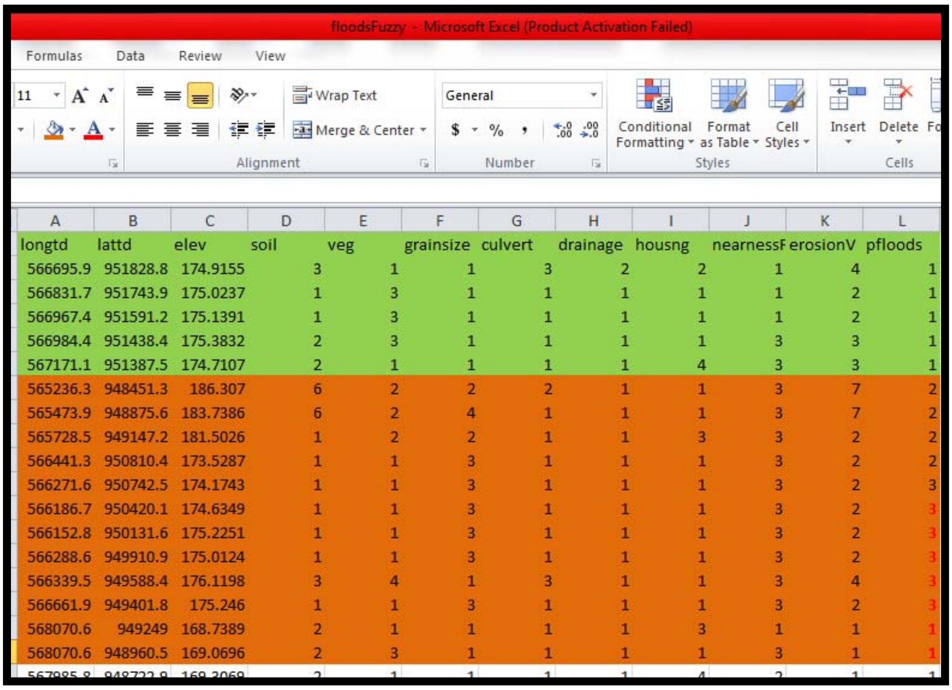


Figure 5: Dataframe used in developing the predictive Model for floods in the Madogo-Mororo areas. The response column is pfloods, coded in numerical values 1, 2 and 3.

The **R** software library of **frbs** was used to map flooding and generate prediction models using neuro-fuzzy inference systems. The display shown hereunder reflects part of the outputs thus generated.

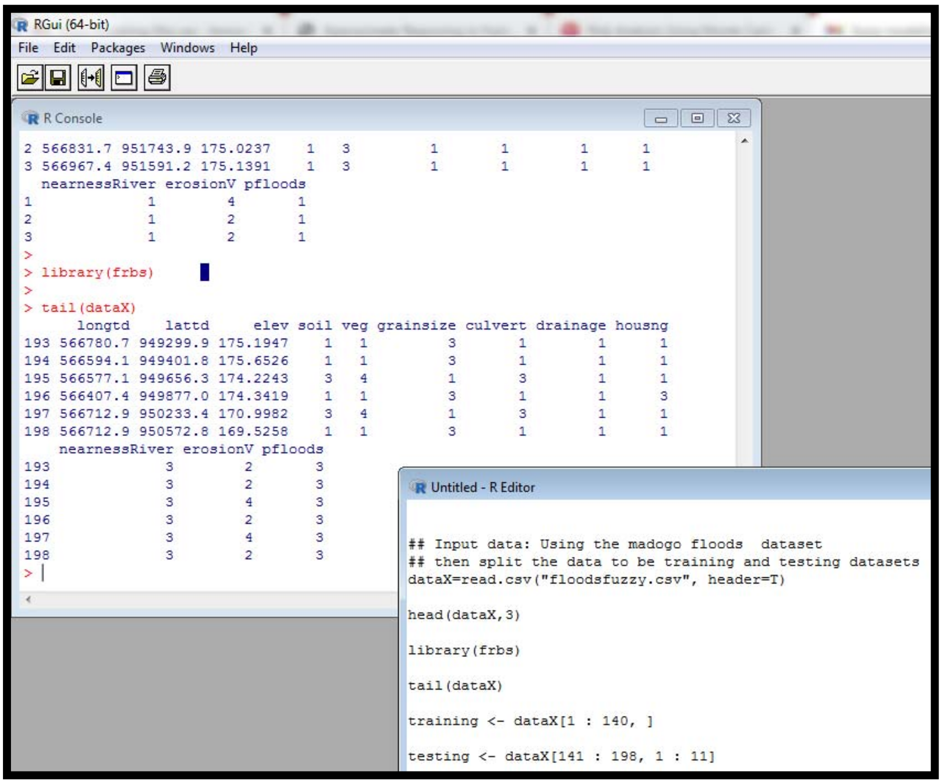


Figure 6: The fuzzy inference data for floods prediction as run on the R software GUI.

The fuzzy regression problem displayed here employed the Wang and Mendel algorithm and the R codes used for the task are shown overleaf:

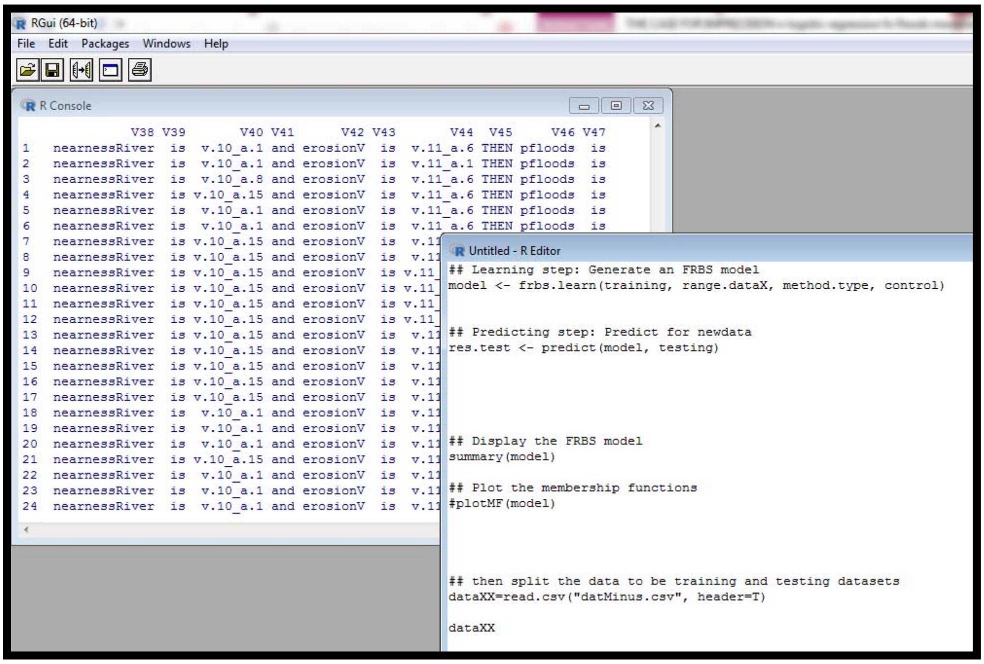


Figure 7: The fuzzy inference OUTPUT for floods prediction as run on the R software GUI. Note the Antecedents and Consequent Linguistic variables for soil erosion status and influence of nearness to seasonal and perennial rivers to floods probability in the study area.

5. Model generated for Floods prediction in study area

The model generated in r is as shown hereunder. It comprises the twelve variables, and is a product of training and testing undertaken on the original dataset.

Learning step: Generate an FRBS model

Model <- frbs.learn(training, range.dataX, method.type, control)

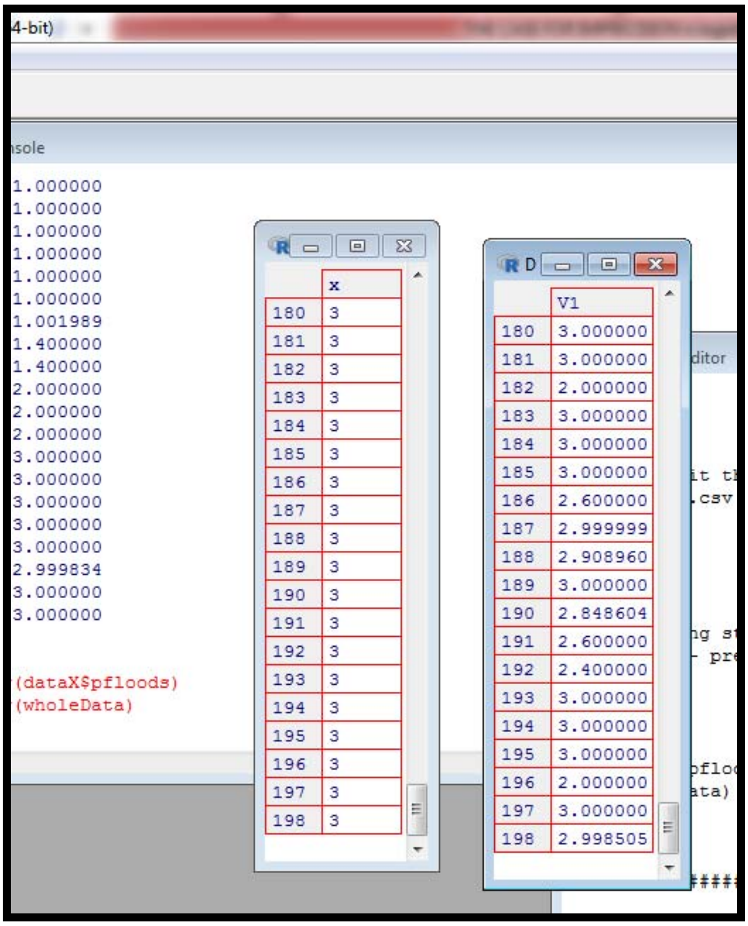


Figure 8: The comparison between predicted and the actual class with only three misclassifications out of 198. This is an accuracy level of 98.4 percent

There are three zones that came up during the study. This model may now be tested on field generated dataset. For example, there area twenty one rows of dataset, with ten for flooded areas, three taken from moderate risk areas and eight taken from areas in the zone deemed floods free. The table appearing in the page overleaf serves to illustrate the new field data table, whose areas are plots to be bought by developers.

The task is to predict the areas for flood risk. The table has no response column which is the probability column for floods risk. This is deliberately so as the model to be used has the column.

	A	B	C	D	E	F	G	H	I	J	K
1	longtd	lattd	elev	soil	veg	grainsize	culvert	drainage	housng	nearnessF	erosionV
2	568036.7	947382.1	167.9247	1	3	1	1	1	1	1	2
3	567476.6	947297.2	174.4414	1	3	1	1	1	1	1	2
4	567799.1	947772.4	168.9324	2	3	1	1	1	1	3	3
5	568087.6	948213.7	166.2457	2	1	1	2	1	4	2	3
6	567867	949707.2	168.5103	2	3	1	1	1	1	3	3
7	568070.6	950997.1	169.7277	2	1	1	1	1	4	2	3
8	567968.8	951387.5	171.0091	2	1	1	1	1	2	1	3
9	565117.5	950250.4	178.9574	6	2	2	1	1	1	3	3
10	565456.9	950488	177.3457	5	2	2	1	1	1	3	3
11	565847.3	950555.9	175.8854	6	2	2	2	1	1	3	3
12	566152.8	950674.7	174.635	6	2	2	1	1	1	3	3
13	566628	950657.7	171.6544	6	2	3	1	1	1	3	3
14	567052.3	950674.7	170.1373	6	2	2	1	1	1	3	3
15	567069.3	950199.4	168.4881	1	1	3	1	1	1	3	2
16	567222	949843	169.2986	1	1	3	1	1	1	3	2
17	566865.6	949639.4	172.4696	2	1	3	1	1	1	3	2
18	566729.8	949860	172.0003	1	1	3	1	1	1	3	2
19	566543.1	950131.6	172.6184	3	4	1	3	1	1	3	4
20	566543.1	950437.1	172.7178	1	1	3	1	1	1	3	2
21	566984.4	950301.3	168.8936	1	1	3	1	1	1	3	2
22	566967.4	950640.7	169.7687	1	1	3	1	1	1	3	2
23											
24											

Figure 9: The table with variables to be predicted for floods risk in study area. The first ten rows were taken from areas deemed flood-risk prone. The next three were taken from areas deemed to possess moderate risk. The final eight rows were areas generally considered risk-free.

The model generated the following output table:

```
RGui (64-bit)
File  Edit  Packages  Windows  Help
R Console
>
> newPred
          [,1]
 [1,] 1.000000
 [2,] 1.000000
 [3,] 1.000000
 [4,] 1.000000
 [5,] 1.000000
 [6,] 1.000000
 [7,] 1.000000
 [8,] 1.001989
 [9,] 1.400000
[10,] 1.400000
[11,] 2.000000
[12,] 2.000000
[13,] 2.000000
[14,] 3.000000
[15,] 3.000000
[16,] 3.000000
[17,] 3.000000
[18,] 3.000000
[19,] 2.999834
[20,] 3.000000
[21,] 3.000000
>
```

Figure 10: The predictions as captured in the R softwares GUI. All the classes expected have been correctly predicted by the Model.

6. *Results*

The neuro-fuzzy model predicts all the new data from the field correctly, assigning them the classes of flood propensity expected from the respective flood zoning classifications anticipated.

It is imperative that the performance of this algorithm be compared with that of another ML algorithm, and since the task is that of classification Logistic regression was considered appropriate. There are three classes and therefore the multinomial classifier was considered appropriate for the work.

V. Logistic Regression Models

A. Multinomial Dependent Variable

This is the algorithm employed for logistic regression when the task involves estimating more than just two classes' in predictions. In the case of the two class tasks, one may employ the binomial predictive models

Whenever the response variable has more than two classes, more than one regression equation will be needed to make predictions. Moreover this number of regression equations would be the equivalent of one less than the number of expected outcomes. It means that from the case where we are predicting three floods class of **flooded, moderate and floods-free,** the equations will be two only, derived from three minus one.

Ordinarily, the process creates more regression coefficients, for each independent variables used, making the process of interpreting and using the model even more difficult. In this case, care must be taken to understand what each regression equation is predicting. Once this is understood, interpretation of each of the respective regression coefficients for each variable can proceed. For the purpose of understanding this model, let it be supposed that floods probability classes 1, 2, and 3 are represented by A, B and C, respectively. Suppose the reference value is C. The two regression equations would be as thus:

$$\ln\left(\frac{p_A}{p_C}\right) = \beta_{A0} + \beta_{A1}X_1 + \beta_{A2}X_2$$

and

$$\ln\left(\frac{p_B}{p_C}\right) = \beta_{B0} + \beta_{B1}X_1 + \beta_{B2}X_2$$

1. Interpreting the above equations

From the above, the coefficients for X1 in the above equations, β A1 and β B1, displays the change in the log-odds of A versus C, and B versus C for a one unit change in X1, respectively.

The codes used in the model are shown hereunder:

```
# Partition Data
set.seed(121)
ind <- sample(2, nrow(dataX),
              replace = TRUE,
              prob = c(0.8, 0.2))
training <- dataX[ind==1,]
testing <- dataX[ind==2,
]

##
# Multinomial Logistic regression with First Two PCs
library(nnet)


training$pfloods <- relevel(training$pfloods, ref = "flooded")
mymodel <- multinom(pfloods~., data = training)
summary(mymodel)



# Confusion Matrix & Misclassification Error - training
p <- predict(mymodel, training)
```

Figure 11: R-Codes used in the model

The accuracy metrics for prediction of floods using the data on flooding in Madogo is hundred percent from the training data as shown in the tabulated confusing matrix shown hereunder:

```
> # Confusion Matrix & Misclassification Error - training
> p <- predict(mymodel, training)
> tab <- table(p, training$pfloods)
> tab

p             flooded floodFree moderateRisk
  flooded          49         0            0
  floodFree         0        60            0
  moderateRisk      0         0           52
>
```

Figure 12: Model Predictions with Floods Training dataset

The accuracy metrics for prediction of floods using the data on flooding in Madogo is 92 % from the testing data as shown in the tabulated confusing matrix shown hereunder:

```
1              flooded floodFree moderateRisk
 flooded            11         0            0
 floodFree           0        15            0
 moderateRisk        0         3            8
 1 - sum(diag(tab1))/sum(tab1)
1] 0.08108108
 1-0.08108108
1] 0.9189189
```

Figure 13: Model Predictions with Floods Testing dataset

The accuracy metrics for prediction of floods using the data on flooding in Madogo is 98.4% from the WHOLE data as shown in the tabulated confusing matrix shown hereunder:

```
3               flooded floodFree moderateRisk
 flooded             60         0            0
 floodFree            0        75            0
 moderateRisk         0         3           60
 1 - sum(diag(tab3))/sum(tab3)
1] 0.01515152

 # accuracy score using training data
 accuracy=sum(diag(tab3))/sum(tab3)
 print(accuracy)
1] 0.9848485
```

Figure 14: Model Predictions with the whole floods dataset

Only three cases of mis-classifications arose using the whole dataset, just like was the case with the neuro-fuzzy regression model predictions. Both Logistic regression and neuro-fuzzy models may be used in mapping and predicting flood occurrences on Madogo –Mororo areas.

GIS Model 1-Floods –this is the flood-risk map of study area. The eastern portion is inherently at risk owing to the numerous lagha units-these are seasonal streams whose flow is ephemeral. The rivers drain southeast wards towards the perennially-flowing river Tana. All the settlements encountered on the course of this flow are at an inherent risk of flooding. The values of 1 represent flood risk. A Value of 2 represents moderate risks to floods. The value of 3 represents floods free area.

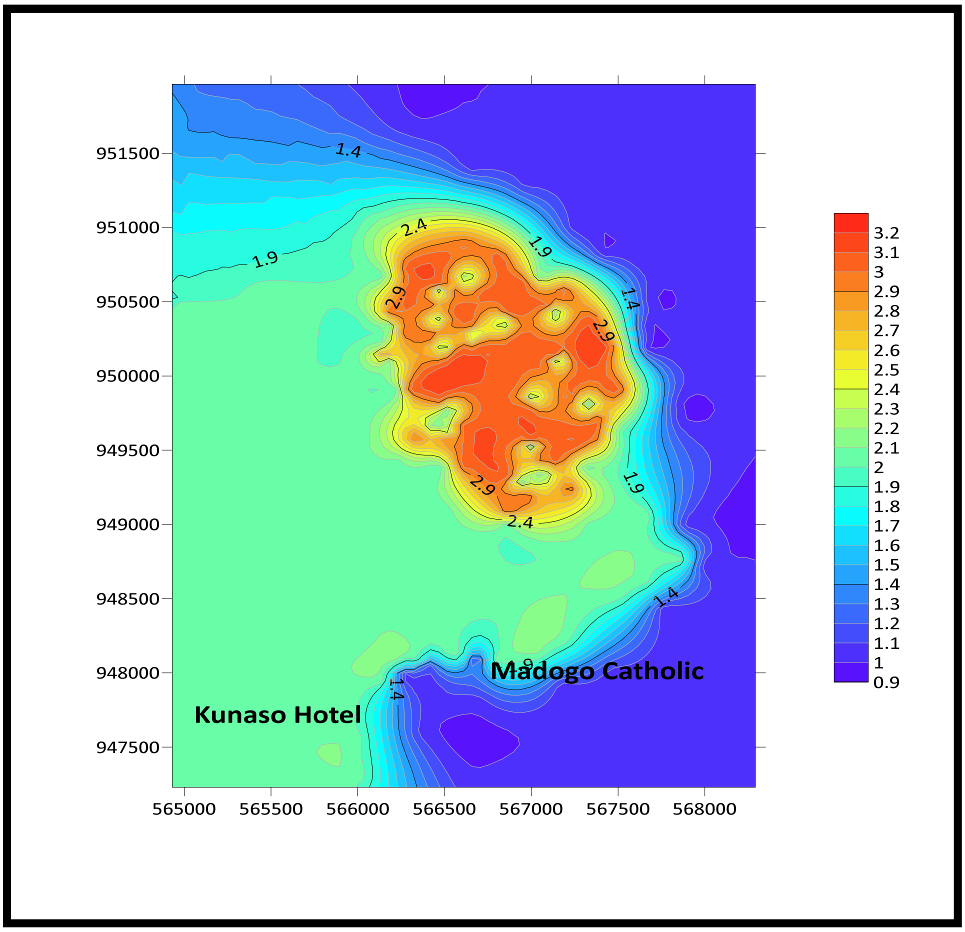


Figure 15: Model Showing the Flood Risk Areas With respect to Topography.

VI. ACKNOWLEDGMENT

I wish to acknowledge the Ministry of Water, Sanitation and Irrigation of the Republic of Kenya for giving its support in conducting the field investigations and surveys in Mororo and Madogo. In addition, my sincere gratitude goes to the Water Resources Authority, Lower Tana Sub-Catchment for providing the Secondary data used in the research project. I also wish to acknowledge the support of my Research assistant, **Mr Jibreel Shune** for assistance with field work in mapping and coding soils and the river details.

VII. RECOMMENDATIONS AND CONCLUSIONS

The study vindicates the use of both neuro-fuzzy inferences and logistic regression, a statistical technique in decision making on matters town planning, so that persons may not be allocated land for building residential plots for areas predicted to be highly susceptible to flooding in Madogo area. This way, people will not in future lose lives unnecessarily to poor decisions. Also, it is time for the county government to relocate all persons residing very close to seasonal streams that are usually very prone to floods and also the ones living within the River Tana riparian corridors. The farming and grazing within the catchment exacerbate the siltation and erosion activities, which end up being deposits in the river. The reservoir volume of the riverbed has progressively diminished to the point whereby whenever it rains; the inflow cannot be transported downstream, on this account. It thus occasions a backflow which floods the surrounding houses and infrastructures. The longitudes and latitudes of the places where people wish to settle at, as well as the soil characteristics and nearness to the river beds, are such a critical factor in helping save lives, using the models developed. The information may be used as input for the model to predict flooding risks thereof.

REFERENCES

[1] Acreman, M., & Holden, J. (2013). How wetlands affect floods. *Wetlands*, *33*(5), 773-786.

[2] Andrews, P., Lord, J. M., & EVANS, E. M. N. (1979). Patterns of ecological diversity in fossil and modern mammalian faunas. *Biological Journal of the Linnean Society*, *11*(2), 177-205.

[3] Jia, X., Morel, G., Martell-Flore, H., Hissel, F., & Batoz, J. L. (2016). Fuzzy logic based decision support for mass evacuations of cities prone to coastal or river floods. *Environmental modelling & software*, *85*, 1-10.

[4] 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.

[5] Mousavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 1536.

[6] Reisenbüchler, M., Bui, M. D., Skublics, D., & Rutschmann, P. (2019). An integrated approach for investigating the correlation between floods and river morphology: A case study of the Saalach River, Germany. *Science of the Total Environment*, *647*, 814-826.

[7] Sivakumar, S. S. (2015). Flood mitigation strategies adopted in Sri Lanka a review. *International Journal of Scientific and Engineering Research*, *3*(6).

[8] Zhong, M., Wang, J., Gao, L., Lin, K., & Hong, Y. (2019). Fuzzy risk assessment of flash floods using a cloud-based information diffusion approach. *Water Resources Management*, *33*(7), 2537-2553.