Linking reported drought impacts with drought indices, water scarcity, and aridity: the case of Kenya

Marleen R. Lam[1], Alessia Matanó[2], Anne F. van Loon[2], Rhoda Odongo[2], Aklilu D. Teklesadik[3], Charles N. Wamucii[1], Marc J.C. van den Homberg[3], Shamton Waruru[4], and Adriaan J. Teuling[1]

[1]Hydrology and Quantitative Water Management (HWM), Wageningen University & Research (WUR), Wageningen, the Netherlands
[2]Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam, Amsterdam, the Netherlands
[3]510, an initiative of the Netherlands Red Cross, Anna van Saksenlaan 50, 2593 HT Den Haag, the Netherlands
[4]National Drought Management Authority (NDMA), Lonrho House, 8th Floor, Standard Street, P.O Box 10304, G.P.O. 00100 Nairobi, Kenya

Correspondence: Marleen R. Lam (marleen.lam@icloud.com), Adriaan J. Teuling (ryan.teuling@wur.nl)

Abstract. The relation between drought severity, as expressed through widely used drought indices, and drought impacts is complex. In particular in water-limited regions where water scarcity is prevalent, the attribution of drought impacts is difficult. This study assesses the relation between reported drought impacts, drought indices, water scarcity, and aridity across several

5 counties in Kenya. The monthly bulletins of the National Drought Management Authority in Kenya have been used to gather drought impact data. A Random Forest (RF) model was used to explore which set of drought indices best explains drought impacts on: pasture, livestock deaths, milk production, crop losses, food insecurity, trekking distance for water, and malnutrition. The findings of this study suggest a relation between drought severity and the frequency of drought impacts, whereby the latter also showed a relation with aridity, whilst water scarcity did not. The results of the RF model reveal that drought impacts

10 can be explained by a range of drought indices across regions with different aridity. While the findings strongly depend on the availability of drought impact data and the socio-economic circumstances within a region, this study highlights the potential of linking drought indices with text-based impact reports. In doing so, however, spatial differences in aridity and water scarcity conditions have to be taken into account.

1 Introduction

15 Drought events are among the world's most impact-full disasters (Stahl et al., 2016) and are receiving increasing attention across different scientific disciplines because of their complex links to both natural and socio-economic processes (Van Dijk et al., 2013; Van Loon et al., 2016a, b). Drought can be characterized as a slow-onset event whose impacts build up over time and extend spatially in relation to a range of contextual factors (Heinrich and Bailey, 2020). For instance, differences in societal and political characteristics can lead to a different range and magnitudes of impacts even though the intensity and duration of

20 drought are similar (Savelli et al., 2022). At the same time, catchment characteristics also strongly influence the severity and propagation of drought events (Van Loon, 2015).

Due to the projected increase in drought frequencies (Change, 2014), each successive drought event can result in increased destabilization, triggering insecurity and resource-based conflicts (Peng et al., 2020). Monitoring and early warning (M&EW) is one important measure to enhance drought resilience. The goal of M&EW is to provide reliable and timely information on

25 drought conditions (using a wide range of drought indices) to enable local society to better prepare and act accordingly (Wilhite et al., 2007). However, there is a gap between forecasting a hydro-meteorological event and the understanding of its potential impacts, as recognized by the World Meteorological Organization (WMO, 2015). Linking drought impacts to drought indices can contribute to the ongoing development and improvements of the M&EW, aiming to reduce human and financial losses arising from a drought event.

30 Drought can occur in different parts of the terrestrial hydrological cycle. Therefore, droughts are generally classified as meteorological, hydrological, or soil moisture (agricultural) drought, which are related to, and influenced by, both natural processes and human activities (Van Loon, 2015; Van Loon et al., 2016b). Meteorological drought is typically defined only by the precipitation deficit, thereby not taking into account any actual conditions at or below the land surface. The precipitation anomaly can then propagate through the hydrological cycle, resulting in soil moisture, streamflow and groundwater deficit and

35 hence, soil moisture and hydrological drought, respectively. Hydrological drought refers to anomalously low levels in surface water (lakes, reservoirs, rivers) and/or groundwater. The detection and quantification of drought is usually done by so-called standardized drought indices (Kchouk et al., 2022; Yihdego et al., 2019). By expressing the anomaly with respect to the mean and variability of the local climate, drought characteristics can be compared across regions with different climate conditions. In addition, accumulation periods can be used to account for time lags and memory encountered in hydrological stores (Sutanto

40 and Van Lanen, 2022). The most simple drought indices use only meteorological data while others include soil moisture or streamflow data (Yihdego et al., 2019). Models using drought indices to forecast drought can detect climate signals into soils and hydrology. Yet, the link between drought indices and socio-economic impacts has rarely been analyzed, in spite of its importance for developing future measures to reduce vulnerability to drought.

The assessment and monitoring of drought impacts is complex given: (1) the great variety of drought impact categories;

45 (2) their possible propagation throughout the hydrological and social system; and (3) the difficulty of drought impact attribution. For Europe and the USA, drought impact databases have been developed, namely the European Drought Impact report Inventory (EDC, 2013) and the Drought Impact Reporter (NDMC, 2005) respectively. Some studies assessed the link between drought impacts and drought indices, mainly with a focus on Europe. For instance, the qualitative dataset of EDII has been used to assess the link between drought impacts and indices at continental (Blauhut et al., 2015), national (Stagge et al., 2015),

50 and regional scale (Bachmair et al., 2015, 2016, 2018). The results of multiple studies suggest that linking drought indices with impacts is time, region, and sector specific (Bachmair et al., 2015, 2016, 2017; Blauhut et al., 2015; Ma et al., 2020; Stagge et al., 2015; Wang et al., 2020). This urges the need to study their relation in other setting such as the Horn of Africa where drought impact data has not been structurally gathered as compared to Europe and the US.

Linking drought impacts with drought indices is regarded difficult as there is often no strong intuitive cut-off within impact

55 categories between non-drought and drought conditions (Hall and Leng, 2019). For instance, water scarcity conditions can be the result of anthropogenic actions and can lead to the same impacts experienced as during drought conditions (Van Loon and

Van Lanen, 2013). Water scarcity is a frequent phenomenon within (semi)-arid regions (Maliva and Missimer, 2012) and it occurs when water demand (both societal and ecological water demand) exceeds water supply (Kimwatu et al., 2021). It often leads to long-term unsustainable use of water resources (Van Loon and Van Lanen, 2013). Whereas aridity, based on the ratio

60 of long-term annual precipitation and potential evapotranspiration rates (UNESCO, 1979), is regarded as a relatively constant value, water scarcity is dynamic in time and related to both decreases in water availability (drought) and increases in water demand. The simultaneous presence of both water scarcity (partly driven by anthropogenic causes) and meteorological drought in an arid region can lead to a difficult attribution of the impacts experienced. However, separation of the causes of impacts is needed to generate reliable information to stimulate early action in the affected sectors during drought events.

65 In this study, we focus on Kenya because of the presence of strong gradients in precipitation, aridity, water yield (i.e., amount of precipitation minus total actual evapotranspiration), and water scarcity (Mulwa et al., 2021; Wamucii et al., 2021), in combination with the availability of reported impacts of recent droughts. The country has experienced several drought events in the recent past: for instance, 2008–2011 was classified as a prolonged severe drought (Mutsotso et al., 2018) and the drought in 2016–2017 was considered a national disaster (Kew et al., 2021; Ondiko and Karanja, 2021), with more than three million

70 people under food insecurity (Thomas et al., 2020). The country has also experienced a diverse range of drought impacts such as cattle mortality, wildlife death, famine, human losses, and severe food shortages (Ondiko and Karanja, 2021). The presence of drought hazard, drought impacts, water scarcity, and aridity makes this country a suitable study area to analyse their relations. In this context, the following main research question is formulated: What is the relation of drought impacts with drought indices and with water scarcity under different arid circumstances?

75 It is expected that drought indices will show a somewhat similar response across different climatic zones in Kenya because of the standardized nature of drought indices, but that drought impact occurrences might vary between climate zones. We hypothesize that drought impacts (and therefore the relationship between drought indices and impacts) will differ across regions with different aridity characteristics in Kenya because of the distinct socio-economic settings, possibly making arid areas more vulnerable than more humid areas (Maliva and Missimer, 2012). It is also expected that water scarcity will show a relation with

80 aridity due to the presence of unreliable water conditions.

2 Data and methods

2.1 Study area

Kenya is situated in East-Africa. Its highest altitudes can be found in the central highlands (over 5000 m above sea level), and low-lying regions can be found in the East, Northwest, and Northeast. The country mostly has an arid and semi-arid

85 climate which comprises about 80% of the territory and hosts about one quarter of the population (FEWS NET, 2013) of approximately 53 million people (The World Bank, 2020). Mean annual rainfall is less than 250 mm in the semi-arid and arid areas and more than 2.000 mm in the mountainous areas. Long rains are occurring from March to May (MAM) while the short rains occur during October to December (OND) (Ayugi et al., 2020). The medium to high potential agricultural areas are in the highland areas in the central and western parts of the country (sub-humid/humid zones) where the population density is six

90   times the country's average. Farming is the primary livelihood (both subsistence as well as commercial) for more than 75%
      of the population. Less than 4% are pastoralists who mainly live in the semi-arid and arid regions which are characterized by
      poorly distributed and unreliable rainfall (FEWS NET, 2013).

      For this study, six counties have been selected according to different aridity levels, livelihood zones, and available drought
      impact information. Figure 1 presents the counties considered in this study (Fig.1a), the aridity (Fig.1b), and the livelihood

95   zones (Fig.1c). Marsabit is an arid county (arid index 0.03–0.20) in the Northern pastoral zone while Baringo, Kitui and Kwale
      are considered semi-arid (arid index 0.20–0.50). Baringo is located in the western part of Kenya and encompasses mostly a
      high potential agricultural zone while Kitui and Kwale are both mostly marginal mixed farming zones. Nyeri is situated in the
      central highlands and encompasses mostly a high potential agricultural zone. However, this study will specifically focus on a
      region in Nyeri, namely Kieni whereby the main livelihood is connected to agropastoral activities (FEWS NET, 2013). Narok

100   mostly consists of (agro)pastoral grounds. Both Nyeri and Narok are regarded sub-humid zone regions (arid index 0.50–0.75).

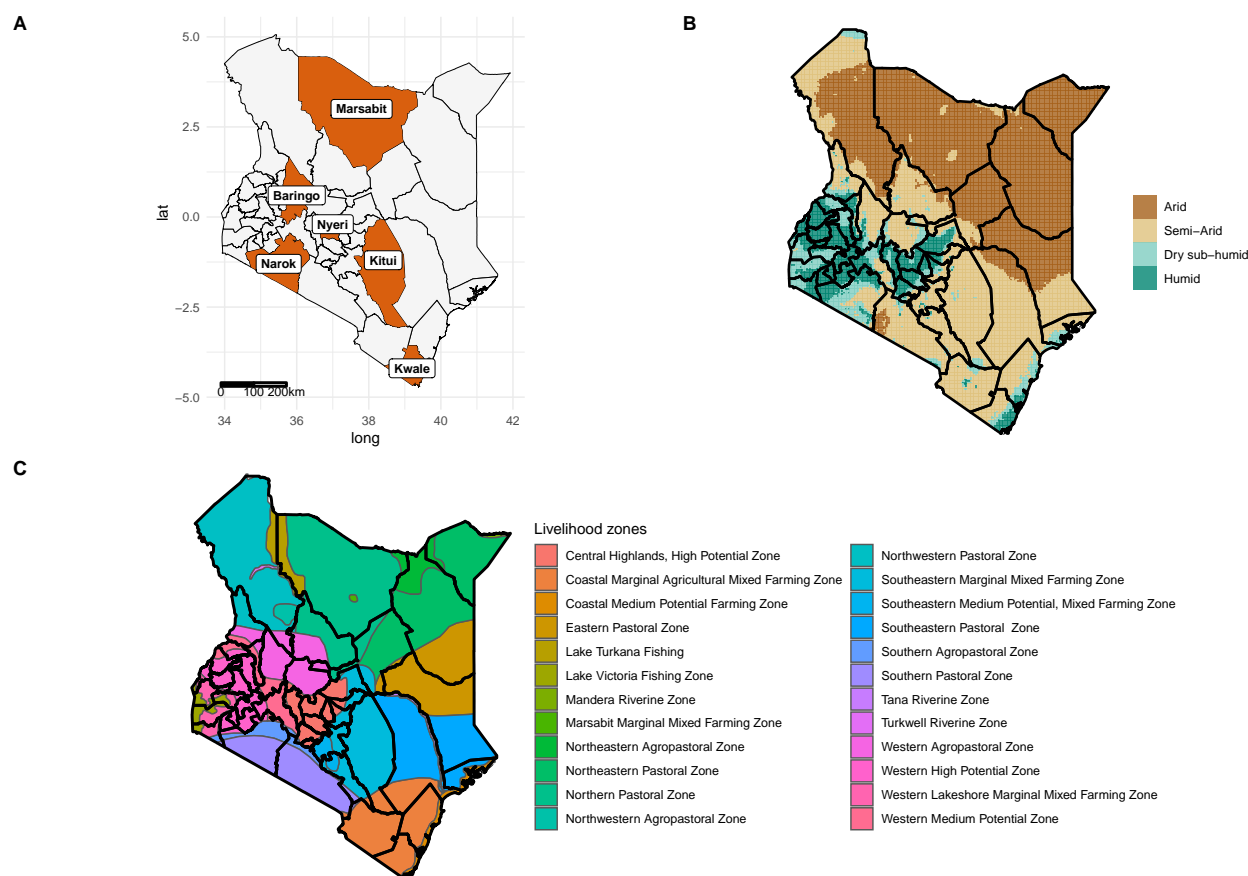



Figure 1. Study area and its main characteristics. a) Counties considered in this study, b) distribution of aridity, and c) distribution of
livelihood zones.

2.2 Data

To study the linkage between drought impacts, drought indices, water scarcity, and aridity, several datasets were used. In this study, we used re-analysis data to analyse several hydro-meteorological variables (Section 2.2.1), national drought bulletins for extracting text-based drought impact data (Section 2.2.2), and a gridded water scarcity dataset from McNally et al. (2019)
105 (Section 2.2.3).

2.2.1 Hydro-meteorological and soil moisture datasets

Precipitation data

Precipitation data has been retrieved from the Multi-Source Weighted-Ensemble Precipitation (MSWEP) version 1.1. This is a global gridded precipitation (P) dataset that takes full advantage of the complementary nature of highest quality gauge-
110 , satellite- and reanalysis-based P estimates, available as a function of timescale and location, by optimally combining them (Beck et al., 2017). MSWEP covers the period 1979–present at 3 hourly temporal and 0.1 degree spatial resolution. This dataset was chosen for this analysis based on its spatial and temporal resolution, good performance in capturing spatial and temporal variation of drought conditions (Xu et al., 2019).

Soil moisture and Potential Evapotranspiration data

115 The Global Land Evaporation Amsterdam Model (GLEAM) version 3.5a (v3.5a) consists of a set of algorithms dedicated to the estimation of land surface evaporation (also referred to as evapotranspiration) and root zone soil moisture from satellite and reanalysis data at the global scale and 0.25 degree spatial resolution (Martens et al., 2017; Miralles et al., 2011). The model uses the latest version of MSWEP precipitation, (v2.8), European Space Agency Climate Change Initiative (ESA-CCI) soil moisture (v5.3), and Vegetation optical depth (VOD) (Martens et al., 2017). The model uses the Priestley and Taylor equation
120 to calculate the Potential Evapotranspiration (PET) based on observations of ERA5 surface net radiation and near surface air temperature. GLEAM datasets have been used in multiple hydro-meteorological applications and recent drought conditions studies in the Horn of Africa (Javadinejad et al., 2019; Nicolai-Shaw et al., 2017; Peng et al., 2020). For this study, the GLEAM potential evaporation (PET) and root zone soil moisture were used (see http://www.gleam.eu) for the period 2010–2020.

Streamflow data

125 Streamflow data has been retrieved from the Global Flood Awareness System (GLOFAS) which consists of global gridded reanalysis river discharge data, with a horizontal resolution of 0.1 degree at a daily time step and time period of 1979–present. It combines surface and sub-surface runoff from the HTESSEL land surface model used within ECMWF's global atmospheric re-analysis (ERA5) with the LISFLOOD hydrological and channel routing model (Harrigan et al., 2020). LISFLOOD calculates a water balance at a six hourly or daily temporal resolution with 0.05 degree spatial resolution (see http://www.globalfloods.eu/).

130 2.2.2 The National Drought Management Authority

The monthly written bulletins containing drought impact data in Kenya were retrieved from the National Drought Management Authority (NDMA). The NDMA has offices in the 23 Arid and Semi-Arid Lands (ASALs) which are considered vulnerable

to drought. The Authority performs sentinel surveillance each month based on rainfall amounts, vegetation and water status. In addition, food security threats are assessed on (1) Availability aspects: cropping area and yield (maize, beans, sorghum

135 etc.), animal body condition, milk production, livestock death, and forage condition; (2) Access factors: market access and performance, food availability in the household and market. This data is fed into a web-based software created by the Kenya's Drought Early Warning System and sent directly to the county director. The director analyzes the data against the Vegetation Condition Index (VCI) (provided to them in the form of charts and graphs) where after he computes the Food Consumption index and the index of malnutrition (using the mid-upper-arm circumference).

140 2.2.3   Water scarcity

This study have utilized water scarcity (WS) data from McNally et al. (2019), a monthly water scarcity dataset with a spatial resolution of 0.1 degree for Africa between March 2018 and present. The water scarcity dataset is based on hydrological data from the Famine Early Warning System Network (FEWS NET) Land Data Assimilation System (FLDAS) and gridded population data from WorldPop (2015). The different classes of water scarcity are defined by the Falkenmark index. This index

145 categorises the amount of renewable freshwater available for each person per year, as shown in Table 1. The water scarcity dataset of McNally et al. (2019) provides monthly water scarcity data despite the yearly values of the Falkenmark index. However, McNally et al. (2019) used the yearly values of the Falkenmark index to classify the water scarcity on a monthly basis by using a 12-month running total of the streamflow data. The water scarcity dataset have been aggregated for the whole of Kenya whereafter monthly average values per county have been calculated and classified by the Falkenmark index. More

150 information about the water scarcity dataset can be found in Appendix A.

2.3   Methods

2.3.1   Drought indices

There are several widely-used standardized drought indices to characterize meteorological, hydrological and soil moisture drought. The Standardized Precipitation Index (SPI), devised by McKee et al. (1993), allows quantification of precipitation

155 deficits/ surpluses on a range of different accumulation periods. The SPI was calculated by summing daily MSWEP precipitation over n months (termed accumulation periods) obtaining a monthly temporal resolution. Monthly precipitation values were then ranked and their percentiles calculated. The number of zeros was taken into consideration following recommendations

Table 1. Falkenmark index.

Category	m^3/year /capita
No stress	>1700
Stress	1000–1700
Scarcity	500–1000
Absolute scarcity	<500

from Stagge et al. (2015). Thereafter, the values were fitted through a standard normal distribution to standardize the values between −3 and 3. Finally, the gridded SPI values were spatially aggregated to county resolution by averaging the SPI values of

160 all grid cell per county, to match the spatial resolution of the recorded impacts. A similar procedure was used in the calculation of the indices mentioned below.

The Standardized Precipitation Evapotranspiration Index (SPEI) is similar to SPI but also incorporates temperature by including evaporation anomalies (Vicente-Serrano et al., 2010). SPEI is calculated from the difference between precipitation and potential evapotranspiration. Thus, it provides a water balance and does not have the zero precipitation problems encountered

165 by SPI. The Standardized Soil Moisture Index (SSMI) is based on mean monthly GLEAM rootzone soil moisture content. Finally, the Standardized Streamflow Index (SSI) is based on mean monthly GloFAS discharge values (Nalbantis, 2008). A mask was created with mean monthly discharge values above 1 m^3/s. This mask was then used for the calculation of the SSI. SSI and SSMI are often used to take into account drought propagation through the hydrological cycle and are therefore able to better represent catchment memory compared to SPI and SPEI.

170 All the four drought indices (SPI, SPEI, SSMI and SSI) were calculated on a monthly timescale at the original grid scale with an accumulation period of 1, 3, 6, 12 and 24 months. The drought indices were calculated for the period 1980–2020. However, for investigating drought indices-impact relationships, we used drought indices between 2014–2020, in accordance with the availability of drought impact data.

2.3.2 Drought impact data

175 This research gathered impact data from the National Drought Management Authority (NDMA) for the above specified counties in Kenya, between 2014 and 2020. The NDMA was established by the Kenyan government in 2016 with the aim to set up and operate early warning drought systems and to develop drought preparedness strategies and contingency plans (Barrett et al., 2020). Their website provides monthly bulletins assessing food security in 23 regions using socio-economic and biophysical factors. These text-based impact reports provide the input for the impact categories considered in this study. The impact cate-

180 gories are based on the available information from the NDMA and can therefore be regarded as categories of socio-economic relevance for Kenya. The text-based impact data were turned into quantitative binary values by looking if the drought impact category value was different from the normal values for that time of the year (compared to previous years). The following impact categories were considered:

– Pasture (i.e. livestock migration pattern, quality and quantity pasture, livestock body condition);

185 – Livestock deaths;

– Milk production;

– Food insecurity;

– Crop losses;

    – Trekking distance to gather water for households;

190   – Malnutrition.

Jaccard similarity for binary values was used to measure the similarities between the occurrence of drought impact categories (Niwattanakul et al., 2013).

2.3.3 Random Forest Modelling

A machine learning algorithm, namely Random Forest (RF), have been used to assess the relation between the drought impact
195 categories and drought indices (Rpackage randomForest, version 4.6-14). It is a fairly new technique for linking drought indices with impacts but showed high potential in the studies of (Bachmair et al., 2016, 2017). The RF algorithm, proposed by Breiman (2001), combines several randomized decision trees and aggregates their predictions by averaging. It is designed to minimize the overall classification error which is irrespective of the class distribution (Elreedy and Atiya, 2019). Therefore, the datasets with amount of drought impacts per impact category and the corresponding drought indices was balanced by using a
200 synthetic minority oversampling technique (SMOTE) and randomized under-sampling (RUS). The min-max method was used as a normalization technique to scale the datasets. The drought impact datasets were aggregated by aridity: Marsabit (arid), Baringo, Kwale and Kitui (semi-arid) and Narok/Nyeri (sub-humid). We developed a random forest model for each of the drought impact categories aggregated per county with the same degree of aridity. Model performance was evaluated using a subset (25%) of the original dataset as test data. The area under the ROC (Receiver Operating Characteristic) curve (AUC)
205 describes the model's ability to predict the occurrence and non-occurrence of events correctly. A more detailed explanation about the RF model and the tuning of parameters can be found in Appendix C.

3 Results

3.1 Drought indices and drought impacts

To visually represent the relationships between drought impacts and drought indices, a timeframe from 2016 to 2020 is con-
210 structed to include the drought of 2016/2017. Most drought impacts were reported in Marsabit and Kitui while Baringo and Nyeri reported the lowest amount of impacts (Table 2). Table 2 presents the share of each impact category (in %) with respect to the total number of impacts per county. Pasture and Milk production are the most reported drought impacts across the counties, with values between 17.8 and 31.8%. Noticeable is that Nyeri has the highest share in pasture-related impacts: Pasture impacts are 29.6% and Milk production impacts are 31.8% of the total impacts for Nyeri.
215 The least reported drought impacts are on Crop losses, Livestock deaths and Food insecurity with average values of 3.1%, 7.8% and 10.1% respectively. Impacts related to Malnutrition are the highest in Baringo (17.8%) and Marsabit (16.9%), while Nyeri has by far the lowest amount of Malnutrition impacts (6.8%). Baringo has the highest share of impacts concerning Trekking distance for water (20.0%) while Nyeri has the lowest percentage (9.1%).

Table 2. Total amount of reported drought impacts between 2016 and 2020 and the share of drought impact categories (%) for each county.

County	Baringo	Kitui	Kwale	Marsabit	Narok	Nyeri
Count	45	93	50	124	51	44
Pasture (%)	17.8	30.1	28.0	20.2	25.5	29.6
Livestock deaths (%)	11.1	5.4	6.0	9.7	9.8	4.6
Milk production (%)	22.2	22.6	26.0	18.6	27.5	31.8
Food insecurity (%)	4.4	10.8	10.0	15.3	3.9	15.9
Crop losses (%)	6.7	1.1	4.0	2.4	2.0	2.3
Trekking distance water (%)	20.0	15.1	12.0	16.9	17.7	9.1
Malnutrition (%)	17.8	15.0	14.0	16.9	13.7	6.8

A timeline of the drought indicator SPEI for different accumulation periods (1, 3, 6, 12 and 24 months) and a timeline
220 with drought impacts are presented for Marsabit and Nyeri in Figures 2a and 2b for the time period 2016–2020. Noticeable
is that Marsabit experienced more extreme drought (in frequency and intensity) than Nyeri: SPEI-03 with a value of −2.22 in

(a) Marsabit   (b) Nyeri

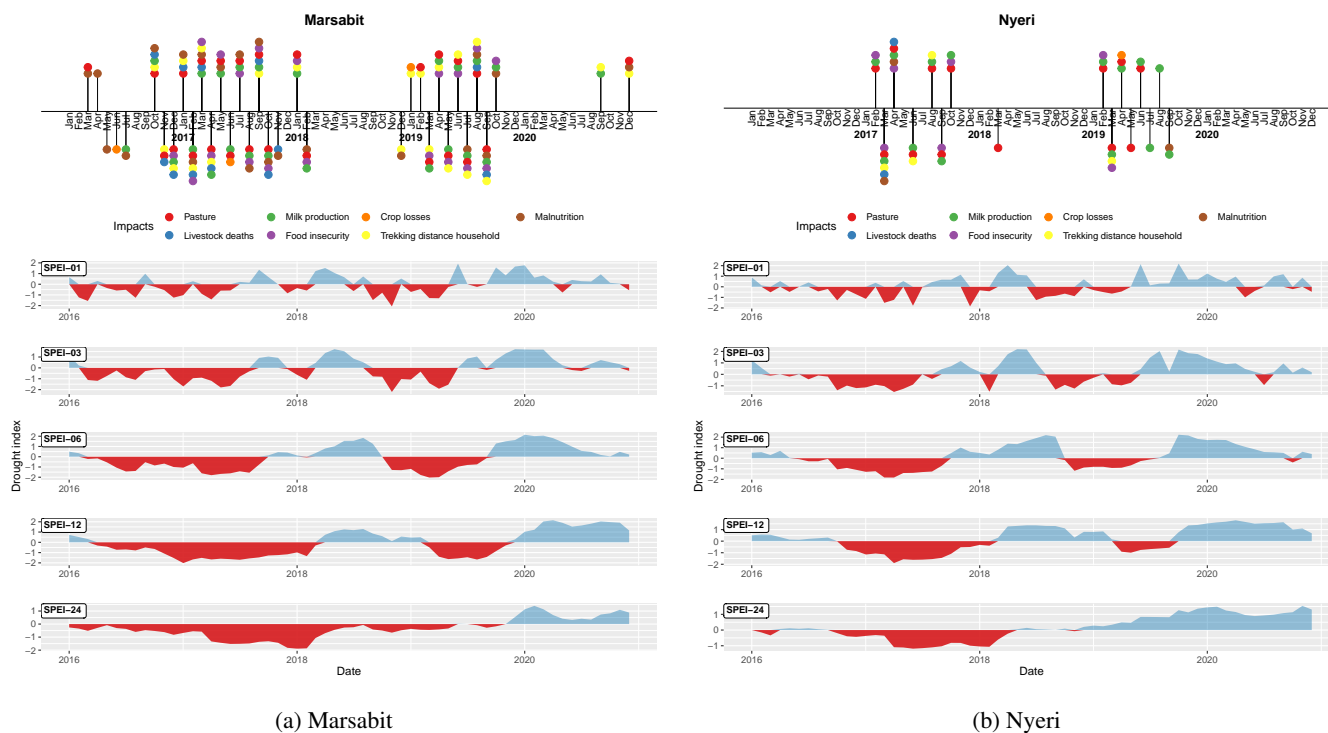



Figure 2. A timeline of the drought indicator SPEI for different accumulation periods (1, 3, 6, 12 and 24 months) and a timeline with drought
impacts for Marsabit and Nyeri. Negative values of SPEI (indicating a drought) are presented in red while positive values are blue.

Table 3. Correlation between the impact categories (Jaccard similarity).

Impact category	Pasture	Livestock deaths	Food insecurity	Milk production	Trekking distance water	Malnutrition
Livestock deaths	0.23					
Food insecurity	0.39	0.27				
Milk production	0.63	0.23	0.42			
Trekking distance water	0.50	0.26	0.29	0.47		
Malnutrition	0.41	0.20	0.27	0.34	0.34	
Crop losses	0.15	0.04	0.00	0.11	0.11	0.11

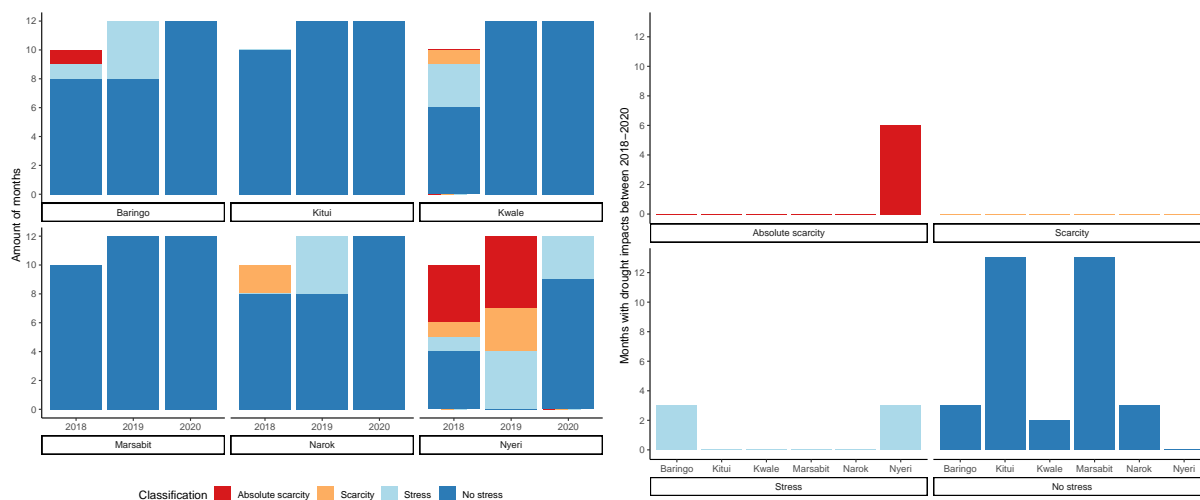
November 2018 was the most extreme drought for Marsabit while SPEI-12 with a value of −1.90 in April 2017 was the most extreme drought for Nyeri.

SPEI-24 indicates that Marsabit experienced a multiyear drought from January 2016 to May 2019. The drought of 2016–
225   2017 is well visible for both counties. In addition, there was a drought at the end of 2018 and 2019 which is more pronounced for Marsabit than for Nyeri. Regarding the drought impacts, Marsabit reported drought impacts (N = 124) from March 2016 until December 2020 with the exception of the periods between March and December 2018 and between November 2019 and August 2020. Nyeri reported drought impacts (N = 44) from February 2017 until September 2019 with only one impact reported between November 2017 and January 2019.

230   Taking the 2016/2017 drought as an example, the drought impacts reported in Marsabit are between March 2016 and February 2018 and highly overlap with SPEI-12, which is prevalent between April 2016 and March 2018. Reported drought impacts for Nyeri are between February 2017 and March 2017 and correspond most with SPEI-12 occurring from October 2016 until April 2018. A direct relationship with the other accumulation periods is not directly visible, except for SPEI-24. In particular, most of the analysed drought impacts occur after the onset of drought identified with accumulation periods of less than 12.

235   In this study, we also explored the relation between reported drought impacts (Table 3) by using the Jaccard similarity for binary values. Pasture and Milk production are a bit related (0.63) while Crop losses are not much related to any other impact category (<0.20). Trekking distance to water points indicates a bit of relation with Pasture (0.50) and Milk production (0.47). Other relations between impact categories are not very prevalent (<0.40).

3.2   Drought impacts and water scarcity

240   The degree of water scarcity per year (in number of months) across the counties is visualized in Figure 3a. Kitui and Marsabit experienced no water stress since March 2018 (start of timeframe WS dataset) while Nyeri experienced stress, scarcity and absolute scarcity during six out of ten months in 2018 and all months of 2019. Baringo, Kwale and Narok did also experience stress and scarcity conditions (respectively 2, 4 and 2 months out of 10 for 2018 (WS dataset starts at March 2018) and 4, 0 and 4 months out of 12 for 2019) but with a lower frequency than Nyeri.

(a) Water scarcity over March 2018 and 2020.

(b) Water scarcity and drought impacts.

Figure 3. The degree of water scarcity per year (2018–2020) across the counties (a) and months with drought impacts in relation to water scarcity (b) (McNally et al., 2019).

Figure 3b shows the amount of months with drought impacts during 2018 and 2020 in relation to the degree of water scarcity. Nyeri experienced 9 months with drought impacts between March 2018 and 2020, of which 6 months with absolute water scarcity and 3 months in a stress situation. Kitui and Marsabit experienced 14 months with drought impacts but did not experience any degree of water scarcity. Baringo had 6 months with drought impacts, of which half of the months were showing stress situations.

3.3 The Random Forest model

The performance of the Random forest (RF) models per impact category is shown in Table 4. The AUC ranges from 0.50 to 1.00. The performance of the models for the drought impacts on Pasture and Livestock have the best fit, with AUC values ranging from 0.87 to 1.00. Models developed for the drought impact of Malnutrition have the worst fit, with all AUC values below 0.60. The models for the arid region of Marsabit (MA) and the sub-humid regions of Narok and Nyeri (NA, NY) had the best overall fit with diverse ranges of performance among the impact categories. For instance, the model on Marsabit concerning Food insecurity has very high performance (1.00) while the one in Narok/Nyeri has a low performance (0.53). For Marsabit and Narok/Nyeri (arid and sub-humid regions respectively), the models related to activities of pasture (Pasture, Livestock deaths and Milk production) have very high performance rates (AUC > 0.87).

The occurrence of drought impacts in Trekking distance for water can be best predicted for the counties Narok and Nyeri with an AUC of 0.88. The models of the semi-arid counties of Baringo, Kitui and Kwale performed relatively poorly with the exception of the model developed for Crop losses (0.84).

Table 4. Performance of the RF model per impact category and arid characteristics (MA = Marsabit, BA = Baringo, KI = Kitui, KW = Kwale, NA = Narok and NY = Nyeri).

	Arid	Semi-arid	Sub-humid zone
	MA	BA, KI, KW	NA, NY
Pasture	0.87	0.62	0.96
Livestock deaths	1.00	0.71	0.97
Milk production	0.80	0.56	0.89
Food insecurity	1.00	0.67	0.53
Crop losses	0.55	0.84	0.64
Trekking distance household	0.55	0.57	0.88
Malnutrition	0.56	0.59	0.50

Figures 4 and 5 show the drought indices which are best linked with the drought impact categories; we take into account only the relations of the best performing models. The MeanDecreaseAccuracy (MDA: in %) represents the importance of the predictor for the model: it expresses how much accuracy the model loses when each variable would be excluded one by one. The mean decrease in Gini coefficient is a measure of how each variable contributes to the homogeneity of the nodes and leaves in the Random Forest. In this regard, homogeneity means that most of the samples at each node are from one class. A higher MeanDecreaseGini (in %) indicates higher explanatory importance.

(a) Pasture	(b) Livestock deaths

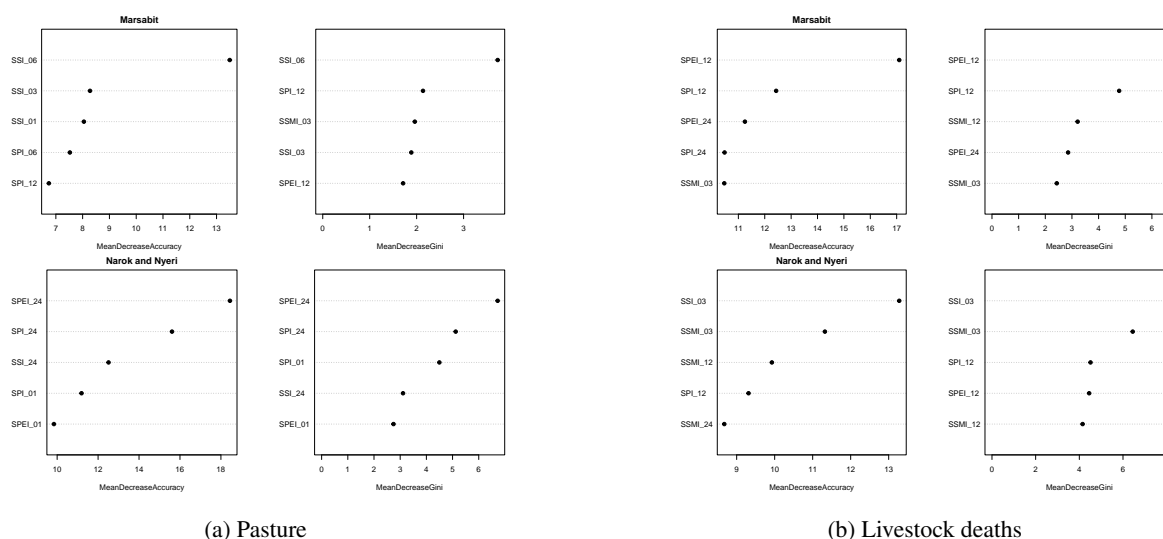



Figure 4. Drought indices best linked with Pasture and Livestock deaths for Marsabit and Narok/Nyeri. MeanDecreaseAccuracy (in %) indicates how much accuracy the model would loose when each variable would be excluded. The MeanDecreaseGini (in %) presents how each variable contributes to the homogeneity of the nodes and leaves in the Random Forest.

(a) Food insecurity (MA)

(b) Crop losses (BA, KW, KI)

(c) Trekking distance water points (NA/NY)

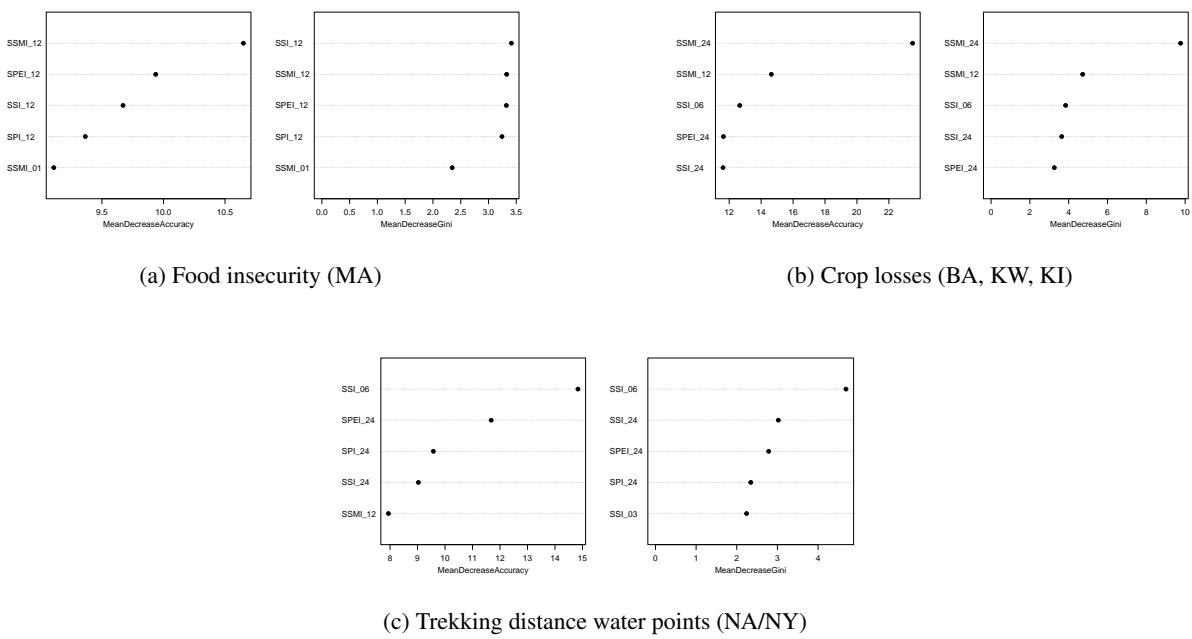



Figure 5. Drought indices best linked with Trekking distance water points (NA/NY), Crop losses (BA, KW, KI) and Food insecurity (MA).

As shown in Figure 4a, Pasture impacts for Marsabit tend to be related to shorter drought anomalies (6 months) than Narok and Nyeri (24 months). Furthermore, SSI is the best predictor for Pasture impacts in Marasbit while the meteorological

270 drought indices SPI and SPEI are the best predictors for Narok and Nyeri. SPEI-24 is by far the best performing drought indicator for Pasture in Narok/Nyeri with a MDA higher than 18%. For Livestock Deaths (Figure 4b), the situation is reversed: meteorological indices such as SPEI and SPI with longer accumulation periods (12–24 months) are the best link for Marsabit while the predictors SSI and SSMI with shorter accumulation periods (3–12 months) are the best link for Narok and Nyeri. SPEI-12 is by far the best performing drought indicator in Marsabit for Livestock deaths.

275 The results show that Food insecurity for the arid region of Marsabit can be well predicted with a range of drought indices (Figure 5a), among which SSMI is the main drought predictor. However, the accumulation period is more or less stable on 12 months. For Baringo, Kwale and Kitui, high accumulation periods (6–24 months) are associated with Crop losses, whereby SSMI, SSI and SPEI are the most prominent indices (Figure 5b). Noticeable is that SSMI-24 is by far the most prominent drought indicator for Crop Losses with a MeanDecreaseAccuracy of 24% and a MeanDecreaseGini of 10%. Trekking distance

280 to water points for Narok/Nyeri can mainly be predicted by SSI, SPEI and SPI with an accumulation period between 6–24 months (Figure 5c).

4 Discussion

4.1 Data sources and methods

This study used the water scarcity dataset of McNally et al. (2019) which is based on regional streamflow data and population data from WordPop 2015. This dataset has never been validated in the Horn of Africa which could be a limitation of this research. In addition, different hydrological datasets were used for the water scarcity dataset and the calculation of the SSI. However, despite some inconsistencies between the datasets, both are following the same pattern which justifies drawing conclusions based on the water scarcity dataset. The comparison between streamflow data of the water scarcity dataset and SSI-01 is included in Appendix B.

Drought impact data have been generated by analyzing the monthly county-specific reports of the NDMA. These reports had some monthly gaps and were sometimes written in different formats which made it hard to objectively appraise the drought impacts and their magnitude. However, iterative processes and focusing on abnormal conditions caused by drought according to the NDMA ensures a reliable list of drought impact occurrences. Despite the great effort and very valuable drought impact data information from the NDMA, this study stresses the need for an impact database for Africa such as the already existing databases EDII and DIR for Europa and the USA respectively. Future research can assess how to build an impact database with enhanced quality in terms of higher spatial and temporal resolution, more impact categories and more quantitative information on the impact. To build such a database for historical events, systematically blending and fusing impact data coming from different sources need to be explored (Majani et al., 2022). Sources that can possibly complement the NDMA bulletins range from global repositories such as EM-DAT or DesInventar, drought appeals from humanitarian organisations such as the Kenya Red Cross Society, index-insurance claims submitted to insurance companies or digital media reports.

We used a Random Forest technique to link drought impacts with drought indices. However, other literature used other techniques such as the Pearson correlation (Wang et al., 2020), Spearman correlation (Ma et al., 2020), and logistic regression (Bachmair et al., 2017; Blauhut et al., 2015; Stagge et al., 2015). Using RF to link drought indices with drought impacts is a fairly new technique and has only been done before by Bachmair et al. (2016, 2017) with a focus on Germany and the UK. These studies indicated a potential of using RF for drought M&EW. This study confirms this, as the performance indices (AUC) were good for several drought impact categories. However, using RF to link drought impacts with drought indices also shows the need to expand the drought impact data collection because of its sensitivity to data availability (Bachmair et al., 2016).

4.2 Relations with aridity

The majority of the drought impact data are livestock- and pasture-related. It was expected that the reported drought impacts were linked with the main livelihood activity of the county. This is partly confirmed by the results of this study. The main livelihood activity of Marsabit is pasture whereby around 50% of the drought impacts are related to this activity. Nyeri (Kieni sub-county) has agro-pastoral activities and more than 60% of the reported drought impacts is related to pastoral activities. Most of the reported drought impacts are related to pastoral activities, this could be due to the drought impact categories considered by the NDMA.

315 Marsabit and Kitui had the highest reported drought impacts while Baringo and Nyeri the least amount of reported drought impacts. This suggests that drought impacts are linked with aridity because Marsabit and Kitui contain larger areas classified as (semi-)arid than Baringo and Nyeri. The interference with socio-economic circumstances is likely to play a role as acute and chronic food insecurity, poverty, lack of economic development, limited access to basic social services, and low education levels are the highest among households in the ASALs (FEWS NET, 2013).

320 Drought impacts often appeared after the onset of a drought, therefore introducing a kind of lag. Maliva and Missimer (2012) stated that arid areas will have more extreme drought due to global warming which will increase the potential evapotranspiration. However, this study can not link drought occurrence to aridity because of the short timeframe (10 years) analyzed. The analysis of longer timeseries could indicate if there is an interannual trend and variability of drought indices, therefore determining whether there is a drying climate or a drought event (Xu et al., 2021). This could be an interesting follow-up research
325 whereby aridity conditions could be analysed in relation to drought occurrences.

4.3 Water scarcity and drought impacts

According to the water scarcity dataset, most drought impacts occurred at times without water stress, with the exception of Nyeri. These findings contrast with the text-based drought impact data on distance from water sources (i.e., Trekking distance for water) from the NDMA bulletins, which could be used as a proxy for water stress conditions. Increased distance from water
330 sources was reported in arid and semi-arid regions during most of the months when meteorological and hydrological drought conditions occurred (Figure 3b). The same impact was reported for three months during the 2017–2018 drought period and one month during the 2019 drought event in Nyeri (Figure 2b).

The discrepancies between the increased distance from water sources and the water scarcity index could be explained by the fact that the streamflow data used for developing the WS dataset were calculated without taking into account the presence
335 of reservoirs, located mainly in the central-western areas of Kenya (Lehner et al., 2011; Mulwa et al., 2021). In addition, the WS dataset uses population data as a proxy for water demand. Since the population density has high values in central-western counties and low values for the ASAL counties (which host only 25% of the population although they cover about 80% of the territory of Kenya; (FEWS NET, 2013)), it is not surprising that the WS index is higher for west-central counties than for ASAL counties.However, low population density does not imply low water stress: pastoral and agricultural livelihoods are
340 predominant in ASAL counties and are highly dependent on water availability. In addition, the staple foods for densely populated areas located in central-western counties are mainly supplied by the ASAL counties, resulting in high water consumption by the latter. Finally, water scarcity is also shaped by political choices, public policies, and social order (Savelli et al., 2021; Van Loon and Van Lanen, 2013). These factors were not accounted for in the development of the WS dataset.

In summary, the WS dataset is apolitical, does not take reservoirs into account, and is highly dependent on population
345 density, which is not a true reflection of water demand. Despite these limitations, interesting conclusions can still be drawn. The WS dataset highlights that water resources were sufficient to meet the water demand in the arid and semi-arid regions of Kenya during drought events. However, water insecurity in the ASAL regions was high during periods of drought (FEWS NET, 2017), possibly due to inefficient water management, poor maintenance of water supply systems (related in turn to corruption

and poverty) (Bellaubi and Boehm, 2018; Jenkins, 2017; Mulwa et al., 2021). The sub-humid central-western counties, on the
350 other hand, should have suffered from water scarcity during periods of drought due to the high population density and hence
the high pressure on available water resources. However, in reality, they experience little water stress thanks to the presence
of reservoirs that buffered the drought conditions (FEWS NET, 2017). This shows that water scarcity can be reversed through
wise usage of the available water resources (Phillip, 2013).

4.4 Drought indices and the Random Forest model

355 The results show that linking drought indices with drought impacts is region-specific, as confirmed by many other studies
(Bachmair et al., 2015, 2016, 2018; Blauhut et al., 2015; Ma et al., 2020; Parsons et al., 2019; Stagge et al., 2015; Wang et al.,
2020). For instance, shorter accumulation periods were found for Pasture at Marsabit (SSI-06) while longer accumulation
periods were found for Narok/Nyeri (SPEI-24). This lag suggests the presence of water buffers in Narok/Nyeri, damming the
sub-annual fluctuations in water availability and therefore generating less influence on the impact category Pasture (Mulwa
360 et al., 2021). On the contrary, Livestock deaths are linked with high accumulation periods in Marsabit (SPEI-12) and short
accumulation periods in Nyeri (SSI-03). These differences between the best match between drought impacts and drought
indices implies therefore a link with human activities. As confirmed by Xu et al. (2019), human activities can interfere with
natural processes and therefore influence the drought propagation time between meteorological and hydrological drought. This
calls for more research towards water management practices in relation to drought indices and drought impacts.

365 Regarding the drought indices, various drought indices are marked as the most optimal indicator: SSI is found in relation to
Livestock deaths (Marsabit), Pasture (Narok/Nyeri) and Trekking distance to water points (Narok/Nyeri), while SSMI is found
in relation to Crop losses (Baringo, Kwale, Kitui) and Food insecurity (Marsabit). Noticeable is that SSI gives a possible link
with water-dependent activities while SSMI shows a possible link with agricultural practices. It is expected that SSI and SSMI
would show a memory in relation to SPI and SPEI because of the propagation through the hydrological cycle, introducing a lag
370 between meteorological, soil moisture and hydrological drought (Wang et al., 2016). Therefore, the time length and duration
of SPI and SPEI can be used to express soil moisture and hydrological drought. In general a 1-month timescale is considered
meteorological drought, 3-6 months as soil moisture drought and 12 months can be considered as hydrological drought (Dai
et al., 2020). This link is partly visible by looking at the drought indices in relation to the accumulation periods. For instance,
SSI-06 is the best match for Trekking distance household which indicates hydrological drought. The best link after SSI-06 are
375 SPEI and SPI with a 24 months timescale, also indicating the presence of a hydrological drought.

Studies that linked drought impacts with drought indices are mainly focused on Europe (Bachmair et al., 2015, 2016, 2018;
Blauhut et al., 2015; Parsons et al., 2019; Stagge et al., 2015) and recently China (Ma et al., 2020; Wang et al., 2020). Com-
parisons with these studies are quite difficult due to the different socio-economic and climatic circumstances. As studied by
Bachmair et al. (2018), SPI and SPEI with an accumulation period of three and four months showed the highest correlation
380 for the impacts on crops in Germany. This is not consistent with the results found in this study in relation to Crop losses for
Baringo, Kitui and Kwale: those accumulation periods are quite high (6–24 months). As stated in the study of Bachmair et al.
(2018), an accumulation period of one month was found to have a notably lower correlation with drought impacts and was often
non-significant which is also confirmed by the results of this study. A reasonable explanation for this is that the occurrence of

impacts lags behind the occurrence of drought. Another study of Bachmair et al. (2016), showed that SPI and SPEI with longer
385 accumulation periods (12–24 months) are best linked to impact occurrence in the UK when using the RF model. In general,
this does match with the results of this study: SPI-12, SPEI-12, SPI-24 and SPEI-24 are the most occurring accumulation
periods, linking the occurrence of drought impacts with the presence of hydrological drought. Our results indicate that impacts
associated with different types of drought have different response times, as confirmed by the distinct differences in drought
indices and impact linkage pattern.

390 It should be noted that adaptation measures can influence the optimal drought index found by using the RF model. The
use of adaptation measures is linked with increasing livelihood resilience whereby smallholders are better prepared for future
challenges (Nyberg et al., 2020). The past years Kenya has experienced several drought events. This can influence the extent of
adaptation measures taken and therefore the resilience against droughts which affects the impacts. It is therefore recommended
to link adaptation measures to drought impacts and indices in order to analyze spatial differences and to map fluctuations over
395 time.

This study contributes to the ongoing debate about the operational needs for drought monitoring by linking multiple drought
indices to reported drought impacts. Results show the best drought index for a given impact which can be combined with other
socio-economic and environmental data to provide enough inputs for the construction of drought impact forecasting, useful for
stakeholders and decision makers (Heinrich and Bailey, 2020; Stagge et al., 2015). In addition, this research takes first steps
400 in analysing the link between drought and water scarcity and aridity, which is valuable information for the existing literature
database on drought and its impacts. However, it is recommended to validate the results in other areas and on finer spatial scales
whereby the influence of human activities on drought propagation and water scarcity can be analyzed. Besides this, research
would benefit from a refinement of the water scarcity dataset in order to better represent human influences on water scarcity
conditions.

405 **5 Conclusions**

Drought is expected to happen more frequently in the future, generating a range of impacts in diverse sectors. This urges the
need to develop early warning systems to mitigate the adverse consequences of drought and thereby reducing the human and
financial costs. However, there is still no full understanding of the relation between drought impacts and drought indices in
Africa. In addition, this continent struggles with water scarcity and the presence of arid regions, which possibly influences
410 the relation between drought hazard and impacts. This paper aimed to fill this knowledge gap by exploring the link between
drought impacts, drought indices, water scarcity and aridity with a focus on Kenya.

The arid region of Marsabit had the most severe drought and the highest amount of drought impacts over a timeframe from
2016 to 2020. Nyeri, classified as a sub-humid region, had lower frequencies and intensities of drought and reported the least
amount of drought impacts. This indicates that drought impacts are linked with drought severity and that the occurrence of
415 drought impacts are related to aridity. The skewed spatial distribution of drought impacts could be related to the fragile socio-
economic conditions in the ASALs of Kenya which makes this region more vulnerable to drought than the sub-humid region

of central-western Kenya. Water scarcity was not found to be related with aridity while this was expected due to the presence of unreliable water conditions. On the contrary, Marsabit (arid) did not experience any water scarcity during the analysed timeframe (March 2018 and 2020) whilst Nyeri (sub-humid) did. In addition, most drought impacts occurred at times without
420 water stress (except for Nyeri) even when increased distance from water sources was reported as a drought impact, which can be used as a proxy for water stress conditions. Reasonable explanations for this can be found in the water scarcity dataset which is apolicital, does not take reservoir into account, and is highly dependent on population density.

With a Random Forest model, a link between drought impacts and drought indices was made. The results indicated that every region, aggregated on aridity, had their own set of predictors for every impact category. Region dependency was found
425 by other studies as well. In relation to drought impacts on Pasture, anomalies were shorter (6 months) for the arid region of Marsabit than for the sub-humid regions of Narok/Nyeri (24 months). For the impacts on Livestock deaths reversed results were found: lower accumulation periods were found for Narok/Nyeri (3-12 months) while longer accumulation periods were present in Marsabit (12-24 months). Drought indices with longer timescales (>12 months), indicating a hydrological drought, were often found to match best with all the drought impact occurrences. The differences in linkages could be related to water
430 management practices, natural characteristics and climatic circumstances.

The predictive ability of indices heavily depends on the spatial and temporal resolution of drought impact data. Therefore, this study stresses the need of systematic drought impact data collection as done by the NDMA. In addition, a finer spatial resolution is needed to capture the regional differences in human influences on water scarcity and drought impacts. Studying other research areas and validating the results of this study on smaller scales will expand the knowledge base on drought and
435 impacts and will substantiate the conclusions of this study. The integration of regional predictions on drought impacts will contribute to the development of early warning systems on drought which help society to better prepare and act accordingly, therefore reducing vulnerability and increasing resilience to drought and impacts.

Appendix A: Detailed explanation of the water scarcity dataset

The water scarcity index from McNally et al. (2019) is based on outputs from the FEWS NET Land Data Assimilation Sys-
440 tem (FLDAS), which is a custom instance of the National Aeronautics and Space Administration (NASA) Land Information
System (LIS). The FLDAS's Noah 3.6 land surface model is driven by the Climate Hazards Group InfraRed Precipitation
with Station (CHIRPS) rainfall and NASA's Modern-Era Retrospective analysis for Research and Applications (MERRA-2)
meteorological forcing. This model partitions rainfall inputs into surface and subsurface runoff (i.e., baseflow), soil moisture
storage and evapotranspiration. Surface runoff is the precipitation in excess of infiltration and saturation capacity of the soil
445 while subsurface runoff is the drainage from the bottom soil moisture layer caused by gravity. The total runoff is routed though
the river network with the Hydrological Modelling and Analysis Platform version 2 (HyMAP-2) river routing scheme. The
definition of catchments are based on boundaries defined by the U.S. Geological Survey (USGS) Hydrological Derivatives
for Modelling Applications (HDMA) database. A Pfafstetter code, based on an hierarchical numbering system, are attributed
to the catchments. For the water scarcity index, Pfafstetter level 6 basins are used in order to represent the relatively local
450 nature of water supplies. Two population datasets are used as a proxy for water demand, namely the WorldPop 2015 dataset
and the European Commission's Joint Research Center's (JRC) Global Human Settlement (GHS) data. To classify the amount
of water scarcity, the Falkenmark index is used. The Falkenmark Index thresholds are specified annually while monthly data
is required for the routinely updated maps about water scarcity. Therefore, a 12-month running total of the streamflow from
the current and 11 previous months are used whereby the Falkenmark index (based on yearly values) can still be used on a
455 monthly resolution. The population estimates are aggregated to Pfafstetter basin level 6 whereafter the 12-month total spatially
aggregated streamflow (m^3) is divided by the population to produce an estimate of m^3/person (McNally et al., 2019).

(a) Marsabit (b) Nyeri

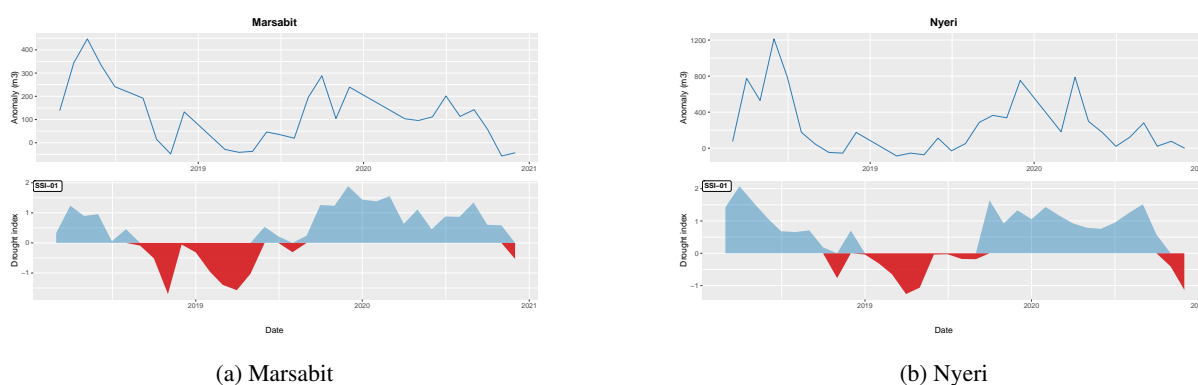



Figure A1. Streamflow anomalies (WS) and SSI-01 between March 2018 and 2020.

Appendix B: The hydrological datasets: the streamflow datasources

Different hydrological datasets were used for the water scarcity dataset and the calculation of the SSI. The SSI index is based on data from GloFAS while streamflow data for the water scarcity dataset is based on outputs from the FLDAS. If there are any discrepancies between the datasets, wrong conclusions could be made. To compare the two different datasets, SSI-01 is plot together with the streamflow anomalies of the water scarcity dataset for Marsabit and Nyeri (Figure A1). The streamflow anomalies are based on the 1982-2016 FLDAS historical record while SSI is based on the period between 1980 and 2010. Despite some irregularities between the datasets, both are following quite the same pattern. This suggest that it is reasonable to compare the results from the two different hydrological datasets.

Appendix C: A detailed explanation about the Random Forest model

Random forest (RF) is a machine learning algorithm whereby a large number of regression or classification trees on boot-strapped sub-samples of the data are constructed. Bootstrapping is a way to resample the dataset which includes replacement from the original dataset. In other words, RF's combines several randomized decision trees and aggregates their predictions by averaging. The goal of a decision tree is to create a model that predicts the value of a target by learning simple decision rules induced from the data features. A decision tree consists of nodes, edges/branches and leaf nodes. The nodes are the test for the value of a certain attribute (predictor), the edges/branches are the outcome of a test and connect to the next node or leaf and the leaf nodes are the terminal nodes predicting the outcome. In order to validate the model, a training data set and a test dataset is constructed with a proportion of 75% and 25% of the original dataset. Each tree is built on a subset of the training data set: approximately two-third of the training dataset is used for building a tree while one third is not used, called the out-of-bag error data. This generates an additional estimate of performance, namely the out-of-bag error which is a method to measure the prediction error of the random forest. For the model, two parameters needed to be tuned, namely the amount of randomized trees (ntree) and the amount of variables available for splitting at each tree node (mtry). The value of ntree has been proven to

Table A1. Tuning of parameters for the RF model: mtry values.

	Arid	Semi-arid	Sub-humid zone
	MA	BA, KI, KW	NA, NY
Pasture	3	4	4
Livestock deaths	6	4	4
Milk production	4	6	4
Food insecurity	6	2	6
Crop losses	3	4	2
Trekking distance household	6	3	4
Malnutrition	2	6	4

have not much effect on the overall accuracy of the model and is set to the default value of ntree = 500. For the best value of mtry, the function tuneRF function of the RandomForest package has been used which aims to lower the OOB error. The tuned
480 parameters for mtry are visible in Table A1.

The performance of the various models for the impact categories were tested by applying the model on the test set data (25% of the original data set). The AUC (Area Under the Curve) ROC (Receiver Operating Characteristics) curve was used to check and visualize the performance. It describes how much the model is capable of distinguishing between the classes: how higher the AUC, the better the model is at predicting 0 classes as 0 and 1 classes as 1. The ROC curve is plotted with the Sensitivity
485 on the y-axis and Specificity on the x-axis. These variables represent the true positives and true negatives respectively.

Data availability. Data is available on the 4tu.ResearchData platform. The DOI and link of access is https://doi.org/10.4121/19620357.

Author contributions. MRL has designed and conducted the research, supervised by RT, AvL, AS and AT. RO calculated, provided and have written the data on the drought indices. MRL has written the manuscript with input from all co-authors. MvdH, CW and SW have reviewed the manuscript. The final version has been approved by all co-authors.

490 *Competing interests.* The contact author has declared that neither they nor their co-authors have any competing interests.

Acknowledgements. We would like to thank Amy McNally for providing the water scarcity dataset. In addition, thanks to the National Drought Management Authority (NDMA) of the Kenyan government for providing the data for the drought impacts.

References

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

Bachmair, S., Kohn, I., and Stahl, K.: Exploring the link between drought indicators and impacts, Natural Hazards and Earth System Sciences, 15, 1381–1397, https://doi.org/10.5194/nhess-15-1381-2015, 2015.

Bachmair, S., Svensson, C., Hannaford, J., Barker, L., and Stahl, K.: A quantitative analysis to objectively appraise drought indicators and model drought impacts, Hydrology and Earth System Sciences, 20, 2589–2609, https://doi.org/10.5194/hess-20-2589-2016, 2016.

Bachmair, S., Svensson, C., Prosdocimi, I., Hannaford, J., and Stahl, K.: Developing drought impact functions for drought risk management, Natural Hazards and Earth System Sciences, 17, 1947–1960, https://doi.org/10.5194/nhess-17-1947-2017, 2017.

Bachmair, S., Tanguy, M., Hannaford, J., and Stahl, K.: How well do meteorological indicators represent agricultural and forest drought across Europe?, Environmental Research Letters, 13, 034 042, https://doi.org/10.1088/1748-9326/aaafda, 2018.

Barrett, A. B., Duivenvoorden, S., Salakpi, E. E., Muthoka, J. M., Mwangi, J., Oliver, S., and Rowhani, P.: Forecasting vegetation condition for drought early warning systems in pastoral communities in Kenya, Remote Sensing of Environment, 248, 111 886, https://doi.org/10.1016/j.rse.2020.111886, 2020.

Beck, H. E., Van Dijk, A. I., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., and De Roo, A.: MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data, Hydrology and Earth System Sciences, 21, 589–615, https://doi.org/10.5194/hess-21-589-2017, 2017.

Bellaubi, F. and Boehm, F.: Management practices and corruption risks in water service delivery in Kenya and Ghana, Water Policy, 20, 388–409, https://doi.org/10.2166/wp.2018.017, 2018.

Blauhut, V., Gudmundsson, L., and Stahl, K.: Towards pan-European drought risk maps: quantifying the link between drought indices and reported drought impacts, Environmental Research Letters, 10, 014 008, https://doi.org/10.1088/1748-9326/10/1/014008, 2015.

Breiman, L.: Random forests, Machine learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Change, I. C.: Synthesis Report. Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change, 151, 2014.

Dai, M., Huang, S., Huang, Q., Leng, G., Guo, Y., Wang, L., Fang, W., Li, P., and Zheng, X.: Assessing agricultural drought risk and its dynamic evolution characteristics, Agricultural Water Management, 231, 106 003, https://doi.org/10.1016/j.agwat.2020.106003, 2020.

EDC: Welcome to the EDII and EDR database, https://www.geo.uio.no/edc/droughtdb/, accessed: 2022-04-29, 2013.

Elreedy, D. and Atiya, A. F.: A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance, Information Sciences, 505, 32–64, https://doi.org/10.1016/j.ins.2019.07.070, 2019.

FEWS NET: Kenya food security brief. United States agency for international development (UASID) famine early warning systems network (FEWS NET), https://fews.net/sites/default/files/documents/reports/Kenya_Food%20Security_In_Brief_2013_final_0.pdf, accessed: 2022-17-05, 2013.

FEWS NET: Kenya Food Security Outlook, https://fews.net/sites/default/files/documents/reports/KE%20FSO%20Feb%20-%20Sep%202017_Final.pdf, accessed: 2022-17-05, 2017.

Hall, J. W. and Leng, G.: Can we calculate drought risk... and do we need to?, Wiley Interdisciplinary Reviews: Water, 6, e1349, https://doi.org/10.1002/wat2.1349, 2019.

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., and Pappenberger, F.: GloFAS-ERA5 operational global river discharge reanalysis 1979–present, Earth System Science Data, 12, 2043–2060, https://doi.org/10.5194/essd-12-2043-2020, 2020.

Heinrich, D. and Bailey, M.: Forecast-based Financing and Early Action for Drought–Guidance Notes for the Red Cross Red Crescent, 2020.

535   Javadinejad, S., Hannah, D., Ostad-Ali-Askari, K., Krause, S., Zalewski, M., and Boogaard, F.: The impact of future climate change and human activities on hydro-climatological drought, analysis and projections: using CMIP5 climate model simulations, Water Conservation Science and Engineering, 4, 71–88, https://doi.org/10.1007/s41101-019-00069-2, 2019.

Jenkins, M.: The impact of corruption on access to safe water and sanitation for people living in poverty, Anticorruption Resource Center, 2017.

540   Kchouk, S., Melsen, L. A., Walker, D. W., and Van Oel, P. R.: A geography of drought indices: mismatch between indicators of drought and its impacts on water and food securities, Natural Hazards and Earth System Sciences, 22, 323–344, https://doi.org/10.5194/nhess-22-323-2022, 2022.

Kew, S. F., Philip, S. Y., Hauser, M., Hobbins, M., Wanders, N., Van Oldenborgh, G. J., Van Der Wiel, K., Veldkamp, T. I., Kimutai, J., Funk, C., et al.: Impact of precipitation and increasing temperatures on drought trends in eastern Africa, Earth System Dynamics, 12, 17–35,

545   https://doi.org/10.5194/esd-12-17-2021, 2021.

Kimwatu, D. M., Mundia, C. N., and Makokha, G. O.: Developing a new socio-economic drought index for monitoring drought proliferation: a case study of Upper Ewaso Ngiro River Basin in Kenya, Environmental Monitoring and Assessment, 193, 1–22, https://doi.org/10.1007/s10661-021-08989-0, 2021.

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., et al.: High-

550   resolution mapping of the world's reservoirs and dams for sustainable river-flow management, Frontiers in Ecology and the Environment, 9, 494–502, https://doi.org/10.1890/100125, 2011.

Ma, M., Lv, J., Su, Z., Hannaford, J., Sun, H., Qu, Y., Xing, Z., Barker, L., and Wang, Y.: Linking drought indices to impacts in the Liaoning Province of China, Proceedings of the International Association of Hydrological Sciences, 383, 267–272, https://doi.org/10.5194/piahs-383-267-2020, 2020.

555   Majani, B. S., Malamud, B. D., and Millington, J.: Use of blended evidence sources to build a history of flooding impact and an impact severity scale: A case study of Nairobi, Kenya, Tech. rep., Copernicus Meetings, https://doi.org/10.5194/egusphere-egu22-12012, 2022.

Maliva, R. and Missimer, T.: Arid lands water evaluation and management, Springer Science & Business Media, 2012.

Martens, B., Miralles, D. G., Lievens, H., Van Der Schalie, R., De Jeu, R. A., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., and Verhoest, N. E.: GLEAM v3: Satellite-based land evaporation and root-zone soil moisture, Geoscientific Model Development, 10, 1903–1925,

560   https://doi.org/10.5194/gmd-10-1903-2017, 2017.

McKee, T. B., Doesken, J., and Kleist, J.: Analysis of Standardized Precipitation Index (SPI) data for drought assessment, Water, 26, 1–72, 1993.

McNally, A., Verdin, K., Harrison, L., Getirana, A., Jacob, J., Shukla, S., Arsenault, K., Peters-Lidard, C., and Verdin, J. P.: Acute water-scarcity monitoring for Africa, Water, 11, 1968, https://doi.org/10.3390/w11101968, 2019.

565   Miralles, D. G., Holmes, T., De Jeu, R., Gash, J., Meesters, A., and Dolman, A.: Global land-surface evaporation estimated from satellite-based observations, Hydrology and Earth System Sciences, 15, 453–469, https://doi.org/10.5194/hess-15-453-2011, 2011.

Mulwa, F., Li, Z., and Fangninou, F. F.: Water Scarcity in Kenya: Current Status, Challenges and Future Solutions, Open Access Library Journal, 8, 1–15, https://doi.org/10.4236/oalib.1107096, 2021.

Mutsotso, R. B., Sichangi, A. W., and Makokha, G. O.: Spatio-temporal drought characterization in Kenya from 1987 to 2016, https://doi.org/10.4236/ars.2018.72009, 2018.

Nalbantis, I.: Evaluation of a hydrological drought index, European Water, 23, 67–77, 2008.

NDMC: Drought Impact Reporter, https://droughtreporter.unl.edu/map/, accessed: 2022-04-29, 2005.

Nicolai-Shaw, N., Zscheischler, J., Hirschi, M., Gudmundsson, L., and Seneviratne, S. I.: A drought event composite analysis using satellite remote-sensing based soil moisture, Remote Sensing of Environment, 203, 216–225, https://doi.org/10.1016/j.rse.2017.06.014, 2017.

Niwattanakul, S., Singthongchai, J., Naenudorn, E., and Wanapu, S.: Using of Jaccard coefficient for keywords similarity, in: Proceedings of the international multiconference of engineers and computer scientists, vol. 1, pp. 380–384, 2013.

Nyberg, Y., Jonsson, M., Laszlo Ambjörnsson, E., Wetterlind, J., and Öborn, I.: Smallholders' awareness of adaptation and coping measures to deal with rainfall variability in Western Kenya, Agroecology and Sustainable Food Systems, 44, 1280–1308, https://doi.org/10.1080/21683565.2020.1782305, 2020.

Ondiko, J. H. and Karanja, A. M.: Spatial and Temporal Occurrence and Effects of Droughts on Crop Yields in Kenya, Open Access Library Journal, 8, 1–13, https://doi.org/10.4236/oalib.1107354, 2021.

Parsons, D. J., Rey, D., Tanguy, M., and Holman, I. P.: Regional variations in the link between drought indices and reported agricultural impacts of drought, Agricultural systems, 173, 119–129, https://doi.org/10.1016/j.agsy.2019.02.015, 2019.

Peng, J., Dadson, S., Hirpa, F., Dyer, E., Lees, T., Miralles, D. G., Vicente-Serrano, S. M., and Funk, C.: A pan-African high-resolution drought index dataset, Earth System Science Data, 12, 753–769, https://doi.org/10.5194/essd-12-753-2020, 2020.

Phillip, M. J.: Combating Water Scarcity in Southern Africa: Case Studies from Namibia, Springer, 2013.

Savelli, E., Rusca, M., Cloke, H., and Di Baldassarre, G.: Don't blame the rain: Social power and the 2015–2017 drought in Cape Town, Journal of Hydrology, 594, 125 953, https://doi.org/10.1016/j.jhydrol.2020.125953, 2021.

Savelli, E., Rusca, M., Cloke, H., and Di Baldassarre, G.: Drought and society: Scientific progress, blind spots, and future prospects, Wiley Interdisciplinary Reviews: Climate Change, p. e761, https://doi.org/10.1002/wcc.761, 2022.

Stagge, J. H., Kohn, I., Tallaksen, L. M., and Stahl, K.: Modeling drought impact occurrence based on meteorological drought indices in Europe, Journal of Hydrology, 530, 37–50, https://doi.org/10.1016/j.jhydrol.2015.09.039, 2015.

Stahl, K., Kohn, I., Blauhut, V., Urquijo, J., De Stefano, L., Acácio, V., Dias, S., Stagge, J. H., Tallaksen, L. M., Kampragou, E., et al.: Impacts of European drought events: insights from an international database of text-based reports, Natural Hazards and Earth System Sciences, 16, 801–819, https://doi.org/10.5194/nhess-16-801-2016, 2016.

Sutanto, S. J. and Van Lanen, H. A.: Catchment memory explains hydrological drought forecast performance, Scientific reports, 12, 1–11, https://doi.org/10.1038/s41598-022-06553-5, 2022.

The World Bank: Population, data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SP.POP.TOTL, 2020.

Thomas, E., Jordan, E., Linden, K., Mogesse, B., Hailu, T., Jirma, H., Thomson, P., Koehler, J., and Collins, G.: Reducing drought emergencies in the Horn of Africa, Science of the Total Environment, 727, 138 772, https://doi.org/10.1016/j.scitotenv.2020.138772, 2020.

UNESCO: Map of the world distribution of arid regions: Map at scale 1:25,000,000 with explanatory note., MAB Technical Notes 7, UNESCO, Paris, 1979.

Van Dijk, A. I., Beck, H. E., Crosbie, R. S., de Jeu, R. A., Liu, Y. Y., Podger, G. M., Timbal, B., and Viney, N. R.: The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society, Water Resources Research, 49, 1040–1057, https://doi.org/10.1002/wrcr.20123, 2013.

Van Loon, A. F.: Hydrological drought explained, Wiley Interdisciplinary Reviews: Water, 2, 359–392, https://doi.org/10.1002/wat2.1085, 2015.

Van Loon, A. F. and Van Lanen, H. A.: Making the distinction between water scarcity and drought using an observation-modeling framework, Water Resources Research, 49, 1483–1502, https://doi.org/10.1002/wrcr.20147, 2013.

610 Van Loon, A. F., Gleeson, T., Clark, J., Van Dijk, A. I., Stahl, K., Hannaford, J., Di Baldassarre, G., Teuling, A. J., Tallaksen, L. M., Uijlenhoet, R., et al.: Drought in the Anthropocene, Nature Geoscience, 9, 89–91, https://doi.org/10.1038/ngeo2646, 2016a.

Van Loon, A. F., Stahl, K., Di Baldassarre, G., Clark, J., Rangecroft, S., Wanders, N., Gleeson, T., Van Dijk, A. I., Tallaksen, L. M., Hannaford, J., et al.: Drought in a human-modified world: reframing drought definitions, understanding, and analysis approaches, Hydrology and Earth System Sciences, 20, 3631–3650, https://doi.org/10.5194/hess-20-3631-2016, 2016b.

615 Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index, Journal of climate, 23, 1696–1718, https://doi.org/10.1175/2009JCLI2909.1, 2010.

Wamucii, C. N., van Oel, P. R., Ligtenberg, A., Gathenya, J. M., and Teuling, A. J.: Land use and climate change effects on water yield from East African forested water towers, Hydrology and Earth System Sciences, 25, 5641–5665, https://doi.org/10.5194/hess-25-5641-2021, 2021.

620 Wang, W., Ertsen, M. W., Svoboda, M. D., and Hafeez, M.: Propagation of drought: from meteorological drought to agricultural and hydrological drought, https://doi.org/10.1155/2016/6547209, 2016.

Wang, Y., Lv, J., Hannaford, J., Wang, Y., Sun, H., Barker, L. J., Ma, M., Su, Z., and Eastman, M.: Linking drought indices to impacts to support drought risk assessment in Liaoning province, China, Natural Hazards and Earth System Sciences, 20, 889–906, https://doi.org/10.5194/nhess-20-889-2020, 2020.

625 Wilhite, D. A., Svoboda, M. D., and Hayes, M. J.: Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness, Water resources management, 21, 763–774, https://doi.org/10.1007/s11269-006-9076-5, 2007.

WMO: WMO Guidelines on Multi-hazard Impact-based Forecast and Warning Services, wMO-No. 1150, 2015.

Xu, H.-j., Wang, X.-p., Zhao, C.-y., Shan, S.-y., and Guo, J.: Seasonal and aridity influences on the relationships between drought indices and hydrological variables over China, Weather and Climate Extremes, 34, 100 393, https://doi.org/10.1016/j.wace.2021.100393, 2021.

630 Xu, Y., Zhang, X., Wang, X., Hao, Z., Singh, V. P., and Hao, F.: Propagation from meteorological drought to hydrological drought under the impact of human activities: A case study in northern China, Journal of Hydrology, 579, 124 147, https://doi.org/10.1016/j.jhydrol.2019.124147, 2019.

Yihdego, Y., Vaheddoost, B., and Al-Weshah, R. A.: Drought indices and indicators revisited, Arabian Journal of Geosciences, 12, 1–12, https://doi.org/10.1007/s12517-019-4237-z, 2019.