Enhancing Food Security in Africa with a Predictive Early Warning System on Extreme Weather Phenomena

Alvin M Igobwa
iLab
Jeremy Gachanja
iLab
Betsy Muriithi
iLab
John Olukuru
iLab
Angeline Rehema Wairegi (awairegi@strathmore.edu)
Strathmore University https://orcid.org/0000-0002-7408-3502
Isaac Rutenberg
CIPIT

Research Article

Keywords: Climate Change, Food Security, Extreme Weather Prediction, Agricultural Insurance, Insurance Based Index

Posted Date: March 2nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1343486/v1

Enhancing Food Security in Africa with a Predictive Early Warning System on Extreme Weather Phenomena

Person A[1], Person B[1], Person C[1], Person D[1], Person E*[2], Person F[2]
Institution One [1]; Institution Two[2]
* Corresponding Author
person.a@institution.edu[1], person.b@institution.edu[1], person.c@institution.edu[1], person.d@institution.edu[1], person.e@institution.edu[2], person.f@institution.edu[2].

Enhancing Food Security in Africa with a Predictive Early Warning System on Extreme Weather Phenomena

Abstract

Climate change is predicted to exacerbate Africa's, already, precarious food security. Climate models, by accurately forecasting future weather events, can be a critical tool in developing countermeasures to reduce crop loss, decrease adverse effects on animal husbandry and fishing, and even help insurance companies determine risk for agricultural insurance policies – a measure of risk reduction in the agricultural sector that is gaining prominence. In this paper, we investigate the efficacy of various open-source climate change models and weather datasets in predicting drought and flood weather patterns in northern and western Kenya and discuss practical applications of these tools in the country's agricultural insurance sector. We identified two models that may be used to predict flood and drought events in these regions. The combination of Artificial Neural Networks (ANNs) and weather station data was the most effective in predicting future drought occurrences in Turkana and Wajir with accuracies ranging from 78% to 90%. In the case of flood forecasting, Isolation Forests models using weather station data had the best overall performance. The above models and datasets may form the basis of a more objective and accurate underwriting process for agricultural index-based insurance, as we expound in the paper.

Keywords: Climate Change, Food Security, Extreme Weather Prediction, Agricultural Insurance, Insurance Based Index

1. Introduction

Agriculture is the most important economic activity in Africa, accounting for 30% Gross Domestic Product (GDP) across the continent (23% of Sub – Saharan Africa's GDP) and employing 60% - 65% of the continent's workforce (Goedde, Ooko - Ombaka and Pais 2019). While agricultural production has increased steadily in Africa since the 1970s, there has been no significant improvements in the production factors, i.e., labor and land; consequently, the sector's growth has been insufficient to adequately address poverty, attain food security, and lead to sustained GDP growth on the continent (Bjornlund, Bjornlund and Van Rooyen 2018, Food and Agriculture Organization (FAO) 2009, Pfister, et al. 2011). Climate change and a population boom have further contributed to food insecurity.

The impact of climate change in Africa is predicted to be greater than anywhere else on the globe. Sub-Saharan Africa is especially vulnerable to climate change due to a confluence of biophysical, political, and socioeconomic stresses that interact to increase the region's susceptibility and constrain its adaptive capacity. Climate change in Sub-Saharan Africa is expected to manifest in an increase in temperatures (an average increase of 3–4 °C over the next century, greater than the projected global annual mean), changes in rainfall intensity, increase in extreme weather phenomena such as droughts and flood, increase in desertification, and changes in the spatial and temporal transmission of infectious diseases – events that are likely to cause disruptions in growing seasons, reductions in the area suitable for agriculture, and declines in agricultural yield (Boko, et al. 2007, Chen, et al. 2006, Muller, et al. 2011, Niang, et al. 2014, Reich, et al. 2001, Richard, et al. 2001, Sarr 2012, Songok, Kipkorir and Mugalavai 2011, Thomas, et al. 2007). A majority of its population is still dependent upon these local natural resources for their livelihood; engaging in activities such as agriculture, pastoralism, and fishing for sustenance and revenue (Osbahr, et al. 2008, Wlokas 2008). Climate change (experienced over longer time frames as changes in climatic norms and over shorter periods as changes in the frequency and severity of extreme weather events) impedes farmers', pastoralists', and agro-pastoralists' ability to grow crops and rear livestock, and, in general, has a negative impact on the livelihoods of communities, especially those in rural areas (Connolly-Boutin and Smit 2016, Kebede, Hasen and Negatu 2011, Songok, Kipkorir, et al. 2011, Thompson and Scoones 2009). Only 5% of farmlands in Africa are irrigated compared to 37% in Asia and 14% in Latin America, making them more dependent on natural precipitation (Ringler, et al. 2010).

Solutions to food insecurity have, in the past, largely focused on agricultural intensification as a mechanism for greater crop production; farming diversification and the utilization of emerging technologies to increase crop yield or protect crop loss have been less utilized (Hazell and Wood 2008). In recent years, however, stakeholders, in both the public and private sectors, have sought more innovative approaches - including the use of AI and big data - to protect and improve crop yield. Artificial Intelligence (AI) functions: spatial planning, soil and weather analysis, plant disease and pest detection, precision application of herbicides and pesticides, seasonal forecasting, etc., have the potential to increase agricultural productivity across the continent. A number of these measures are already in use on the continent. Agricultural insurance, used to protect farmers from

risks posed by extreme weather phenomena, is another tool that has gained increased usage across the globe.

The ability to predict climatic changes using climate models is a critical tool that can be used to develop countermeasures to reduce crop loss, decrease adverse effects on animal husbandry and fishing, and even help insurance companies determine risk for agricultural insurance policies. In this paper, we investigate the efficacy of various open-source climate change models and weather datasets in predicting drought and flood weather patterns in northern and western Kenya. In Section 1, we give a brief overview of climate change models for drought and flood prediction and the weather phenomena experienced in the regions studied; Section 2 and 3 outline the methodology and results of the study; in Section 4, we discuss how the model(s) and dataset(s) determined in Section 3 may be deployed in risk assessment for agricultural insurance, and finally in Section 5, we detail conclusions of the study.

1.1. Extreme Weather in Kenya.

Kenya's reliance on agriculture and other rain-fed activities to sustain its economy makes it particularly vulnerable to climate extremes. For example, recurring floods in areas along rivers flowing from Lake Victoria: Budalang'i in Busia County, Kano Plains in Kisumu County, and the lower parts of the Tana River area, have led to humanitarian and fiscal disasters (Opere 2013). The variability of extreme rainfall in the country has led to socio-economic challenges in urban and rural settings, including damage to infrastructure, loss of agricultural yields, and adverse effects on human health, e.g., the prevalence of water-related diseases (Juma, et al. 2020). Kenya has also experienced a notable increase in severe and frequent drought events in recent decades. Drought events are predominantly experienced in the eastern and north-eastern regions of the country, as well as parts of the coast and Rift Valley - arid and semi-arid lands (ASAL) that form about 80% of Kenya's land cover (Mutsotso, Sichangi and Makokha 2018, Owuor 2015). The occurrence of drought events in these regions has been exacerbated by the continuous decline in March-May (MAM) seasonal rainfall (Ayugi, Tan and Niu, et al. 2020b). The communities in these areas, mainly pastoralists and agro-pastoralists, are made vulnerable by the adverse effects of the droughts, which include acute malnutrition, famine, increased conflict, interrupted food chains, and economic losses due to poor meat and milk production (Owuor 2015)

Research shows that over the past decade, developing countries, on the whole, have incurred over 35 billion USD a year in damages from natural disasters, 20 times the cost sustained in the developed world (Mirza 2003). There is growing concern regarding the increase in frequency and magnitude of these extreme weather events. These losses are expected to increase with time. These projections call for continuous assessment and monitoring of the evolving extreme weather patterns in the country, to develop effective adaptation measures to reduce risk and ensure that the region is prepared when unforeseen weather events occur (Kundzewicz, et al. 2014).

1.2. Climate Change Modelling for Drought and Flooding Weather Patterns

Predicting floods and droughts is important in developing mitigation strategies to manage their impact. The planning of emergency responses and early warning systems rely on flood and drought forecasts at hourly to seasonal time scales (Mizra and Ahmed 2010). Climate change models simulate the future impact of greenhouse gases based on the current understanding of atmospheric

physics and chemistry (Hannah 2015). Climate modelling of floods and droughts are typically done separately, primarily because the two types of weather extremes evolve from different processes. Floods are fast phenomena triggered by excess precipitation, snowmelt, high initial soil moisture, or a combination of these factors, while droughts occur slowly triggered by a precipitation deficit and modulated by positive temperature anomalies (Berghuijs, et al. 2016, Brunner, et al. 2021, Hanel, et al. 2018, Wasko and Nathan 2019, Woodhouse, et al. 2016). Two techniques are commonly employed to predict flood and drought weather patterns: driving a hydrological model with meteorological data from a global/regional circulation model (GCM or RCM) or employing hybrid statistical-dynamical techniques using projections as covariates within a statistical modeling framework (Addor, et al. 2014, Hakala, et al. 2019, Madadgar, et al. 2016, Mendoza, et al. 2015, Mizukami, et al. 2016, Slater and Villarini 2018, Wilby 2010).

Unfortunately, GCMs often have coarse space resolution and cannot represent the fine-scale detail that characterizes the climate in many regions of the world; RCMs, meanwhile, are plagued with uncertainties from boundary conditions, the size of the integration domain, and natural variability within the RCMs and RCM formulation (Luhunga, Botai and Kahimba 2016, Meier, et al. 2011, Min, et al. 2013). In recent years, anomaly detection methods (techniques for identifying unexpected events or patterns that differ from the norm in a given set) have been employed to identify extreme climate events and climate change including floods (Çelik, Dadaser-Çelik and Dokuz 2011, Chalapathy and Chawla 2019, Wibisono, et al. 2021). Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, nearest neighbourhood-based techniques, and deep learning methods are examples of anomaly detection methods utilized for this purpose. DBSCAN and other clustering-based algorithms identify anomalies as data instances that do not belong to any cluster. Climate modeling studies utilizing the DBSCAN algorithm have shown that this approach is superior to traditional statistical modeling approaches - that the algorithm can identify both extreme and rare anomalies, and it can handle more complex data handling needed for multivariate outlier detection (Çelik, Dadaser-Çelik and Dokuz 2011, Wibisono, et al. 2021). It should be noted that if anomalies were to form clusters, due to the inherent nature of the DBSCAN algorithm, it would be unable to detect them (Chandola, Banerjee and Kumar 2009). Nearest neighborhood-based techniques identify anomalies as instances that occur furthest from their closest neighbors. Studies have shown that the temporal and spatial analysis of anomalies using a k^{th} nearest neighbour (KNN) technique in global climate data could also explain extreme events like drought and flood events at specific locations on earth (Das and Parthasarathy 2009). Compared to clustering-based techniques, nearest neighbourhood-based techniques are better optimised to find anomalies since each instance is evaluated with respect to its local neighbourhood rather than cluster membership (Chandola, Banerjee and Kumar 2009). Finally, deep learning techniques have been used to study anomalies in rich climate data, and are in fact, more adept than traditional statistical methods in dealing with the complex data structures (as they can identify obscure and complex correlations in data sets for which more traditional analysis methods have not determined the underlying model) that characterise climate data, large data volumes and automate feature learning for anomaly prediction (Chalapathy and Chawla 2019).

Statistical methods have traditionally been used for drought forecasting. Regression, time series analysis, and probability models dominated early attempts in drought forecasting. However, these linear approaches fail to adequately capture the white noise, non-stationary and non-linearity in the time series and thus are gradually being replaced by Artificial Intelligence (AI) models (Fung, et al. 2020). Artificial Neural Networks (ANNs) stand out among the AI models as versatile and well-established drought forecasting tools (Inoubli, et al. 2020). ANNs are flexible, non-linear mathematical modelling systems composed of inter-connected neuron units that mimic the biological nerve system, designed with efficient parallel processing. ANNs learn from the relationships between input parameters and controlled or uncontrolled variables by checking previous trends in historical data as nonlinear regression; the main factors affecting ANNs performance are input adequacy, network architecture and model validation (Deo and Sahin 2015, Fung, et al. 2020).

2. Methods

2.1. Areas of Study

The study focused on three regions in Kenya: Ahero town in Kisumu County, for the flood analysis, and Turkana and Wajir, located in Kenya's most arid area, for the drought analysis as summarized in Fig. 1 The image on the left shows Ahero town, the area of focus for the flood detection study and the image on the right shows Turkana and Wajir counties in Northern Kenya, the areas of interest for the drought study. Ahero is located on the banks of River Nyando. It is commonly known for having low mean rainfall, approximately 7.72mm. However, sporadic high torrential rainfall sometimes leads to pluvial floods in the area. Ahero is among the wards identified in Kenya with high exposure and vulnerability to flooding hazards (Makena, et al. 2021). Turkana is located in the Northern region of Kenya. It has high-temperature levels that range between 36°C and 38°C. Wajir is located in the North-Eastern part of Kenya, where rainfall is scarce and there is low vegetation cover. According to the country's National Drought Management Authority, both of these areas are prone to recurrent drought events (National Drought Management Authority 2020, National Management Authority 2020)

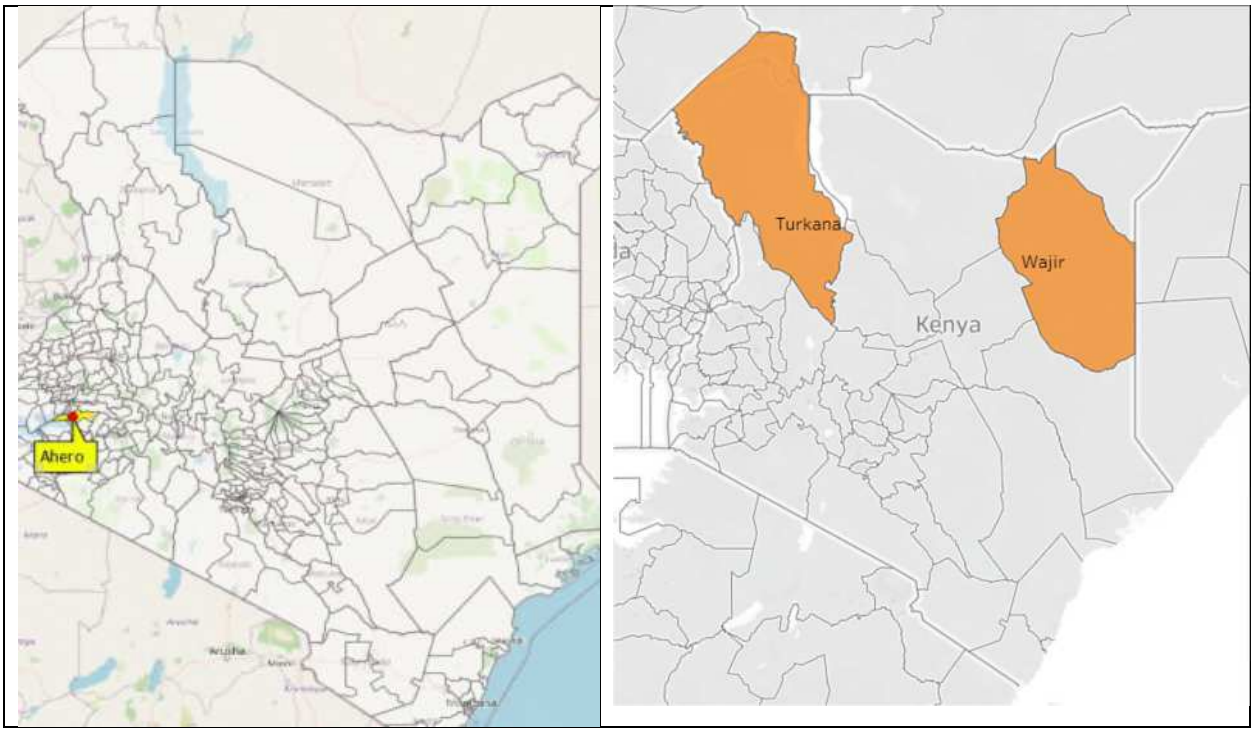


Fig. 1 The image on the left shows Ahero town, the area of focus for the flood detection study and the image on the right shows Turkana and Wajir counties in Northern Kenya, the areas of interest for the drought study

2.2. Datasets

Conventionally, three types of datasets are utilized in climate modelling studies: observational products (satellite or gridded weather station), reanalysis products (obtained by combining observational inputs and climate models based on physical principles), and simulation products (climate models). This study used observational and climate datasets (the most commonly used datasets in data-sparse regions). Monthly precipitation data was obtained for both types of datasets for the period between 1992-2020.

2.2.1. Climate Model Data

Climatic data products, generally either GCM or RCM, are used in regions where accessing accurate and complete meteorological predictors at a high temporal resolution is difficult. At the regional level, however, GCMs fail to capture complex local processes and land-surface heterogeneities, which demand spatially detailed information at much finer scales than currently affordable with GCMs. Regional Climate Models (RCMs) are often used instead. RCMs are downscaled GCMs to a finer resolution (50km) and consider mesoscale factors such as complex topography, coastlines, inland bodies of water, and land cover distribution over a limited region of interest at a computationally affordable cost (Luhunga, Botai and Kahimba 2016, Meier, et al. 2011). Several RCMs of the Coordinated Regional Downscaling Experiment (CORDEX) Africa have been developed and used for analyses of projections of climatic conditions over East Africa. Three Regional Circulation Models (RCM) were selected from CORDEX (Coordinated Regional Downscaling Experiment) for the study:

- Max Planck Institute for Metrology Germany, MPI-M-MPI-ESM-LR (MPI)
- Canadian Centre for Climate Modelling and Analysis, CCCma-CanESM2 (CCCma)
- Met Office Hadley Centre, MOHC-HadGEM2-ES (Had-GEM2)

The three RCMs were found in previous studies to reproduce the rainfall climatology over the study domain with reasonable accuracy (Ayugi, Tan and Gnitou, et al. 2020, Luhunga, Botai and Kahimba 2016). The three datasets were applied to the flood detection model, while the drought prediction models used the MPI and Had-GEM2.

2.2.2. Satellite Data

Two satellite datasets were utilized: CHIRPS (Climate Hazards Group Infrared Precipitation with Station data) and TAMSAT (Tropical Applications of Meteorology using Satellite data and ground-based observations). CHIRPS is a quasi-global rainfall dataset that incorporates a 0.050 (4.4km) resolution. It is mostly used in drought and flood monitoring. The University of Reading established TAMSAT to enhance the use of satellite-based rainfall estimates in Africa (Maidment, Black and Young 2017). The satellite uses a 4km resolution in the capturing of the rainfall data. Both satellite datasets have data spanning from 1992 to 2020 and are strictly for rainfall and therefore only utilized by the flood detection model.

2.2.3. Weather Station Data

Weather station data provides inferences for a single point compared to satellite or climate model datasets which capture a much larger area (TAMSAT and CHIRPS 4km resolution and RCMs 50km resolution). Weather station data was obtained from the Visual Crossing Weather Platform (https://www.visualcrossing.com), which provides daily historical forecasted data from any geographic location. The platform builds its historical weather data by combining raw surface observational data from various meteorological observational datasets such as the Integrated Surface Database (ISD) and the Meteorological Assimilation Data Ingest System (MADIS). The data acquired for this study spans the period from 1992 to 2020. Two additional datasets were also used: Trans-African Hydro-Meteorological Observatory (TAHMO), which was used to confirm the data bias of the visual crossing data, and World Weather Online data, which provides intra-day periods of data (3-hour data up to 1-hour data) for finer forecasts of the data. Weather station data have both temperature and rainfall data and are therefore applicable to the drought and flood models.

2.3. Anomaly Detection Techniques and Performance Evaluation for Flood Detection

Four unsupervised anomaly detection techniques were utilized in the flood study: K-Nearest Neighbours (KNN), Histogram-Based Outlier Score (HBOS), Cluster-Based Outlier Factor (CBLOF), and Isolation Forests.

K-Nearest Neighbours (KNN*)* is one of the simplest anomaly detection methods. It is a supervised machine learning algorithm that takes an unsupervised approach to anomaly detection. This density-based measure assumes that normal data points occur around a dense neighbourhood and anomalies lie far away. The algorithm performs a nearest neighbour search by computing distances of every two data points then compares them to an arbitrary threshold value beyond which observations are identified as anomalies (Chandola, Banerjee and Kumar 2009). The main

advantage of the KNN technique is that it is straightforward to adapt to different data types and only requires the definition of K and an appropriate distance measure for the given data. The study utilised a contamination rate of 10% with a K threshold of 5 and the Minkowski distance measure, a generalisation of the Euclidean and Manhattan distance measures.

Histogram-Based Outlier Score (HBOS) is a statistical-based anomaly detection technique (Goldstein and Dengel 2012). HBOS assumes that the data set features are independent and models the feature densities using normalised histograms (maximum height = 1) with a static or dynamic bin width. The height of every single bin of the histogram represents the density estimation. The outlier score for every data point is a multiplication of the inverse of the estimated densities, defined as:

$$HBOS(x) = \sum_{i=0}^{d} \log(\frac{1}{hist_i(x)})$$

Equation 1: Density Estimation represented by a single bin in Histogram Based Outlier Score (HBOS) calculation

where: d represents the number of features in the data set, $hist_i(v)$ is the density estimation of each feature instance, and x is the vector of features. Higher scores represent anomalies that would intuitively be assigned to bins with low density. This technique is fast compared to other techniques. For instance, it has a linear time complexity $\mathcal{O}(n)$ compared to KNN's $\mathcal{O}(n^2)$. HBOS is also a suitable option for treating global anomalies. However, it is a poor choice for local anomalies.

Cluster-based Local Outlier Factor (CBLOF) is an anomaly detection technique that combines a nearest neighbour-based technique, Local Outlier Factor (LOF), and clustering for data pre-processing and anomaly identification (He, Xu and Deng 2003). The algorithm assigns each data point a CBLOF, which is determined by the distance of the data point to its nearest neighbour and the size of its cluster. Anomalies are identified as instances that belong to small or sparse clusters. CBLOF was designed to address shortcomings in clustering-based methods such as DBSCAN and nearest neighbour-based approaches such as Local Outlier Factor that were deemed ineffective. The technique is better at capturing local anomalies than HBOS. Additionally, its computational cost is minimised by distributing large data sets into meaningful clusters. To calculate the CBLOF, k-means is used to cluster the dataset. Then a heuristic procedure is applied to split the clusters into two categories based on their density. The anomaly score is defined as:

$$CBLOF(x) = \begin{cases} |C_i|^* \min\left(distance\left(x, C_j\right)\right), if\ x \in C_i, C_i \in SC\ and\ C_j \in LC\ for\ j = 1\ to\ b \\ |C_i|^* \min\left(distance(x, C_i)\right), if\ x \in C_i and\ C_i \in LC \end{cases}$$

Equation 2: Anomaly score in Cluster-based Local Outlier Factor (CBLOF) calculations.

where: C_i and C_j represent clusters, LC = $C_i \mid i \leq b$ represents large clusters, SC = $C_j \mid j > b$ represents small clusters, and b is the cluster boundary (small or large).

Isolation Forests is an anomaly detection method for continuous data (Liu, Ting and Zhou 2012). The isolation forests technique makes two assumptions about anomaly data: anomalies vary

significantly from normal observations making it easy to isolate, and anomaly data are rare in the dataset. Given these assumptions, isolation of anomalies is implemented using an isolation tree (iTree), which is a binary tree structure that considers whether an observation is an anomaly or not. The algorithm recursively portions a dataset to build an ensemble of iTrees. The anomaly score is derived from the path length, averaged over the isolation trees' ensemble. Anomalies are identified as instances with short average path lengths on the iTrees. Isolation Forests have several advantages over other anomaly detection techniques. First, given its base assumptions, the algorithm can exploit subsampling, making it fast and scalable. Second, Isolation Forests can detect both clustered and scatted anomalies in the global context of the entire dataset. Isolation Forests requires two training parameters: the number of trees to build and sub-sampling size and one evaluation parameter: the tree height limit during evaluation. The anomaly score for every data point x is defined as:

$$s(x,n) = 2^{\frac{E(h(x))}{c(n)}}$$

Equation 3: Anomaly score for individual data points in the Isolation Forest calculation.

where: $E\big(h(x)\big) = \frac{\sum_{i=1}^{t} h_i(x)}{t}$ is the average path length of x over t iTrees, $c(n) = 2H(n-1) - \left(\frac{2(n-1)}{n}\right)$ with $H(i) = \ln(i) + \gamma$ (γ is $Euler's contant$), the average path length of unsuccessful search in Binary Search Tree (BST). If $s(x,n)$ of x is close to 1, x is considered an anomaly or if less than 0.5, x is considered normal. The isolation forest contamination rate for the study was set to 10% such that the confidence interval was at 90% for the flood detection rate to ensure high precision for anomaly identification.

Model performance was done by comparing model data with the Standardised Precipitation Index (SPI). SPI quantifies precipitation anomalies for long-term normal conditions on multiple time scales (McKee, Doesken and Kleist 1993). Computation of the SPI involves fitting a probability distribution to an aggregated monthly precipitation time scale that may range from 3-48 months. The probability function is then transformed into a normal standardised index that is traditionally classified into flood classes that characterise the flood severity at each place and time scale. This study computed the SPI by fitting a logistic distribution for the different precipitation time series data. The data was converted to monthly intervals to maintain low granularity. SPI values greater than two were treated as the true observations of the occurrence of a flood. Using the SPI values to label the occurrence of flood events, the study measured how well the anomaly detection models predicted the actual class of the data point using the following metrics:

- Fraction Accuracy, which measured the fraction of the flood occurrences discovered by the outlier detection models over the number of floods the SPI found.
- Specificity, which measured how well the model was able to determine when floods would not occur.

$$\frac{TN}{TN + FP}$$

- Sensitivity, which measured how well the model detected flood occurrences.

$$\frac{TP}{TP + FN}$$

Note: TP is defined as the true positives from the model, TN is the model's true negatives, FP the false positives and FN the false negatives from the model.

- AUC-ROC, which measured the ability of the anomaly detection methods to distinguish between flood and normal observations.

2.4. ANN Model Development and Performance Evaluation for Drought Prediction

The ANNs were developed using the output from the Standardised Precipitation Evapotranspiration Index (SPEI) values for 28 years (1992-2020). SPEI is computed using precipitation and the potential evapotranspiration (PET) data to define anomalous wet and dry conditions by normalising the monthly (or weekly) difference between water supply (precipitation) and demand (potential evapotranspiration). This study calculated the PET following the Hargreaves equation (Hargreaves and Samani 1985). The Hargreaves equation uses the daily difference between the maximum and minimum as a proxy to estimate net radiation and simplifies the mass transfer term with a constant. Additionally, the study found the Hargreaves equation to be more suited to monthly durations.

$$PET = 0.0023 \times R_A \times \mathrm{TD}^{0.5}(TC^{\circ} + 17.8)$$

where: R_A measures the extra-terrestrial radiation calculated by knowing the station latitude, TD is the mean maximum minus the mean minimum temperature, TC is the mean temperature in degrees Celsius. The accumulated water profit or loss series, D_n^k, is computed using the following formula:

$$D_n^k = \sum_{i=0}^{k-1} P_{n-i} - PET_{n-i}, \qquad n \geq k$$

where: P_i is the monthly precipitation, PET_i is the monthly potential evapotranspiration, k is the time scale in months, and n is the calculation frequency. The D series was standardised using the log-logistic distribution based on the behaviour of the extreme values (Vicente-Serrano, Beguería and López-Moreno 2010). A log transformation was performed to normalise the input values and control for extreme values in the data. The study converted the SPEI time series into binary data indicating whether a given period experienced drought or normal conditions.

We utilized a multi-layer perceptron (MLP), neural network model, along with a user-created neural network. The selected ANNs used a two-layer neural network with four input layers, eight hidden layers and one output layer. The models used feed-forward backpropagation (BPN) as the training algorithm. The BPN utilised a sigmoid activation function since the output is of a binary form. A learning rate of 0.0001 was used to maximise the creation of the best log loss slope – for all the models, the same parameters were put in place to see how they all differ.

The data was split into a train-test and validation set up for each data set, where 90% of the data was used for training and testing (80:20). The remaining 10% was used in the validation. The training data consisted of the minimum and maximum temperature and precipitation variables. The

SPEI value grading for drought analysis was capped at < -1.5 to signal the start of drought occurrences.

The performance of the ANN models in predicting the monthly SPEI was statistically evaluated using the accuracy score and the AUC-ROC.

- The accuracy score measures the fraction of correct predictions out of the total number of observations.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

- The AUC-ROC measures the models' capacity to distinguish between drought and normal observations.

3. Results

3.1. Flood Detection in Ahero County in Western Kenya

Four anomaly detection models (CBLOF, HBOD, KNN, IF) and six datasets (CHIRPS, MPI, Had-GEM2, TAMSAT, CCCma, Weather Station Data) were analysed to determine their accuracy in detecting floods in Ahero county. Table 1 summarises descriptive statistics of precipitation data obtained from the selected open-source observational and simulation datasets. The datasets varied greatly in terms of precipitation distribution. TAMSAT data recorded the highest mean precipitation of 3.84mm (±.79), while the weather station data had the highest maximum precipitation recorded over the study period, 236mm. The RCMs recorded the lowest precipitation over the period compared to the observational products. The Met Office Hadley Centre Regional Climate Model (Had-GEM2) had the lowest average rainfall of 0.57 mm (±2.43). The differences between observed and simulated annual rainfall have been attributed to the poor skill of RCMs in reproducing the regional climate over East Africa (Ayugi, Tan and Gnitou, et al. 2020, Bichet, et al. 2020, Mumo and Yu 2020).

Table 1: A comparison of precipitation data for Ahero county across the 6 different datasets analyzed.

	MPI	**CCCMA**	**HAD-GEM2**	**CHIRPS**	**TAMSAT**	**WEATHER STATION**
NUMBER OF DAYS	10,593	10,593	10,593	10,593	10,593	10,573
MEAN PRECIPITATION	0.95	0.65	0.57	0.79	3.84	1.13
SD PRECIPITATION	3.48	2.7	2.43	4.10	4.79	8.44
MIN PRECIPITATION	0.00	0.00	0.00	0.00	0.00	0.00

MAX PRECIPITATION	48	34	35	70.07	35.90	236.00

Fig. 2 illustrates the precipitation distribution in Ahero over the observation period and highlights the differences in precipitation data across the datasets. Precipitation over the region features a positive skew, with few instances of extreme rainfall. This distribution is carried across all datasets. However, we find that the precipitation data obtained from RCMs recorded no or low precipitation in the region over the observation period compared to the other datasets. Previous evaluations of RCM performance over the East African region found RCMs underestimate rainfall, especially during the March-April-May (MAM) seasons (Luhunga, Botai and Kahimba 2016). Previous studies report the poor performance of RCMs of the region in reproducing rainfall events when compared to observational data is attributed to various factors, including their different abilities to simulate mechanisms behind rainfall formation over the region due to the paucity of in situ data required for model parameterization (Mumo and Yu 2020).

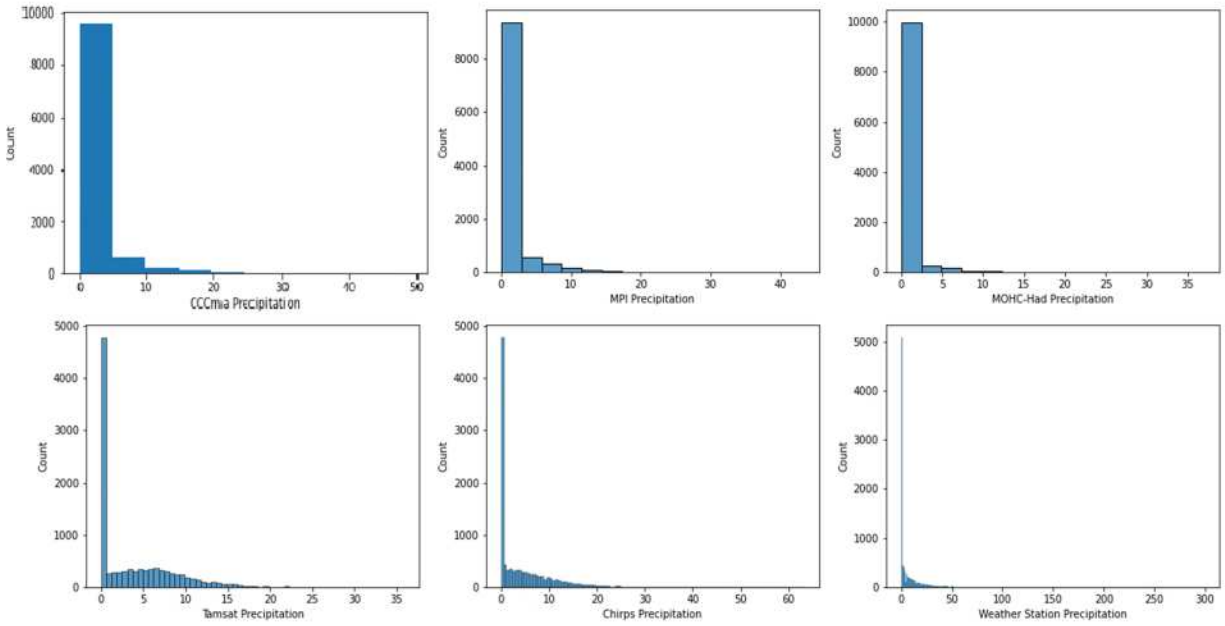


Fig. 2 *Distribution of precipitation across the selected datasets: CCCma, MPI, Had-GEM2 TAMSAT, CHIRPS, and Weather Station Data*

The number of floods identified by the SPI varied with each dataset. The variation in the number of floods across the selected datasets is due to the differing mechanisms in which the data products estimate rainfall. Each model performance was assessed against the unique characteristics of the data used; model performance varied with each dataset. In general, the anomaly detection models were able to identify more than half of the flood anomalies in the CHIRPS dataset but performed poorly in identifying flood anomalies in the climate model datasets. The best performance was obtained by the CBLOF anomaly detection method using the CHIRPS dataset. 80% of the anomalies detected had been identified as flood events based on the SPI. Precipitation distribution in the RCM datasets may be the reason for their poor performance in flood detection; most of the observations in the RCM data had no or low precipitation values recorded. Table 2 details the

number of flood anomalies detected by each model as a fraction of the number of floods identified by the SPI (last row):

Table 2: Algorithm performance calculated as a function of the number of floods detected compared to the number identified by SPI.

	CHIRPS	WEATHER STATION	TAMSAT	CCCMA	MPI	HAD-GEM2
CBLOF	80%	34%	25%	5%	1%	0%
HBOD	67%	46%	-	5%	1%	0%
IF	67%	33%	20%	1%	1%	0%
KNN	67%	40%	50%	5%	1%	-
No. Floods SPI	30	19	17	28	23	24

Model specificity was determined for each type of dataset. All models performed well in terms of classifying non-flood events correctly when using weather station data but were less accurate when utilizing climate model data. Table 3 summarizes the model specificity:

Table 3: Model specificity in classifying non-flood events

	WEATHER STATION	CHIRPS	TAMSAT	CCCMA	HAD-GEM2	MPI
HBOD	1	0.17	0	0	0	0
CBLOF	1	0.14	0	0	0	0
IF	1	0.17	0	0	0	0
KNN	1	0.25	0	0	0	0

Additionally, model sensitivity was determined against each type of dataset. All four models were able to identify flood anomalies correctly for each of the six dataset types (Table 4). However, this may be attributed to the zero-inflation (distribution that allows for frequent zero-valued observations) of the TAMSAT and the climate model data, indicating how skewed these datasets are in the detection of floods.

Table 4: Model sensitivity that measures the model ability to identify flood anomalies.

	CHIRPS	WEATHER STATION	TAMSAT	CCCMA	HAD-GEM2	MPI
HBOD	0.97	1	1	1	1	1
CBLOF	0.96	1	1	1	1	1
IF	0.97	1	1	1	1	1
KNN	0.944	0	1	1	1	1

Finally, Table 5 summarizes the area under the curve receiver operating characteristic (AUC-ROC) performance measure, which provides information on the models' ability to differentiate between flood and non-flood events. Isolation Forests model using weather station data had the best performance in distinguishing between flood events with an AUC-ROC measure of 0.8. Unlike HBOD, CBLOF and KNN, IF models identify anomalies independent of the underlying data distribution; the other anomaly detection algorithms build a profile of normal instances based on distance or density; high-density regions or short distances measured in a local context could be anomalies in the global data set or vice versa. Isolation forest path-length-based isolation traverse data to identify anomalies in local and global contexts. This may account for the discrepancies in performance between the 4 models. The CBLOF model also had good performance with the same dataset. The models performed especially poorly using data from the regional climate models. All models could not differentiate between flood and non-flood events using RCM data. The poor performance may be attributed to the precipitation distribution, which underestimated rainfall in the region, leading to an inflation of zeros in the dataset.

Table 5: Algorithm performance in differentiating between flood and non-flood events as ascertained using the AUC- ROC performance measure.

	CHIRPS	WEATHER STATION	TAMSAT	CCCMA	HAD-GEM2	MPI
HBOD	0.57	0.5	0.5	0.5	0.5	0.5
CBLOF	0.55	**0.732**	0.5	0.5	0.5	0.5
IF	0.566	**0.8**	0.5	0.5	0.5	0.5
KNN	0.597	0.5	0.5	0.5	0.5	0.5

3.2. Drought Detection in Turkana and Wajir Counties in Northern Kenya

3.2.1. Drought Detection using SPEI

SPEI was computed for the selected region using the three datasets. Fig. 3, Fig. 4, Fig. 5 summarize the 1-month SPEI variability in the three datasets for the period between 1992-2020 in Turkana and Wajir counties. Positive SPEI values indicate greater than normal precipitation, while negative values indicate less than normal precipitation. Therefore, extreme peaks and troughs in the graphs indicate flood or drought events, respectively. A cutoff point of -1.5 was used to identify extreme drought events in the two regions in the study. There is evidence of within-season variability with

experiences of moderate to severe and moderate to extreme drought cases within the period. Examining Fig. 3, 1-month SPEI values computed from weather station data shows dry events in the region occur towards the end of 2002 and early 2003; another long dry event starts in 2012 and ends in 2015 with intermittent wet periods, and again in the period from 2019-2020. Extreme drought events occurred in 2012 and the start of 2020.

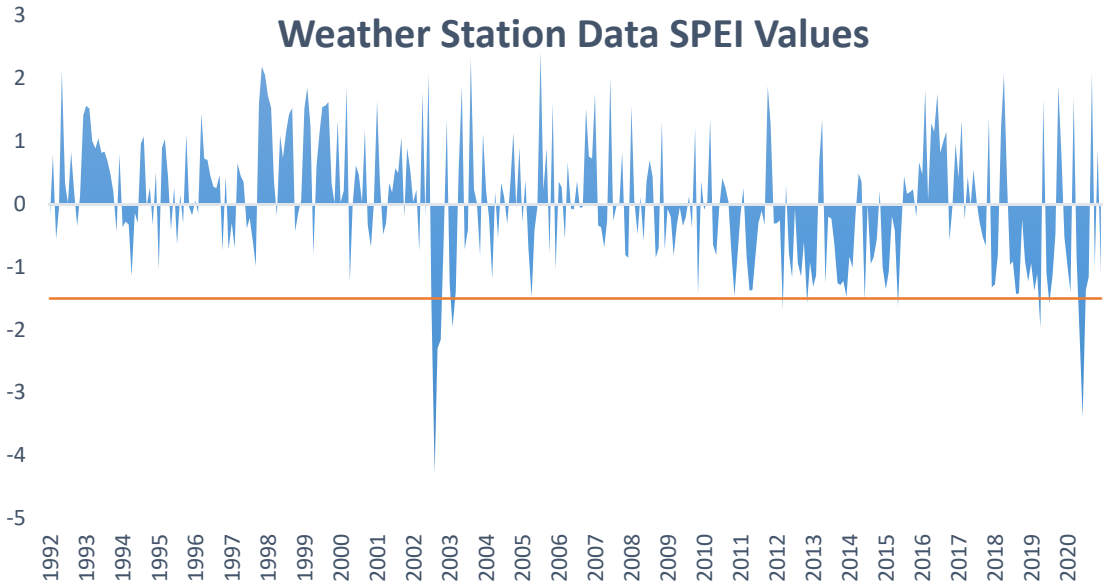
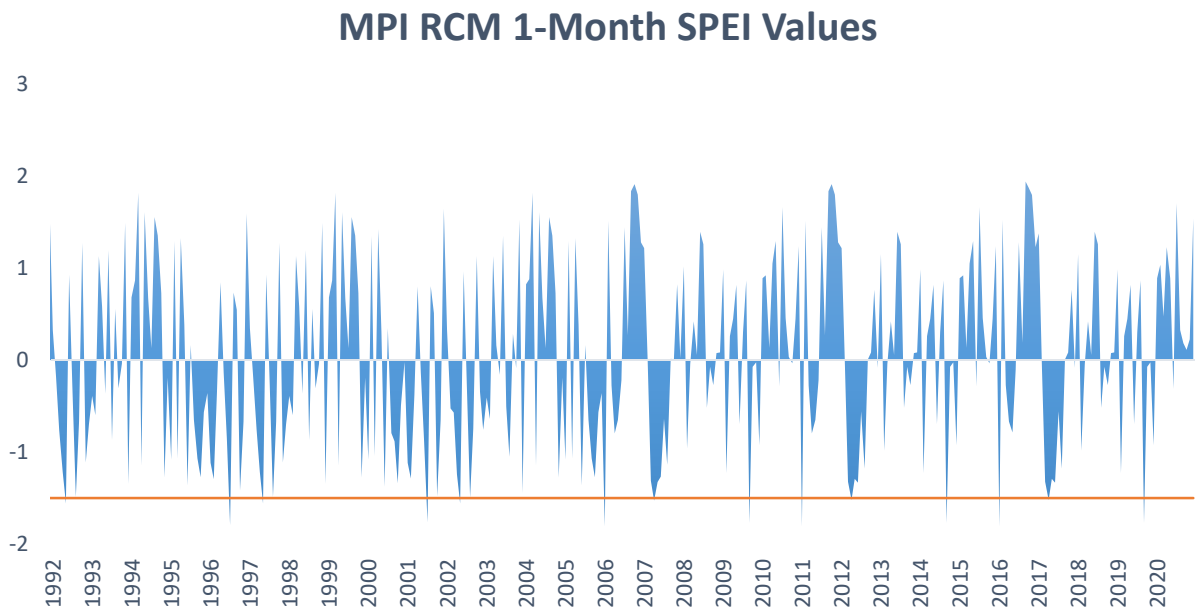


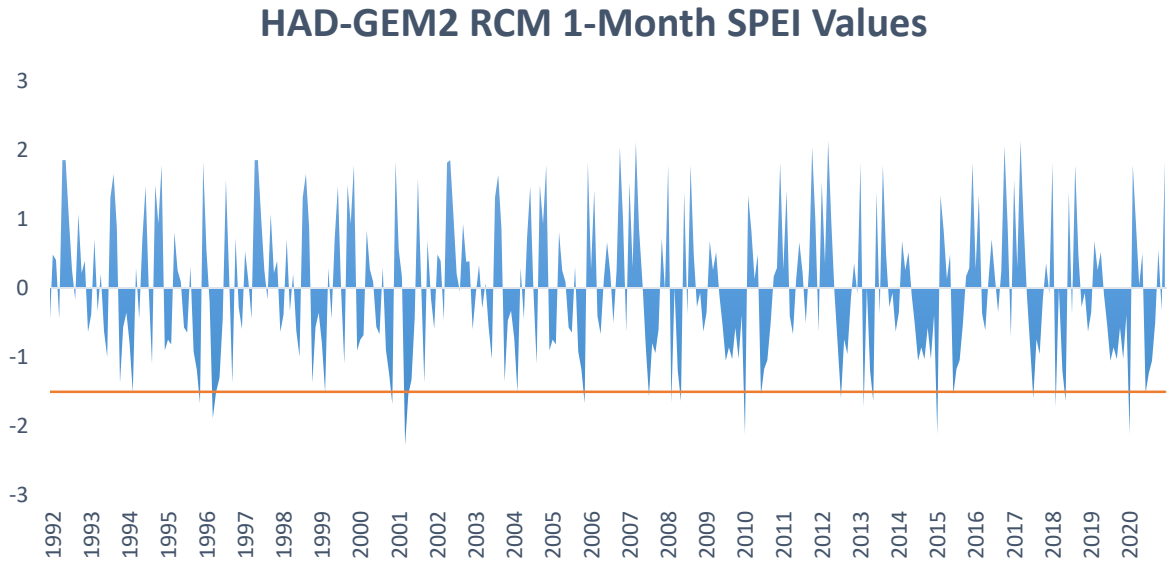
Fig. 3 *1-month SPEI variability in Weather Station Data for the years 1992-2020*

The MPI and Had-GEM2 RCMs capture several instances of drought events as summarized in Fig. 4 and Fig. 5 respectively. Moderate drought events were observed during the period 1992, 1996, 1997, 2001, 2002, 2007, 2009, 2011, 2012, 2014, 2016, 2017 and 2019 for MPI. While the Had-GEM2 captured extreme drought events in 2001, 2010, 2015 and 2020 in addition to the moderate events captured by the MPI dataset.



459

460 ***Fig. 4*** *1-month SPEI variability in MPI RCM for the years 1992-2020*



461

462 ***Fig. 5*** *1-month SPEI variability in Had-GEM2 RCM for the years 1992-2020*

463 The droughts captured in the observational data and the regional climate models correspond to
464 drought incidences reported in the literature (Ayugi, Tan and Gnitou, et al. 2020, Ayugi, Tan and
465 Niu, et al. 2020b, Owuor 2015). However, the time and frequency of drought instances differed
466 between the datasets. The discrepancies can again be attributed to varying precipitation distribution
467 between the observed weather station data and simulated RCM datasets.

3.2.2. ANN Drought Forecasting Model

Different ANNs were tested for forecasting the SPEI using weather station data and climatic model data. Table 6 lists the accuracy scores for the training and validation period. There is some reduction in the evaluation metric during the validation period. The low values in the validation period indicate that there is possibly some zero inflation, where the model predicts one scenario overall. Weather station data using the MLP neural network gave the best performance for 1-month drought forecasting. It had the highest training and validation accuracy scores and the lowest reduction in accuracy scores between training and validation. Among the regional climate models, Met Office Hadley Centre, MOHC-HadGEM2-ES (Had-GEM2) provided the best validation accuracy at 57%.

Table 6: Accuracy Scores of ANNs for 1-month Drought Forecasting of SPEI

	WEATHER STATION DATA		**MPI**		**HAD-GEM2**	
	Training	Validation	Training	Validation	Training	Validation
USER-CREATED NEURAL NETWORK	90%	48%	87%	45%	81%	57%
MULTI-LAYER PERCEPTRON (MLP) NEURAL NETWORK	78%	63%	63%	54%	70%	57%
VALIDATION ACCURACY	73%		63%		40.8%	

ANN models using weather station data had the highest AUC-ROC scores - indicating that the developed models could adequately distinguish between drought and non-drought events (Table 7). The regional climate models had low scores in this respect except for the MOHC-HAD data using the MLP neural network. Additionally, AUC-ROC increased to 0.65 in the validation period for weather station data, indicating the model's generalisability. The regional climate models, MPI and MOHC-HAD, had much lower scores in this respect, with the exception of the MOHC-HAD data using the MLP neural network. MLP performance was superior to the user-generated neural network due to the use of an Adam optimiser. The discrepancies between the RCM and Weather Station data results are difficult to attribute to a specific source. RCMs have deficiencies in projecting the observable climate in the East African region, but weather station data accuracy is also limited (generally, by the number of weather stations used for interpolation, incompleteness, or inaccuracies of station records).

Table 7: AUC-ROC Scores for the ANNs

	WEATHER STATION DATA	**MPI**	**HAD-GEM2**
USER-CREATED NEURAL NETWORK	0.79	0.48	0.44

MULTI-LAYER PERCEPTRON NEURAL NETWORK (MLP)	0.6	0.45	0.62

It is worth noting that model and dataset performance could only be evaluated against SPEI values as there are no drought or flood repositories with historical weather data on these phenomena in Kenya. Ground truth data is needed to validate the above findings.

4. Discussion and Conclusion

The main resource for agricultural production, the land, is heavily affected by agro-climatic conditions - meteorological events heavily influence crop farming and animal husbandry outcomes. Thus, extreme weather events often adversely impact farmers and pastoralists; and contribute to rural poverty. Farmers can lose their crops, means of production, and even access to their land through extreme weather events like floods and droughts. An increase in the intensity and frequency of these extreme weather events due to climate change will result in a corresponding increase in the negative impacts on agricultural productivity. Agricultural insurance, provided by government or private entities, can be a risk mitigation tool for farmers.

Agricultural insurance reduces the risks of lending to the farming sector by enabling farmers and pastoralists a means of repaying loans; eases budget volatility of agriculture-related fiscal expenditures by transferring climatic risk to the private sector; increases fiscal stability during crisis events and stimulates growth of the agricultural sector (Baskaran and Maher 2021). In Kenya, for example, as part of the response to the 2008 – 2011 drought, the government implemented a National Disaster Risk Financing Strategy, under which an agriculture insurance program, implemented in partnership with the private sector, targets vulnerable farmers; as of early 2021, more than half a million farmers were insured against agricultural losses under the program (Baskaran and Maher 2021). For the process to work successfully, however, insurance companies must have access to high-quality data to determine the risks of insuring farmers, agro-pastoralists, pastoralists, and their small businesses – a process known as underwriting. Underwriting, in addition to determining the risks posed by a prospective policyholder, also determines the cost of coverage. We contend that climate change models, with their ability to predict drought or flooding conditions, could be an indispensable part of the underwriting process.

Traditionally, calculation of risk for agricultural insurance involved analysis of yield data. However, this process is often undermined by the lack of high-quality yield data especially in rural regions, like the three regions in this study. An alternative to yield data is employing index-based insurance (IBI), e.g., rainfall-based index insurance. Index-based insurance relies on the measured value of an objective and independent index; the payment to the insured is based not on the assessment of the insured's actual loss but on the measured value from the index, i.e., the index serves as a proxy for the actual loss (Kenya Insurance Regulatory Authority 2015). For example, indemnity payments for the Kilimo Salama insurance in Kenya, Tanzania and Rwanda, and the Nyala Insurance Share Company (NISCO) in Ethiopia, are made for the insured crop, when actual rainfall in the cropping season, recorded in the nearest weather station, falls below predefined threshold levels (Dercon, et al. 2014, Wairimu, Obare and Odendo 2016). IBI can reduce the costs and difficulties of administering and delivering agricultural insurance. Index-based insurance

would necessitate the development of a model that related weather phenomena such as floods or drought, linearly or non-linearly, to risk. An important step in this process would be accurately determining the occurrence, and intensity, of future drought and flood events.

In this study, we identified two models that may be used to predict flood and drought events. The combination of ANNs and weather station data was the most effective in predicting future drought occurrences in Turkana and Wajir counties (both drought-prone regions) with accuracies ranging from 78% to 90%. In the case of flood forecasting, Isolation Forests models using weather station data had the best overall performance. The study also found that the regional climatic models (Max Planck Institute for Metrology Germany, MPI-M-MPI-ESM-LR (MPI; Canadian Centre for Climate Modelling and Analysis, CCCma-CanESM2 (CCCma), and Met Office Hadley Centre, MOHC-HadGEM2-ES (Had-GEM2)) utilized struggled to identify flood anomalies in Ahero county (all three model performance of monthly aggregation was very low) but the RCMs from CORDEX-Africa (Max Planck Institute for Metrology Germany, MPI-M-MPI-ESM-LR (MPI) and Met Office Hadley Centre, MOHC-HadGEM2-ES (MOHC-HAD)) utilised in the study had adequate performance in drought forecasting. Superior weather station data performance, in both flood and drought forecasting, could be attributed to the singe-point measurements provided by weather station data which are usually more accurate when compared to reanalysis or simulation data products. This suggests that improvements in coverage of weather station data in data-sparse regions would increase the accuracy in drought and flood forecasting of climate models which would, in turn, increase accuracy in the underwriting process.

IBI powered by predictive climate change models could democratize risk information. Since IBI would be based on an independent variable determined by the climate change models, and not the actual crop or animal loss, then the insurance companies could make information on risk and the resultant cost of coverage available to farmers ahead of the farming season for a particular crop. This would give farmers ample time to decide: (i) whether they needed coverage, (ii) what type of coverage they required, and (iii) how they would finance the cost of the policy. Coupling climate change models (with high predictive accuracy) to the IBI underwriting process would reduce basis risk – the risk that what is predicted by the index differs from farmers' experiences. A reduction in basis risk would increase farmers' trust in insurance coverage and is likely to increase the number of farmers who elect to obtain agricultural insurance coverage. This would lead to greater economic stability for individual farmers and rural communities (who predominantly rely on agricultural activities for revenue) as a whole. Using climate modelling in the underwriting process will ensure that the risk calculation process is an objective one – reducing ethical issues of human bias in determining coverage and cost of coverage. Finally, flood and drought trends predicted by the climate models may be used by the insurance companies to generate additional mitigation factors in a policy to aid farmers in reducing crop or animal loss from these extreme weather phenomena.

We intend, in the second phase of this project, to develop a model that utilizes the two identified drought and flood predictive climate change models, ANN and Isolated Forest models with weather station data, respectively, to calculate risks of crops grown in the regions of the study

mentioned in this paper. This model will be analysed against current methods of calculating risk for agricultural insurance policies in Kenya to determine which process is more efficacious.

References

Addor, N., O. Rössler, N. Köplin, M. Huss, R. Weingartner, and J. Seibert. 2014. "Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments." *Water Resources Research* 50 (10): 7541 - 7562. doi:https://doi.org/10.1002/2014WR015549.

Ayugi, B., G. Tan, G. T. Gnitou, M. Ojara, and V. Ongoma. 2020. "Historical evaluations and simulations of precipitation over East Africa from Rossby centre regional climate model." *Atmospheric Research* 232: 104705. doi:https://doi.org/10.1016/j.atmosres.2019.104705.

Ayugi, B., G. Tan, R. Niu, Z. Dong, M. Ojara, L. Mumo, H. Babaousmail, and V. Ongoma. 2020b. "Evaluation of meteorological drought and flood scenarios over Kenya, East Africa." *Atmosphere* 11 (3): 307. doi:https://doi.org/10.3390/atmos11030307.

Baskaran, G., and B. Maher. 2021. *Agricultural insurance: The antidote to many economic illnesses.* Brookings Institute. May 26. Accessed September 20, 2021. https://www.brookings.edu/blog/future-development/2021/05/26/agricultural-insurance-the-antidote-to-many-economic-illnesses/.

Berghuijs, W. R., R. A. Woods, C. J. Hutton, and M. Sivapalan. 2016. "Dominant flood generating mechanisms across the United States." *Geophysical Research Letters* 43 (9): 4382-4390. doi:https://doi.org/10.1002/2016GL068070.

Bichet, A., A. Diedhiou, B. Hingray, G. Evin, N. Touré, K. N. A. Browne, and K. Kouadio. 2020. "Assessing uncertainties in the regional projections of precipitation in CORDEX-AFRICA." *Climatic Change* 162 (2): 583-601. doi:https://doi.org/10.1007/s10584-020-02833-z.

Bjornlund, V., H. Bjornlund, and A.F. Van Rooyen. 2018. "Why agricultural production in sub-Saharan Africa remains low compared to the rest of the world – a historical perspective." *International Journal of Water Resources Development* 36 (1): 20 - 53. doi:https://doi.org/10.1080/07900627.2020.1739512.

Boko, M., I. Niang, A. Nyong, C. Vogel, A. Githeko, M. Medany, B. Osman-Elasha, R. Tabo, and P. Yanda. 2007. "Africa. Climate Change 2007: Impacts, Adaptation and Vulnerability." In *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental*, by M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, 433-467. Cambridge, UK: Cambridge University Press. https://cgspace.cgiar.org/handle/10568/17019.

Brunner, M. I., L. Slater, L. M. Tallaksen, and M. Clark. 2021. "Challenges in modeling and predicting floods and droughts: A review." *Wiley Interdisciplinary Reviews: Water* 8 (3): e1520. doi:https://doi.org/10.1002/wat2.1520.

Çelik, M., F. Dadaser-Çelik, and A. Ş. Dokuz. 2011. "Anomaly detection in temperature data using DBSCAN algorithm." *2011 International Symposium on Innovations in Intelligent Systems*

612 *and Applications.* Istanbul: IEEE. 91-95.
613 doi:https://doi.org/10.1109/INISTA.2011.5946052.

614 Chalapathy, R., and S. Chawla. 2019. "Deep learning for anomaly detection: A survey." *arXiv*
615 *preprint arXiv:1901.03407.*

616 Chandola, V., A. Banerjee, and V. Kumar. 2009. "Anomaly Detection: A Survey." *ACM*
617 *Computing Surveys (CSUR)* 41 (15): 1-58. doi:https://doi.org/10.1145/1541880.1541882.

618 Chen, H., A.K. Githeko, G. Zhou, J.I. Githure, and G. Yan. 2006. "New records of Anopheles
619 arabiensis breeding on the Mount Kenya highlands indicate indigenous malaria
620 transmission." *Malaria Journal* 5 (17). doi:https://doi.org/10.1186/1475-2875-5-17.

621 Connolly-Boutin, L., and B. Smit. 2016. "Climate change, food security, and livelihoods in Sub-
622 Saharan." *Regional Environmental Change* 16: 385 - 399.
623 doi:https://doi.org/10.1007/s10113-015-0761-x.

624 Das, M., and S. Parthasarathy. 2009. "Anomaly detection and spatio-temporal analysis of global
625 climate system." *Proceedings of the third international workshop on knowledge discovery*
626 *from sensor data.* Paris: Association for Computing Machinery. 142-150.
627 doi:https://doi.org/10.1145/1601966.1601989.

628 Deo, R.C., and M. Sahin. 2015. "Application of the artificial neural network model for prediction
629 of monthly standardised precipitation and evapotranspiration index using
630 hydrometeorological parameters and climate indices in eastern Australia." *Atmospheric*
631 *Research* 161–162: 65 - 81. doi:https://doi.org/10.1016/j.atmosres.2015.03.018.

632 Dercon, S., R. V. Hill, D. Clarke, I. Outes-Leon, and A. S. Taffesse. 2014. "Offering rainfall
633 insurance to informal insurance groups: Evidence from a field experiment in Ethiopia."
634 *Journal of Development Economics* 106: 132-143.
635 doi:https://doi.org/10.1016/j.jdeveco.2013.09.006.

636 Food and Agriculture Organization (FAO). 2009. "FAO: Technical papers from the Expert
637 Meeting on How to Feed the World in 2050." November 11. Accessed September 8, 2021.
638 http://www.fao.org/wsfs/forum2050/wsfs-forum/en/on.

639 Fung, K., Y. Huang, C. Koo, and Y. Soh. 2020. "Drough forecasting: A review of modelling
640 approaches 2007-2017." *Journal of Water and Climate Change* 11 (3): 771-799.
641 doi:https://doi.org/10.2166/wcc.2019.236.

642 Goedde, L., A. Ooko - Ombaka, and G. Pais. 2019. *Winning in Africa's agricultural market.*
643 McKinsey and Company. February 19. Accessed September 8, 2021.
644 https://www.mckinsey.com/industries/agriculture/our-insights/winning-in-africas-
645 agricultural-market.

646 Goldstein, M., and A. Dengel. 2012. "Histogram-based outlier score (HBOS): A fast unsupervised
647 anomaly detection algorithm." *35th German Conference on Artificial Inteligence (KI-*
648 *2012).* Germany. 59-63.

649 Hakala, K., N. Addor, C. Teutschbein, M. Vis, H. Dakhlaoui, and J. Seibert. 2019. "Hydrological
650 Modeling of Climate Change Impacts." In *Encyclopedia of Water: Science, Technology,*
651 *and Society*, 1-20. John Wiley & Sons.
652 doi:https://doi.org/10.1002/9781119300762.wsts0062.

653 Hanel, M., O. Rakovec, Y. Markonis, P. Máca, L. Samaniego, J. Kyselý, and R. Kumar. 2018.
654 "Revisiting the recent European droughts from a long-term perspective." *Scientific Reports*
655 8 (1): 9499. doi:https://doi.org/10.1038/s41598-018-27464-4.

656 Hannah, L. 2015. "The Climate System and Climate Change." In *Climate Change Biology (Second*
657 *Edition)*. Cambridge, MA: Academic Press.

658 Hargreaves, G. H., and Z. A. Samani. 1985. "Reference crop evapotranspiration from
659 temperature." *Applied Engineering in Agriculture* 1 (2): 96-99.
660 doi:https://doi.org/10.13031/2013.26773.

661 Hazell, P., and S. Wood. 2008. "Drivers of change in global agriculture." *Philosophical*
662 *Transactions of the Royal Society B: Biological Sciences* 363 (1491): 495-515.
663 doi:https://doi.org/10.1098/rstb.2007.2166.

664 He, Z., X. Xu, and S. Deng. 2003. "Discovering cluster-based local outliers." *Pattern Recognition*
665 *Letters* 24 (9-10): 1641 - 1650. doi:https://doi.org/10.1016/S0167-8655(03)00003-5.

666 Inoubli, R., A. B. Abbes, I. R. Farah, V. Singh, T. Tadesse, and M. T. Sattari. 2020. "A review of
667 drought monitoring using remote sensing and data mining methods." *5th International*
668 *Conference on Advanced Technologies for Signal and Image Processing (ATSIP) 2020.*
669 Tunisia: IEEE. doi:https://doi.org/10.1109/ATSIP49331.2020.9231697.

670 Juma, B., L. O. Olang, M. Hassan, S. Chasia, V. Bukachi, P. Shiundu, and J. Mulligan. 2020.
671 "Analysis of rainfall extremes in the Ngong River Basin of Kenya: Towards integrated
672 urban flood risk management." *Physics and Chemistry of the Earth, Parts A/B/C* 102929.
673 doi:https://doi.org/10.1016/j.pce.2020.102929.

674 Kebede, A., A. Hasen, and W. Negatu. 2011. "A comparative analysis of vulnerability of
675 pastoralists and agro-pastoralists to climatechange: A case study in Yabello Woreda of
676 Oromia Region, Ethiopia." *Ethiopian Journal of Development Research* 33 (1): 61-95.
677 doi:https://doi.org/10.4314/ejdr.v32i2.68611.

678 Kenya Insurance Regulatory Authority. 2015. "The Kenya Index-Based Insurance." June 12.
679 Accessed September 20, 2021.
680 https://www.ira.go.ke/images/docs/THE_DRAFT_KENYA_INDEX_BASED_INSURA
681 NCE_POLICY_PAPER_2015.pdf.

682 Kundzewicz, Z.W., S. Kanae, S.I. Seneviratne, J. Handmer, N. Nicholls, P. Peduzzi, R. Mechler,
683 et al. 2014. "Flood risk and climate change: Global and regional perspectives."
684 *Hydrological Sciences Journal* 59 (1): 1-28.
685 doi:https://doi.org/10.1080/02626667.2013.857411.

686 Liu, F. T., K. M. Ting, and Z. H. Zhou. 2012. "Isolation-based anomaly detection." *ACM*
687 *Transactions on Knowledge Discovery from Data (TKDD).* 6 (1): 1-39.
688 doi:https://doi.org/10.1145/2133360.2133363.

689 Luhunga, P., J. Botai, and F. Kahimba. 2016. "Evaluation of the performance of CORDEX
690 regional climate models in simulating present climate conditions of Tanzania." *Journal of*
691 *Southern Hemisphere Earth Systems Science* 66 (1): 32-54.
692 doi:https://doi.org/10.22499/3.6601.005.

693 Madadgar, S., A. AghaKouchak, S. Shukla, A. W. Wood, L. Cheng, K. L. Hsu, and M. Svoboda.
694 2016. "A hybrid statistical-dynamical framework for meteorological drought prediction:
695 Application to the southwestern United States." *Water Resources Research* 52 (7): 5095-
696 5110. doi:https://doi.org/10.1002/2015WR018547.

697 Maidment, R., E. Black, and M. Young. 2017. "TAMSAT Daily Rainfall Estimates (Version 3.0)."
698 *Dataset.* University of Reading. Accessed September 9, 2021.
699 doi:https://doi.org/10.17864/1947.112.

700 Makena, B., M. Osunga, S. Kingori, and H.S. Abdillahi. 2021. *An Application of Flood Risk*
701 *Analysis for Impact Based Forecasting in Kenya.* Nairobi, Kenya: International Center for
702 Humanitarian Affairs and Kenya Red Cross.

703 McKee, T. B., N. J. Doesken, and J. Kleist. 1993. "The relationship of drought frequency and
704 duration to time scales." *In Proceedings of the 8th Conference on Applied Climatology* 17
705 (22): 179-183.

706 Meier, H. M., A. Höglund, R. Döscher, H. Andersson, U. Löptien, and E. Kjellström. 2011.
707 "Quality assessment of atmospheric surface fields over the Baltic Sea from an ensemble of
708 regional climate model simulations with respect to ocean dynamics." *Oceanologia* 53: 193-
709 227. doi:https://doi.org/10.5697/oc.53-1-TI.193.

710 Mendoza, P. A., M. P. Clark, N. Mizukami, A. J. Newman, M. Barlage, E. D. Gutmann, R.M.
711 Rasmussen, B. Rajogopalan, L.M. Brekke, and J. R. Arnold. 2015. "Effects of hydrologic
712 model choice and calibration on the portrayal of climate change impacts." *Journal of*
713 *Hydrometeorology* 16 (2): 762-780. doi:https://doi.org/10.1175/JHM-D-14-0104.1.

714 Min, E., W. Hazeleger, G. J. Van Oldenborgh, and A. Sterl. 2013. "Evaluation of trends in high
715 temperature extremes in north-western Europe in regional climate models." *Environmental*
716 *Research Letters* 8 (1): 014011. doi:https://doi.org/10.1088/1748-9326/8/1/014011.

717 Mirza, M. M. Q. 2003. "Climate change and extreme weather events: Can developing countries
718 adapt?" *Climate Policy* 3 (3): 233-248. doi:https://doi.org/10.1016/S1469-3062(03)00052-
719 4.

720 Mizra, M.M.Q., and A.U. Ahmed. 2010. "Global Environmental Changes in South Asia." In *A*
721 *Review on Current Status of flood and drought forecasting in South Asia.*, edited by A.P.
722 Mitra and C. Sharma, 233-243. Dodrecht: Springer. doi:https://doi.org/10.1007/978-1-
723 4020-9913-7.

Mizukami, N., Martyn P. Clark, Ethan D. Gutmann, Pablo A. Mendoza, Andrew J. Newman, Bart Nijssen, B. Livneh, L.E. Hay, J.R. Arnold, and L. D. Brekke. 2016. "Implications of the methodological choices for hydrologic portrayals of climate change over the contiguous United States: Statistically downscaled forcing data and hydrologic models." *Journal of Hydrometeorology* 17 (1): 73-98. doi:https://doi.org/10.1175/JHM-D-14-0187.1.

Muller, C., W. Cramer, W.L. Hare, and H. Lotze-Campen. 2011. "Climate change risks for African agriculture." Edited by Robert W. Kates. *Proceedings of the National Academy of Sciences.* 4313-4315. doi:https://doi.org/10.1073/pnas.1015078108.

Mumo, L., and J. Yu. 2020. "Gauging the performance of CMIP5 historical simulation in reproducing observed gauge rainfall over Kenya." *Atmospheric Research* 236: 104808. doi:https://doi.org/10.1016/j.atmosres.2019.104808.

Mutsotso, R. B., A. W. Sichangi, and G. O. Makokha. 2018. "Spatio-temporal drought characterization in Kenya from 1987 to 2016." *Advances in Remote Sensing* 7 (2): 125. doi:https://doi.org/10.4236/ars.2018.72009.

National Drought Managaement Authority. 2020. *Turukana County Drought Early Warning Bulletin for December 2020.* OCHA Services, Relief Web. December. Accessed September 12, 2021. https://reliefweb.int/report/kenya/turkana-county-drought-early-warning-bulletin-december-2020.

National Management Authority. 2020. *Wajir County: Drought Early Warning Bulletin for December 2020.* OCHA Services, Relief Web. December. Accessed September 12, 2021. https://reliefweb.int/report/kenya/wajir-county-drought-early-warning-bulletin-december-2020.

Niang, I., O.C. Ruppel, M.A. Abdrabo, A. Essel, C. Lennard, J. Padgham, and P. Urquhart. 2014. "Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change." 1199-1265. Cambridge: Cambridge University Press.

Opere, A. 2013. "Floods in Kenya." *Developments in Earth Surface Processes* (Elsevier) 16: 315-330. doi:https://doi.org/10.1016/B978-0-444-59559-1.00021-9.

Osbahr, H., C. Twyman, N. Adger, and D.S.G. Thomas. 2008. "Effective livelihood adaptation to climate change disturbance: Scale dimensions of practice in Mozambique." *Geoforum* 39 (6): 1951-1964. doi:https://doi.org/10.1016/j.geoforum.2008.07.010.

Owuor, P. 2015. "The disaster profile of Kenya." Unit for Research in Emergency and Disaster, University of Oviedo, Spain, 1-45.

Pfister, S., P. Bayer, A. Koehler, and S. Hellweg. 2011. "Projected water consumption in future global agriculture: Scenarios and related impacts." *Science of the Total Environment* 409 (20): 4206–4216. doi:https://doi.org/10.1016/j.scitotenv.2011.07.019.

760 Reich, P.F., S.T. Numben, R. Almaraz, and H. Eswaran. 2001. "Land resources stresses and
761 desertification in Africa." *Agro-Science* 2 (2): 1-10.
762 doi:https://doi.org/10.4314/as.v2i2.1484.

763 Richard, Y., N. Fauchereau, I. Poccard, M. Rouault, and S. Trzaska. 2001. "20th century droughts
764 in southern Africa: spatial and temporal variability, teleconnections with oceanic and
765 atmospheric conditions." *International Journal of Climatology: A Journal of the Royal*
766 *Meteorological Society* 21 (7): 873 – 885. doi:https://doi.org/10.1002/joc.656.

767 Ringler, C., T. Zhu, X. Cai, J. Koo, and D. Wang. 2010. *Climate change impacts on food security*
768 *in sub-Saharan Africa. Insights from Comprehensive Climate Change Scenarios.*
769 International Food Policy Research Institute, Washington, D.C.: International Food Policy
770 Research Institute (IFPRI), 28.

771 Sarr, B. 2012. "Present and future climate change in the semi-arid region of West Africa: A crucial
772 input for practical adaptation in agriculture." *Atmospheric Science Letters* 13 (2): 108-112.
773 doi:https://doi.org/10.1002/asl.368.

774 Slater, L. J., and G. Villarini. 2018. "Enhancing the predictability of seasonal streamflow with a
775 statistical-dynamical approach." *Geophysical Research Letters* 45 (13): 6504-6513.
776 doi:https://doi.org/10.1029/2018GL077945.

777 Songok, C.K., E.C. Kipkorir, and E.M. Mugalavai. 2011. "Integration of Indigenous Knowledge
778 Systems into Climate Change Adaptation and Enhancing Food Security in Nandi and
779 Keiyo Districts, Kenya." In *Experiences of Climate Change Adaptation in Africa. Climate*
780 *Change Management*, edited by Filho W. Leal, 69-95. Hamburg, Berlin: Springer.
781 doi:https://doi.org/10.1007/978-3-642-22315-0_5.

782 Songok, C.K., E.C. Kipkorir, E.M. Mugalavai, A.C. Kwonyike, and C. Ngweno. 2011. "Improving
783 the Participation of Agro-Pastoralists in Climate Change Adaptation and Disaster Risk
784 Reduction Policy Formulation: A Case Study from Keiyo District, Kenya." In *Experiences*
785 *of Climate Change Adaptation in Africa. Climate Change Management.*, edited by Filho
786 W. Leal, 55–68. Hamburg, Berlin: Springer. doi:https://doi.org/10.1007/978-3-642-22315-
787 0_4.

788 Thomas, D.S.G., C. Twyman, H. Osbahr, and B. Hewitson. 2007. "Adaptation to climate change
789 and variability: Farmer responses to intraseasonal precipitation trends in South Africa."
790 *Climatic Change* 83 (3): 301 - 322. doi:https://doi.org/10.1007/s10584-006-9205-4.

791 Thompson, J., and I. Scoones. 2009. "Addressing the dynamics of agri-food systems: An emerging
792 agenda for social science research." *Environmental Science & Policy* 12 (4): 386-397.
793 doi:https://doi.org/10.1016/j.envsci.2009.03.001.

794 Tschakert, P., R. Sagoe, G. Ofori-Darko, and S.N. Codjoe. 2010. "Floods in the Sahel: An analysis
795 of anomalies, memory, and anticipatory learning." *Climatic Change* 103 (3): 471-502.
796 doi:https://doi.org/10.1007/s10584-009-9776-y.

Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno. 2010. "A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index." *Journal of Climate* 23 (7): 1696-1718. doi:https://doi.org/10.1175/2009JCLI2909.1.

Wairimu, E., G. Obare, and M. Odendo. 2016. "Factors affecting weather index-based crop insurance in Laikipia County, Kenya." *Journal of Agricultural Extension and Rural Development* 8 (7): 111-121. doi:https://doi.org/10.5897/JAERD2016.0769.

Wasko, C., and R. Nathan. 2019. "Influence of changes in rainfall and soil moisture on trends in flooding." *Journal of Hydrology* 575: 432 - 441. doi:https://doi.org/10.1016/j.jhydrol.2019.05.054.

Wibisono, S., M. Anwar, A. Supriyanto, and I. Amin. 2021. "Multivariate weather anomaly detection using DBSCAN clustering algorithm." *Journal of Physics: Conference Series* (IOP Publishing) 1869 (1). doi:https://doi.org/10.1088/1742-6596/1869/1/012077.

Wilby, R. L. 2010. "Evaluating climate model outputs for hydrological applications." *Hydrological Sciences Journal* 55 (7): 1090-1093. doi:https://doi.org/10.1080/02626667.2010.513212.

Wlokas, H.L. 2008. "The impacts of climate change on food security and health in Southern Africa." *Journal of Energy in Southern Africa* 19 (4): 12-20.

Woodhouse, C. A., G. T. Pederson, K. Morino, S. A. McAfee, and G. J. McCabe. 2016. "Increasing influence of air temperature on upper Colorado River streamflow." *Geophysical Research Letters* 43 (5): 2174-2181. doi:https://doi.org/10.1002/2015GL067613.

Statements and Declarations

Funding

The research detailed in the submitted paper was made possible by a grant (Grant No. 109675-001/002) from the International Development Research Centre (IDRC).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author's Contribution

Dr.Angeline Wairegi and Dr. John Olukuru provided guidance on the study conceptualization, design, and execution. Material preparation, data collection and analysis were performed by Alvin M. Igobwa and Jeremy Gachanja. The manuscript was written by Dr. Angeline Wairegi and Betsy Muriithi. Dr. Angeline Wairegi, Dr. John Olukur, Alvin M. Igobwa, Jeremy Gachanja, Betsy Muriithi, and Dr. Isaac Rutenberg were involved in reviewing and editing of the manuscript, and approved the final manuscript.

Data Availability

The datasets generated and analysed during the current study are available on the Zenodo platform, DOI 10.5281/zenodo.5730074, [https://zenodo.org/record/5730074#.YgOeSN9BzIU].

Ethical Approval

Not Applicable

Consent to Participate

Not Applicable

Consent to Publish

All authors listed above consent to have the submitted manuscript published in the Climate Change journal.