Response to reviewers

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

General response

We would like to thank the reviewers for providing a detailed and in-depth review of our manuscript. We are pleased to see the reviewers confirming that our study is novel and can be a valuable contribution to drought studies with a focus on the African continent. Based on the comments, we have made some major revisions that have led to a significant improvement of our manuscript. We hope that the revised version provides enough detail and clarity to cover the original concerns. In the following pages, a detailed point-by-point response to the comments of each reviewer is given. Our response is shown in blue. The line numbers mentioned in the responses refer to the revised version of the manuscript.

Response to RC1

General comments:

In this study, Lam et al. explore statistical links between drought impacts, drought indices, and water scarcity in five counties in Kenya from 2014 to 2020. The authors use seven impact factors retrieved from the monthly bulletins of the National Drought Management Authority in Kenya and four drought indices at various time scales. A random forest model was used to measure statistical associations between the impacts and drought indices. Their result shows the drought indices that are best linked with each of the impact factors and indicate that these associations are also influenced by the aridity level of the regions.

This study on droughts is in the scope of the NHESS journal and is a good contribution to the field of drought impacts, mainly as it is one of the few studies focusing on the African continent. However, the method and the analysis in the manuscript are not sufficient to support some of the authors' arguments, and the overall presentation of the manuscript (some figures and discussions) needs to be improved. I suggest some major revisions before considering a possible publication in the journal. Below find my comments.

- We thank the reviewer for their comments and suggestions, and we are pleased to see the reviewer confirming that our study is one of the few studies focusing on the African continent in relation to linking drought indices and drought impacts. The reviewer provides very useful comments and suggestions that are addressed in the revised manuscript. All the changes are provided below whereby our responses are shown in blue and in italics. The line numbers mentioned in the response refer to the revised version of the manuscript and are indicated within brackets [xx].

Specific comments:

- **Abstract**

It would be nice to add two or three more sentences about the results (for example, which indices were included in the study, which one is the most associated with which impact factor, etc.) to give more information to the readers. As it is presented, the abstract sounds very general, and hard to see what the significant contribution of this study is. Also, if the authors' study is the first one (or one of the few) that explores the link between drought impacts and drought indices in the African continent, that should be emphasized.

- We thank the reviewer for this remark. We agree and have changed the abstract as follows:

  [4-16] The relation between drought severity and drought impacts is complex and so far relatively unexplored in the African continent. This study assesses the relation between reported drought impacts, drought indices, water scarcity, and aridity across several counties in Kenya. The monthly bulletins of the National Drought Management Authority in Kenya provided drought impact data. A Random Forest (RF) model was used to explore which set of drought indices (Standardized Precipitation Index, Standardized Precipitation Evapotranspiration Index, Standardized Soil Moisture Index and Standardized Streamflow Index) 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 positive relation with aridity. A relation between water scarcity and aridity was not found. The RF model revealed that every region, aggregated by aridity, had their own set of predictors for every impact category. Longer timescales (≥ 12 months) and the Standardized Streamflow Index were mostly present, indicating the importance of hydrological drought to predict drought impact occurrences. 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

Line 22) "Due to the projected increase in drought frequencies": In which regions and when will this increase become significant? Note that the increases in drought frequency and intensity are not globally equal. See Chapter 11 in the new IPCC The Physical Science Basis report (Seneviratne et al., 2021) and cite this new report instead of the 2014 version.

- We thank the reviewer for this insight. We agree and have changed the citation from IPCC (2014) to the newest version, referring to Chapter 11 in the IPCC The Physical Science Basis report (Seneviratne et al., 2021). In addition, we rephrased and combined the sentences on line 22-23 (of the old version).

  [38] "Due to the projected increase in drought frequency in some regions around the world (Seneviratne et al., 2021), the probability of successive drought events might rise, resulting

in increased destabilization, insecurity and resource-based conflict in contexts with high vulnerabilities (Peng et al., 2020)."

Line 26) What do the authors mean by "a hydro-meteorological event"? "Hydrometerological events" in general or "drought" or "extreme hydrometeorological events"?

- Thank you for this accurate remark. We mean extreme hydrometeorological events and its impacts, so we added the word '*extreme*' [43].

Line 27) I am not convinced here how "Linking drought impacts to drought indices" could contribute to "the development and improvement of the M&EW .". Will it improve drought forecasting (in which way)? Help with better preparedness? Please elaborate on this sentence better.

- Thank you for the valuable comment. The sentence is adjusted to the following:

  [49] "Linking drought impacts to drought indices can contribute to drought risk and vulnerability assessments which are useful for the development of drought monitoring and early warning systems. These systems inform national and international organizations in providing timely and relevant assistance."

Line 30-43) This paragraph should go after the first paragraph (from line 22), as it introduces drought indices before discussing "Linking drought impacts to drought indices."

- We thank the reviewer for the suggestion. Accordingly, we changed the order of the paragraphs whereby drought indices are introduced before linking drought impacts to drought indices.

Line 31) In this study, as impacts on society are considered, I would also add a definition of socio-economic drought, which occurs when water as an economic good cannot meet demands due to a weather-related shortfall. For the definition, see, for example, Mishra and Sigh (2010).

- By prioritizing the addition of a more thorough overview of past studies on the link between drought impacts and drought indices (as suggested by the reviewer a few lines below) and to avoid a lengthy introduction, we have decided to shorten the paragraph about the different drought types. In fact, we believe that readers are already familiar with the definition of the different drought types. Additionally, we have provided various references in the text that could help the reader dive into the subject further (Kchouk et al., 2022; Mishra and Singh, 2010; Yihdego et al., 2019).

Line 33-35) Add a citation for this argument.

- A citation for lines 33-35 is no longer needed as we decided to shorten the paragraph about the different drought types (see explanation above).

Line 41) I suggest adding an explanation of some drought indices associated with each drought type after Yihdego et al. (2019) (at least the indices the authors use for this study). So the readers can grasp a better idea of which indices are usually used in drought studies.

- We thank the reviewer for this valuable remark, and we agree that further elaboration would benefit the manuscript. We have re-written the third paragraph (lines 30-43) and moved it to be the second paragraph. It now reads as follows:

  [25-37] Although the multifaceted nature of drought drivers, drought detection and quantification usually rely primarily on the analysis of climatic and hydrological variables (Kchouk et al., 2022; Mishra and Singh, 2010; 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 and Van Lanen, 2022). The most simple drought indices only use meteorological data while others include soil moisture or streamflow data (Yihdego et al., 2019). Meteorological and/or soil moisture (agricultural) drought are often expressed by the Standardized Precipitation Index (SPI), Standardized Precipitation Evapotranspiration Index (SPEI) and the Palmer Drought Severity Index (PDSI) (e.g., Baig et al. (2022); Kamruzzaman et al. (2022); Zhou et al. (2022)) while the Standardized Streamflow Index (SSI), the Standardized Runoff Index and the Standardized Groundwater level Index (SGI) can be applied for hydrological drought (Van Loon, 2015). However, the analysis of hydro-meteorological variables alone may not be sufficient for the identification of the actual impacts of drought as the listed drought indices do not take into account the vulnerability of the system under analysis (Bachmair et al., 2015). To better evaluate and communicate about drought risk, it is necessary to establish reliable links between drought indices and impacts.

Line 41) "Models using drought indices to forecast drought can detect climate signals": It is unclear what the authors mean by models (which models?) and climate signals. Please elaborate on this better.

- When rewriting this paragraph, it has been decided to remove this sentence to avoid unclarity.

Line 42) "Yet, the link between drought indices and socio-economic impacts has rarely been analyzed": But the authors have mentioned several papers on linking drought indices and impacts in the discussion and conclusion section (Blauhut et al., 2015; Bachmair et al., 2015, 2016, 2018, Stagge et al., 2015, Ma et al., 2020; Wang et al., 2020, among others.), right? Even in the next paragraph, the authors claimed that some works on the topic already exist and they are focused on Europe and the U.S. If these previous works had motivated the authors to perform this study, these works should be acknowledged and elaborated better in the introduction: which indices and methods were the previous studies used? What do their results say about the link between droughts and drought impacts? What is the difference between these studies and the authors' study? Also, was there any previous research analyzing the impacts of droughts over Kenya or the African continent?

- We fully agree about the importance to elaborate and acknowledge the studies mentioned. The work of previous authors is further elaborated as follows (fourth paragraph):

  [50] Several studies exist whereby drought impacts have been linked to drought indices. For instance, the qualitative dataset of the European Drought Impact Report Inventory (EDII) has been used to assess the link between drought impacts and indices at continental (Blauhut et al., 2015), national (Stagge et al., 2015), and regional scale (Bachmair et al., 2015, 2016, 2018). Linking

indices to drought impacts has been done using several methods, such as logistic or linear regression (Bachmair et al., 2018; Blauhut et al., 2015; Gudmunsson et al., 2014; Parsons et al., 2019; O'Connor et al., 2022; Stagge et al., 2015), correlation analysis (Bachmair et al., 2016;; Ma et al., 2020; Wang et al., 2020) and an ensemble regression tree approach (random forest) (Bachmair et al., 2016, 2017; Wang et al., 2020). A multitude of drought indices, mostly SPI and SPEI with accumulation periods ranging between 1 and 24 months, were linked to drought impact categories applicable for the research area, for example wildfire activity (Gudmundsson et al., 2014) and agriculture (Parsons et al., 2019). However, according to our knowledge, there are no similar studies with a focus on the Horn of Africa. Linking drought impacts with indices in that region would generate new insights, because other types of drought impact categories are more applicable such as food insecurity, livestock hunger/death/migration, diseases, and conflict (Quandt, 2021).

Line 48) What does EDII stand for?

- The abbreviation of EDII has been added, namely the European Drought Impact Report Inventory (EDII) [47].

Line 65-67) I would remove "because of." and split the sentence in two: "In this study, we focus on Kenya. The country is characterized by strong gradients in precipitation …".

- We changed the sentence as suggested by the reviewer [75].

Line 74) I am not sure what the authors mean by "arid circumstances.". Drought? Aridity index?

- With "arid circumstances" is meant the different regions in term of aridity level which changes the question to:

  [83] What is the relation of drought impacts with drought indices and with water scarcity under different aridity levels?

Line 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": For me, it is not very clear why the authors expect this. Standardizing indices standardize a water balance variable with respect to the climate of a region to facilitate the comparison across the different climates. But this does not make all regions have similar spatial and temporal variability of a drought index or respond similarly to drought drivers. Droughts can be local or regional events caused by different drivers. In addition, different surfaces, vegetation, and geographical conditions affect the spatiotemporal characteristics of droughts. Thus, even neighboring regions can present different drought conditions. For example, see the plots in Spinoni et al. (2015) which show different drought frequencies and extent (using SPI, SPEI, and RDI) in Europe.

- We thank the reviewer for these observations. We agree with the reviewer's observation that droughts can be local or regional and caused by different drivers. We also agree that, through the standardization process, the occurrence of drought events has the same average frequency across regions with different climates. In order to avoid possible uncertainty and confusion with our previous sentence, we have changed the text as follows:

[85] It is expected that drought events and impact occurrences 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). Furthermore, it is expected that water scarcity will show a relation with aridity due to the presence of unreliable water conditions.

2. Data and Methods

2.1. Study Area

Line 98) "This study will specifically focus on a region in Nyeri, namely Kieni...". This sentence is confusing. So, if I understood it correctly, inside of Nyeri, the authors will focus only on Kieni. Then, I would rephrase a sentence like, "This study specifically focuses on one region in Nyeri, namely Kieni. From now on, with Nyeri, we refer to only the Kieni region.".

If this is the case, why is only Kieni selected in the Nyeri region?

- We agree to rephrase the sentence. In addition, to avoid unclarity, we also explained why we focused our study only on the Kieni district, in Nyeri county::

  [107] This study specifically focuses on one district in the Nyeri county, namely Kieni, according to the availability of drought impact data provided by the National Drought Management Authority (NDMA). From now on, with Nyeri, we only refer to the Kieni district.

2.2. Data

Line 108) Add a citation for the MSWEP version 1.1.

- We have added a citation for the MSWEP version 2, which is according to Beck et al., 2019 [118].

Line 111) The spatial resolution of MSWEP is 0.25 degree based on Beck et al., 2017.

- We thank the reviewer for this accurate remark. We used the MSWEP version 2 according to Beck et al., 2019 with a 0.1 degree spatial resolution. The citations are adjusted accordingly.

Line 117) Add a citation for MSWEP v2.8.

- We have added a citation for the MSWEP v2.8, which is according to Beck et al., 2017a [128]. To avoid lengthy abbreviations, the European Space Agency Climate Change Initiative (ESA-CCI) soil moisture is changed to satellite-observed soil-moisture, to read:

[127] The model uses MSWEP data set (Beck et al., 2017a), satellite-observed soil moisture, reanalyis air temperature and radiation, and Vegetation optical depth (VOD) (Liu et al., 2011) to produce terrestrial evaporation and root-zone soil moisture (Martens et al., 2017).

Line 117) Add a citation for Priestley and Taylor equation.

- According to the comment of the reviewer, a citation has been added for the Priestley and Taylor equation (1972) [130].

Line 118) Add a citation for ERA5.

- According to the comment of the reviewer, a citation has been added for ERA5, namely Dee *et al.* (2011) [132].

Line 125) Add a citation for GLOFAS.

- A citation has been added for GLOFAS, according to Harrigan et al., 2020. [138].

Line 127) What do HTESSEL and LISFLOOD stand for? Also, add a citation for each of the data and models, including ERA5, if the ERA5 dataset used in this part is a different version from the previously mentioned one.

- We thank the reviewer for this accurate remark. In order to improve the ease of reading, it is decided to avoid the use of many (unnecessary) abbreviations. Therefore, HTESSEL and LISFLOOD are being removed from the text. In addition, the following citations has been added for the ECMWF ERA5 global reanalysis (Balsemo et al., 2009; Hersbach et al., 2020) and the hydrological rainfall-runoff channel routing model (Van Der Knijff et al., 2010; Hirpa et al., 2018). This reads as follows:

[139] It combines the land surface model runoff component of the ECMWF ERA5 global reanalysis (Balsamo et al., 2009; Hersbach et al., 2020) with a hydrological rainfall-runoff channel routing model (Van Der Knijff et al., 2010; Hirpa et al., 2018) (see http://www.globalfloods.eu/).

Line 128) The last sentence about LISFLOOD can be removed if this is only the input model used to generate the GLOFAS data.

- According to comment of the reviewer, the last sentence about LISFLOOD has been removed.

Line 130) I would suggest changing the subtitle to "Impact data from the National Drought Management Authority" to make it more consistent with the other subtitles.

- According to the comment of the reviewer, the subtitle is changed to “Impact data from the National Drought Management Authority” [143].

Line 138) Add a citation for the VCI Index and a brief description of this index (how is this index estimated?)

- A brief description of the VCI and citations has been added (e.g., Atzberger et al., 2014; Atzberger and Eilers, 2011; Klisch and Atzberger, 2016) to the text which reads as follows:

  [151] The director analyzes the data against the three-monthly Vegetation Condition Index (VCI3M) provided at county level and on a monthly basis. The VCI is obtained from an advanced filtering method for Moderate Resolution Imaging Spectroradiometer (MODIS) normalized difference vegetation index (NDVI) at pixel level developed and implemented by University of Natural Resources and Life Sciences (BOKU) (Kilsch and Atzberger, 2016). The MODIS NDVI data undergoes offline smoothening based on the Whittaker smoother (Atzberger and Eilers, 2011) to daily NDVI values, and near real time filtering based on available observations within the past 175 days (Atzberger et. al., 2014). Finally, the director calculates the percentage of children under five years with malnutrition using the Mid-Upper-Arm Circumference (MUAC) color codes against the long term average.

Line 138) Add a citation after the Food Consumption Index and briefly describe this (how is this index estimated?)

- A brief description with citation of the Food Consumption Score has been added, to read:

  [159] Lastly, the Food Consumption Score is computed based on food frequency and diversity based on a seven-day recall of food consumed at the household level, taking into account the relative nutrition importance of different food groups (WFP, 2008). A poor FCS means a lack of vegetable consumption every day and low consumption of protein rich food such as dairy and meat.

Line 139) Add a citation for the malnutrition index.

- More elaboration about the malnutrition index has been added, to read:

  [157] In addition, the director calculates the percentage of children under five years with malnutrition using the Mid-Upper-Arm Circumference (MUAC) color codes of the United Nations Children's Fund (UNICEF) against the long term average.

Line 140) I suggest splitting the sentence into two. For example, "This study has utilized water scarcity (WS) data from McNally et al. (2019). The data is a monthly water scarcity dataset with a spatial resolution of 0.1 degree for Africa from March 2018 to the present."

- According to the comment of the reviewer, the following has changed:

  [164] This study has utilized water scarcity (WS) data from McNally et al. (2019). The data is a monthly water scarcity dataset with a spatial resolution of 0.1 degree for Africa from March 2018 to the present.

Line 143) FLDAS is driven by the satellite-based CHIRPS precipitation. I can see that the authors mention this in the Appendix, but also add a sentence about it in this section to make it consistent with the description of other datasets.

- We thank the reviewer for this accurate remark. FLDAS is driven by the satellite-based CHIRPS precipitation, the next sentence has been added to the manuscript:

  [167] The FLDAS's Noah 3.6 land surface model is derived from 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.

CHIRPS is a different precipitation dataset and is not included in generating the MSWEP data. How would this WS based on a different set of precipitation affect your final result? This could be briefly discussed in the Discussion section.

- We thank the reviewer for this suggestion. The following has been added to the Discussion about Data sources and methods:

  [353] Furthermore, the computation of the meteorological drought indices (SPI/SPEI) and the WS dataset are based on different satellite-based precipitation products, namely MSWEP and CHIRPS respectively. The two datasets showed good performances on global level (Beck et al., 2017b) and more specifically for East Africa (Cattani et al., 2021). Although the slightly underestimation of the MSWEP data compared to CHIRPS over East Africa, both precipitation products showed considerable agreement (Cattani et al., 2021), thereby justifying the simultaneously use of both products.

Line 144) Add a citation for the Falkenmark index.

- According to the comment of the reviewer, a citation has been added for the Falkenmark index (Falkenmark et al., 1989) [170].

Caption for Table 1) Mention here that the Falkemark index is for the water scarcity level. Something like: "Falkenmark index for the water scarcity level".

- In accordance with the comment of the author, the following has been added to the caption "… *for the water scarcity level*" (p. 6).

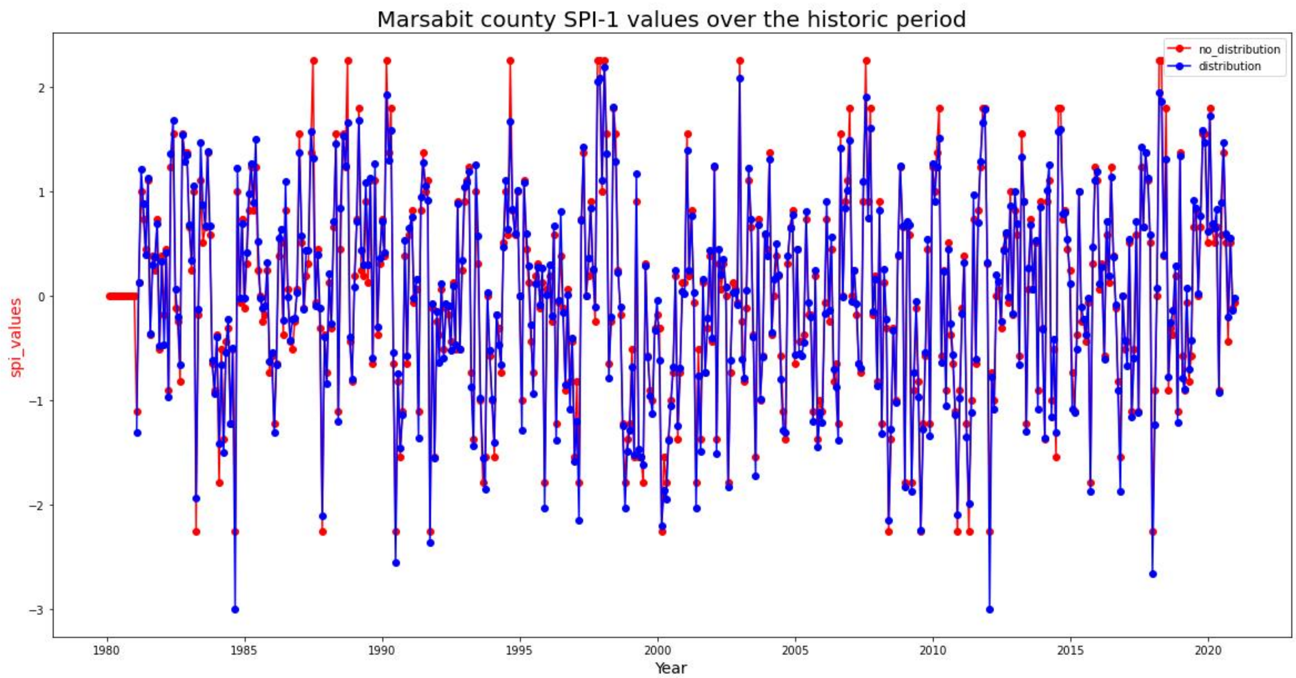
2.3. Methods

2.3.1. Drought Indices

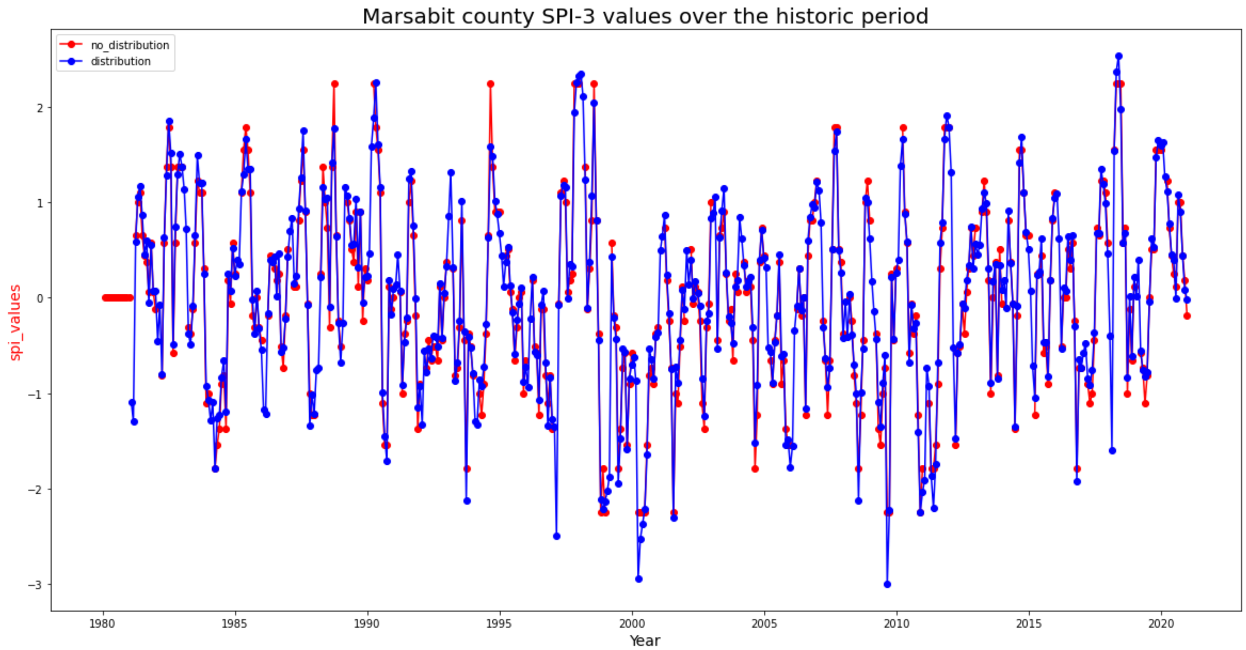
Line 158) The values are not fit through a standard normal distribution from the beginning. The cumulative precipitation is transformed to a normalized distribution by first "fitting a parametric statistical distribution" (Stagge et al., 2015a), which in general, they are a Gamma or Pearson III distribution for SPI (2015a) and a log-logistic distribution for SPEI (Vicente-Serrano et al., 2010; Beguería et al., 2014). Read Stagge et al. (2015a) and Stagge et al (2015b) to see different types of distributions for precipitation. Which type of distribution did the authors use?

- We thank the reviewer for this comment. For the calculation of the standardized indices, we ranked the precipitation values and then transformed them to a normal distribution without fitting a parametric distribution (as tested by Stagge et al., 2015, and others). This is justified in our case because the distribution is approximately normal, in which case the transformation would not add much. When we compared between SPI calculated with a statistical distribution and SPI with ranking, the results were similar; see below plots. The following has been added to the text:

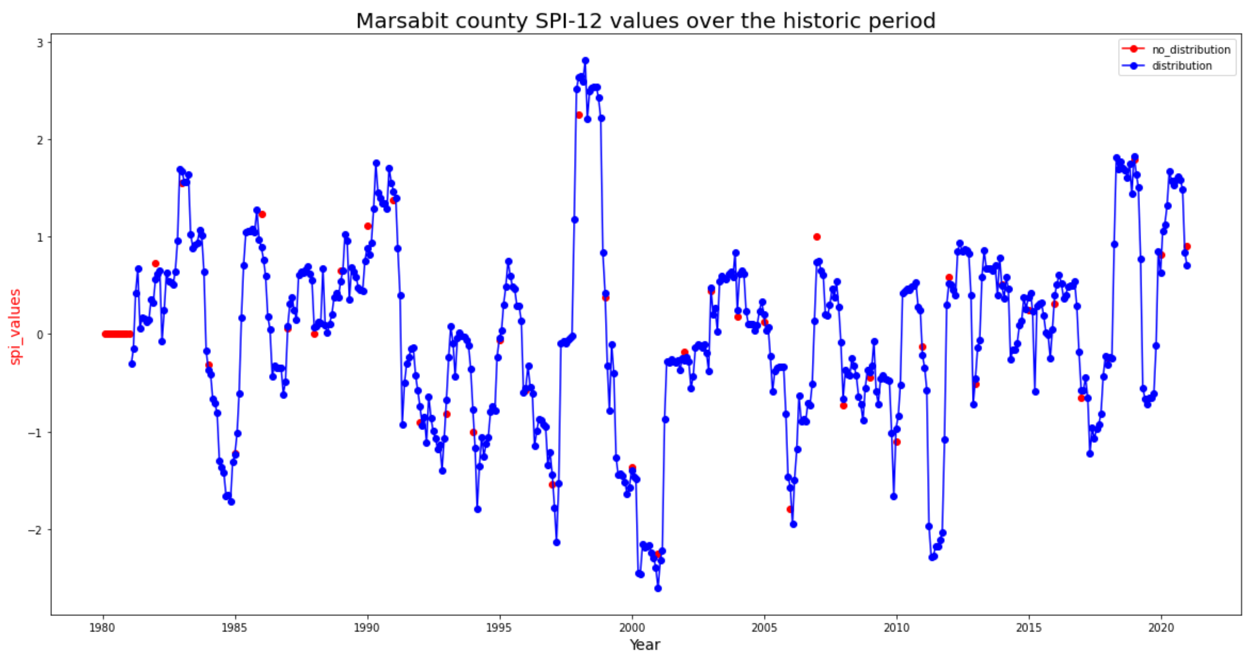
> *[183] Thereafter, the values are standardized to a normal distribution with values between -3 and 3 by ranking, so without fitting a parametric statistical distribution (as tested by Stagge et al., 2015, and others). This is justifiable in our case because the distribution is already approximately normal. When we compared the SPI calculated with a statistical distribution and SPI with ranking, the results were similar.*



a)



b)



c)

Fig a, b, c: comparison of SPI calculated with a statistical distribution (gamma) and SPI with ranking.

Line 159) Are the values area-weighted averaged? If yes, mention it.

- The values are not area weighted. Based on the resolution size and the areas of the administrative units used we assumed the average county level values would not be greatly influenced by not weighting the grid level values.

Lines 160) Is the same procedure for SPI and SPEI (cumulating monthly values, fitting a specific type of parametric distribution, then normalizing the distribution) also applied to SMI and SSMI? If yes, which parametric distributions were used for these two variables?

- For the calculation of all the standardized indices (SPI, SPEI, SSMI and SSI), the values are ranked and then transformed to a normal distribution without fitting a parametric distribution. More elaboration is given in the text [180-190] and the comment above.

Line 162) "incorporates temperature by including evaporation anomalies": It is the other way around: SPEI incorporates the effects of potential evapotranspiration, which depends strongly on the temperature. And for the calculation, "potential evapotranspiration" is included, not "evaporation anomalies". See Beguería et al. (2014).

- We agree with the comment of the reviewer and changed the text to the following:

  [191] The Standardized Precipitation Evapotranspiration Index (SPEI) is similar to SPI (Vicente-Serrano et al., 2010), but instead of precipitation it uses the difference between precipitation and potential evapotranspiration as input (Beguería et. al., 2014).

Line 166) Is GLOFAS or GloFAS? Previously, the authors used GLOFAS (Line 125). Be consistent with the acronyms.

- We thank the reviewer for this accurate remark. The abbreviation is GloFAS which is changed at line 137.

Line 179) The authors never define what they consider "drought". What is the threshold the authors set to define drought periods?

- We thank the reviewer for this comment. We consider a region in drought when the drought index is <- 0. The following has been added to the text, section 2.3.1 Drought indices:

  [186] Negative values of the drought indices indicate dryer than average conditions while positive values indicate wetter than average conditions. An area is considered in drought when the drought index is below 0.

General comments about Section 2.3.1) As the authors already mentioned in the introduction, each drought index represents different types of droughts: SPI is usually for meteorological drought that only considers the effect of precipitation, while SPEI and SMI are agricultural or soil moisture droughts, and SSI is the hydrological drought (See Yihdego et al., 2019). This difference between the indices should be clearly explained here.

- It is indeed important that the differences between the indices should be clearly explained. We consider SPEI as a meteorological index as it is calculated using climate data (precipitation minus potential evapotranspiration, PET; Vicente-Serrano et al., 2010). SPEI computed with long accumulation periods (e.g., SPEI-12) is often used as proxy for soil moisture or hydrological drought (Dai et al., 2020; Seneviratne et al., 2012; Wanders et al., 2017) but that does not make it a soil moisture drought index. More elaboration about the drought indices and drought type is given in section 2.3.1, to read:

  > *[203] In this study, SPI and SPEI represent meteorological drought, as they are based on precipitation and evapotranspiration anomalies. SSMI represents soil moisture drought, while SSI represents hydrological drought, as they are based on soil moisture and streamflow anomalies respectively (Yihdego et al., 2019). We also used SPI and SPEI with longer accumulation periods as a proxy for soil moisture and hydrological drought (Dai et al., 2020; Seneviratne et al., 2012; Wanders et al., 2017).*

2.3.2. Drought Impact data

Line 176) A citation or a link to this website should be added if this is an official website. Also, is this impact data categorized by the authors available online?

- A link to the website have been added to the text, namely https://www.ndma.go.ke/ [209]. Furthermore, data about the impacts, water scarcity, and drought indices are available online, mentioned at page 23, *data availability*. The DOI and link of access is https://doi.org/10.4121/19620357.

Lines 181-183) Could the authors explain more about how this impact data was transformed into a binary value? With binary values, it means 1 if there was an impact and 0 with no impact? What do the authors mean by normal values compared to the previous years? How are these normal values defined (based on the years with no drought?), and what is the threshold to distinguish impact and no impact? (For example, when there is 30% or 50% -or whatever percentage – of crop failures compared to the normal no-drought year?)

- The National Drought Management Authority provides monthly drought impact bulletins by county with textual information on the status of several biophysical and socio-economic indicators. Furthermore, NDMA informs about the severity of drought by categorizing the state of the monitored indicators which is in general related to the following five levels: very good, normal, moderate, severe, and extreme. We converted this information into binary data by assigning a value of 0 to very good and normal conditions and a value of 1 to moderate, severe, and extreme conditions. The following has been changed/added:

  > *[217] Furthermore, the early warning bulletins inform about the severity of drought by categorizing the state of the monitored biophysical and socio-economic indicators. This categorization is in general related to the following five levels: very good, normal, moderate, severe, or extreme conditions. We converted this information into binary data by assigning a value of 0 to very good and normal conditions and a value of 1 to moderate, severe, and extreme conditions.*

Line 181) How do the authors define drought impacts? Bachmair et al. (2016) define a drought impact as "a negative environmental, economic, or social effect experienced under drought conditions (based on EDII).".

- We thank the reviewer for this comment. Our definition of drought impacts is based on information from the National Drought Management Authority (NDMA) of Kenya. The headline of the monthly county early warning bulletins of the NDMA provides information on the drought severity through using the following classification: 'normal', 'alert', 'alarm', 'emergency' or 'recovery'. Drought severity level is defined by the County Steering Group (CSG), coordinated by the NDMA (involves NDMA county staff, key ministries such as agriculture and livestock and the county Kenya Meteorological Department, and NGOs), based on biophysical variables (SPI), vegetation status data (VCI) and socio-economic variables of food security (Mwangi et al., 2022). Only when 'alert', 'alarm' or 'emergency' were mentioned, the bulletin was considered for this analysis. In addition, as explained at the answer on the comment above, only when an indicator was signed as moderate, severe, or extreme conditions, we considered it as an impact. We have added the following to the text:

  [214] The heading of the monthly county early warning bulletins provides information on the drought phase classification, according to the following levels: 'normal', 'alert', 'alarm', 'emergency' or 'recovery'. This classification is based on biophysical variables, such as SPI and the Vegetation Condition Index (VCI), and socio-economic indicators of food security (Mwangi et al., 2022). Only the bulletins mentioning the phases 'alert', 'alarm' or 'emergency' were considered for this analysis.

  [220] This study defines a drought impact as a negative or adverse effect on economic, environmental or social level which are experienced under drought conditions (Erian et al., 2021).

Did the authors initially categorize all the impacts from NMDA, then select only drought impacts (impacts under drought conditions) for the analysis?

- We elaborated on this question in the answers given above. Only the months with relatively severe warning stages are considered for this study whereby only drought impacts under drought conditions are selected.

Line 187) What is exactly food insecurity? How is this impact factor measured, and how independent is this factor from crop losses and milk production?

- The food insecurity is measured based on deviations in the Food Consumption Score (FCS). More elaboration about the FCS is given at section 2.2.2. The inter relatedness of the different drought impact categories is measured by the Jaccard similarity method: Milk production had a statistical significant Jaccard similarity of 0.42 with Food insecurity while Crop losses had a value of 0.00*. Following the logical lag between impact categories, food insecurity could happen after crop losses and/or decreased milk production. However, a reduction in crop losses does not always lead to food insecurity, for instance when aids are put in place, crop losses or decreased milk production do not translate in food insecurity. When mentioning the considered drought impact categories, the following has been added for 'food insecurity':

[226] Food insecurity (based on the Food Consumption Score, FCS);

Line 191) Add a brief explanation of Jaccard similarity. How does this method work?

- We added further elaboration about the Jaccard similarity:

[230] The Jaccard similarity coefficient for binary values, first developed by Paul Jaccard in 1901 (Jaccard, 1912) was used to measure the similarities between the occurrence of drought impact categories (Niwattanakul et al., 2013). It measures the size of intersection of the two binary sets divided by the size of the union, the following equation is given:

$$Jaccard(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

2.3.3. Random Forest model

General comment about the section) The overall explanation of RF should be improved, and many parts of the appendix should be included in this section. For instance, partitioning the data in training and testing sets is essential information for readers to know how the RF model was built. Have a look at Bachmair et al. (2016) to see how they introduce the RF method for their work. And more technical parts related to sampling (lines 196-201) could be moved to the appendix.

- We thank the reviewer for this remark. We agree and moved many parts from the appendix to the method section. In addition, the section 2.3.3. (p. 9) is completely rewritten as: (1) an additional 10-fold cross-validation has been performed, similar to Bachmair *et al.* (2016), (2) the following performance metrices are introduced: precision, recall, the F1-score, and the "Area under the ROC curve" (AUC), (3) the 'variable importance' function (varImp) of the caret package in Rstudio (version 6.0-93) and (4) the point biserial cross correlation are introduced.

About the RF model) Usually, machine learning techniques aim to deal with large datasets. How sensitive is the output of an RF model to the number of total values or the length of each variable? Is there any metric that shows how statistically significant is the fitting of your data?

I'm asking this as I can see the Kenya dataset is only for 7 years, which is 84 months in total. Based on the Appendix, 75% of the dataset is used to feed and train the model (63 months) and the rest of the months to validate the output (11 months).

How can the authors ensure that the RF is a good tool for this dataset?

- We thank the reviewer for this comment. According to the author's knowledge there is no metric that can show how statistically significant the fitting of our data is. The RF model is sensitive for the amount of data and class imbalances, what we tried to tackle by using a synthetic minority oversampling technique (SMOTE) and randomized under-sampling (RUS). In addition, as the reviewers suggested, we conducted another method (the point-biserial correlation) to validate the results of the RF model. These two different methods showed similarities which increases the reliability of output of the RF model. More elaboration about this method and additions to the text are given below.

Could the authors repeat their analysis by using a very simple approach, such as a sort of correlation analysis (Spearman rank or Pearson correlation analysis, similar to Bachmair et al., 2016, or a simple linear regression model.) between the impact data and drought indices? This analysis takes into account the number of datasets in the significance test (by considering the degree of freedom). This additional analysis can be compared to the output of the RF, helping to see whether the result from the RF is robust enough and whether the sample size affects the final output.

- We thank the reviewer for this valuable suggestion. As suggested by the reviewer, we conducted an extra analysis, namely the point-biserial cross correlation. This is a method to measure the relationship between a binary and continuous variable, namely the drought impact data and the values of the drought indices. We included the results of this method in Appendix C 'Point-biserial correlation'(p. 23). In general, the results of the point-biserial correlation and the 'importance' function of the RF model match with each other, confirming the reliability of the RF model analysis. Furthermore, the following has been added to the text:

  [265] The AUC describes whether the model was able to predict the occurrence and non-occurrence of impacts correctly. We validated the results of the RF model by conducting a point-biserial correlation. This method measures the direction and strength of a relationship between a continuous and categorical variable (Essen and Akpan, 2018).

  [375] This study confirms this, as the performance metrics were good for several drought impact categories and all the discussed models in terms of 'variable importance' showed similarities with the results of the point-biserial correlation (Appendix C).

Line 201) Mention here more clearly that the regions are divided by their aridity indices or levels to assess the relationship between aridity levels and drought indices.

- The following sentence has been added at the beginning of the section to state the aim of the RF analysis:

  [238] A machine learning algorithm, namely the classification type of Random Forest (RF), has been used to assess the drought indices best linked to drought impacts per region with the same aridity level.

Also, what do the authors mean by aggregated? The three regions are averaged to be one input data in the RF?

- We mean with 'aggregated' that the datasets of the regions with the same aridity level were combined. So, the regions are not averaged. We changed the word 'aggregated' to '*clustered according to aridity levels*'.

Line 204) Add a citation for the ROC method and explain briefly how this is calculated.

- The following has been added in relation to the "Area under the ROC curve" (AUC):

[261] The following model performance metrics were used to see how the RF model performed on the test set; precision, recall, the F1-score, and the "Area under the ROC curve" (AUC) (Hanley and McNeil, 1982).

[265] The AUC describes whether the model was able to predict the occurrence and non-occurrence of impacts correctly.

3. Result

3.1. Drought Impacts

Line 225 - 234) The authors mainly discuss the impacts and droughts in Masarabit and Nyeri. But what about other regions? Why are the other three regions not included in the result? The plots for impacts and drought indices of other regions should also be included, at least in the appendix or supplement.'

- Thank you for this remark. We made a choice for Marsabit and Nyeri because of their distinct characteristics in aridity levels and climate. The figures only serve illustrative purposes, therefore plots for the other counties are not included in the manuscript but we provided those figures in the supplement (Figures S1-S4), as the reviewer suggested. At line [271] the words '*to illustrate'* have been added and the reason to visualize those counties have been added:

  [281] We choose to visualize the drought impacts and SPEI time series of those counties because of their contrasting aridity levels.

  [282] Similar figures for the other counties are included in the Supplement (Figures S1-S4).

Line 210) "Drought impacts": The same as the comment for line 181. Are these the total impacts during 2014-2020, or only the impacts during droughts? How do the authors define a drought impact? As I can see from Fig 1., some impacts occur during the non-drought period in SPEI (for example, September 2020).

- We consider the drought impacts from the county early warning bulletins of the NDMA when the phase classification is 'alert', 'alarm' and 'emergency' (more elaboration on the answers of the comments on page 12). This phase classification is based on biophysical variables (such as SPI and the VCI) and socio-economic indicators in relation to drought. Therefore, this drought definition of the NDMA does not necessarily have to match with our considered drought indices and accumulation periods.

Table 2) The same comment as above. Are these "drought impacts" or the total impacts?

- Table 2 shows the total amount of reported drought impacts between 2016-2020 and the percentage of drought impact category involved per county.

And I would suggest replacing "Count" with "Number of impacts".

- As suggested by the reviewer, "Count" is replaced with "Number of impacts".

Line 219) Replace "A timeline" with a "time series".

- As suggested by the reviewer, “a timeline” is replaced by “a time series” *[280].*

Figure 2 (and also for other figures)):

- Enlarge the texts and labels, please. They are challenging to read. Also, enumerate each subplot (with a. b. c. d.) to indicate the part the authors want to discuss clearly.

- The plots for impacts (the upper two panels) are complicated to read and distinguish the years. I wonder if there is a better way to visualize these plots: for example, by transforming them to line or bar plots, or somehow finding a way to make the font bigger and add a vertical line to divide the years.

- We thank the reviewer for this remark. We adjusted the plots: the lines connected to the drought impacts are dotted grey and the intervals for the months are bigger to increase readability. The plots belong to each other as one figure because of the vertical alignment of the years/months, therefore the plots will not be enumerated (so a. b. c. d.) but will remain a. and b.

- In the caption: "time series" instead of "timeline".

- As suggested by the reviewer, “a timeline” is replaced by “a time series” at the caption of Figure 2.

Line 226) The same comment as line 210: Is this the number of total impacts or drought impacts?

- We mean the drought impacts reported during January 2016 and December 2020, as shown in Figure 2.

Lines 235-138) What is the authors' criterion to claim "a bit of relation" and "no relation"? Why is 0.63 slightly similar but 0.4 not? Also, add a significance level of the Jaccard similarity.

- We thank the reviewer for this comment. Unfortunately, there is no standard significance level of the Jaccard similarity. Only closer to 1 means that the datasets are more similar than each other than closer to 0. We rephrased these lines, so that we avoid the terms “a bit of relation” and “no relation”.

> *[298] The results are shown in Table 3 whereby closer to 1 means that the datasets are more similar to each other than closer to 0. Pasture and Milk production have the highest significant Jaccard similarity of 0.63 while Crop losses are not much related to any other impact category (<0.20). Trekking distance to water points have a significant Jaccard similarity of 0.50 with Pasture and 0.47 with Milk production. Other relations have a Jaccard similarity below 0.40.*

Table 3) How statistically significant are these correlations? One can add asterisks next to the values when they are statistically significant.

- We thank the reviewer for this suggestion. Statistical hypothesis testing using the Jaccard similarity coefficient has seldom been done (Chung et al., 2019). However, Chung et al., 2019 have implemented their proposed methods in an R package called ‘Jaccard’ (https://cran.rproject.org/package=jaccard). This package is used to determine the significance of the correlation whereby p values lower than 0.05 are considered statistically significant. Asterisks have been added to table 3. The following has been added to the caption of Table 3:

> *(p. 12) Correlation between the impact categories (Jaccard similarity): the asterisks indicate the statistical significance (p value < 0.05) in accordance with (Chung et al., 2019).*

Line 230) What do the authors mean by "a direct relation"? Could you clarify it?

- We mean with "a direct relationship" that the drought impact occurrence and drought event happen simultaneously (one on one). So, most of the impacts occur within the timeframe of the drought event. The following has been changed to improve the clarity:

> *[295] In general, drought impact occurrence does not happen simultaneously with the drought time period of the other accumulation periods, except for SPEI-24.*

Line 231) What and when is the onset of drought? This question is also a continuation of the comment in line 179: To mention the onset of drought, what the authors define as drought periods should be clarified first.

- We have elaborated on this question at the answer on the comment above (line 179, p. 12).

Overall comment about the section) I'm quite skeptical about including the indices with 2 years time scale (24 months) in this study. A higher time scale index is appropriate to represent relatively long dry and wet conditions. Even though these indices have a monthly time scale, one monthly value of the index represents two-year hydroclimate condition (Hence, also showing a strong persistence.).

The Kenya dataset is too short (7 years) to include these indices for the analysis. If one drought occurs in one month in this index, this indicates a dry period of almost 30% (2 years) of the study period. The authors can also see this in the time series of SPEI-24 for Masarabit: more than 50-60% of the time series shows a drought condition covering a large portion of the impacts. The plot basically indicates that there is only one long drought event for the analysis, and I don't think one can derive a significant statistical association from this time series.

Another study, for instance, Bahmair et al. (2016) can include this long-time scale index, as their dataset starts in 1970.

- We thank the reviewer for this consideration. The goal of section 3.1 is to visualize/illustrate the relationship between drought impacts and SPEI with different accumulation periods for the time period 2016-2020. For SPEI-24, this results into one long drought event, covering around 50% or more of the drought impacts in the timeline 2016-2020. However, the RF analysis considers a bigger time period, namely July 2013 to 2020 whereby multiple drought events occur according to SPEI-24 dependent on the considered county (see the Figures below). It is good to note that the computation of SPEI-24 is based on a time period 1980 to 2020 and not only over the 7 years for which impact data are available. In addition, with our analysis of the RF model, we try to predict when an impact occurred **per time step** whereby drought is defined differently in accordance with the different drought indices and accumulation periods. In this case, we do not define an event: we predict the (non-)occurrence of drought impacts based on drought indices with different accumulation periods (so different drought types). The classification tree predicts the occurrence of drought events (1 or 0) based on numeric values of the drought indices with different accumulation periods. So, indeed as the reviewer suggested, the statistical model would have huge uncertainty if there is only one event to predict drought impacts. However, this is not the case with the RF method. Also, even with drought covering 50% or more of the period for which we have impact data, it is still possible to have a clear relation. This possibility should not a priori be dismissed. Therefore, we decided to still include the drought index SPEI-24 in our study. The 'variable importance' function of the caret package in Rstudio (version 6.0-93) will show how important a certain drought index with accumulation period is for the model to make accurate predictions.

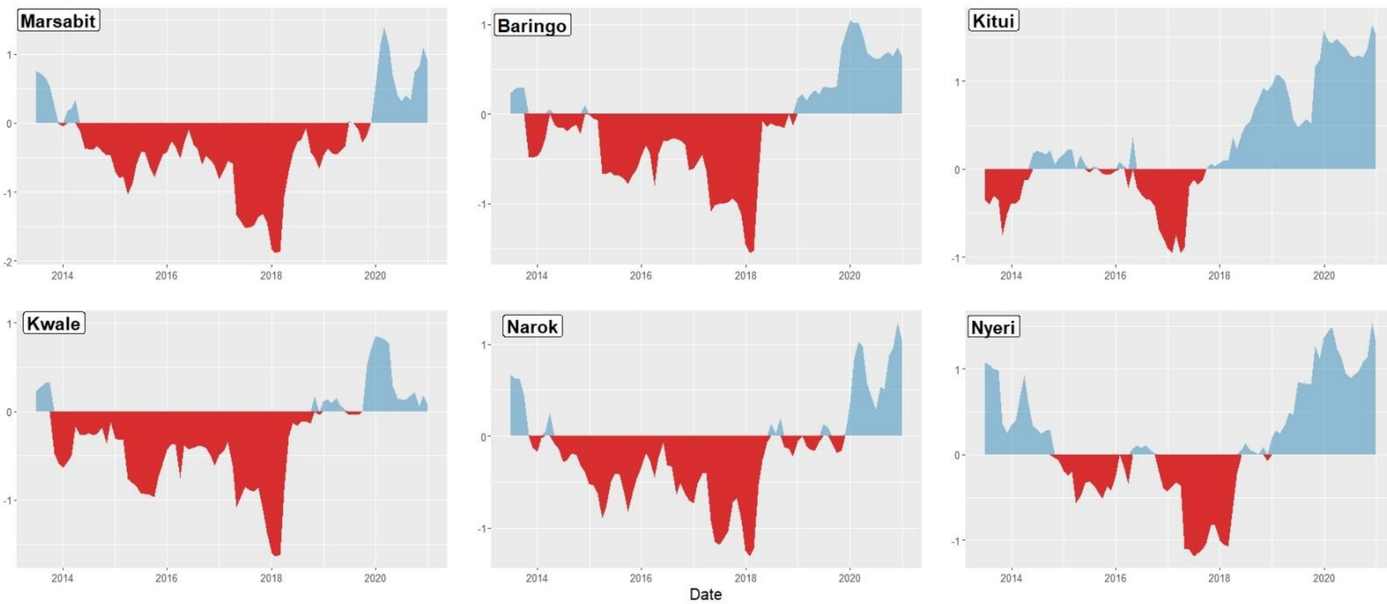


Figure 1. Time series of SPEI-24 between July 2013 and December 2020 for the counties: Marsabit, Baringo, Kitui, Kwale, Narok and Nyeri. Drought (SPEI-24 below 0) covers 73.6% of the period for Marsabit, 64.8% for Baringo, 36.3% for Kitui, 71.4% for Kwale, 70.3% for Narok and 40.7% for Nyeri.

3.2. Drought impacts and water scarcity.

Line 240) Start the paragraph mentioning that the association between WS and droughts is done from 2018 due to the length of the WS dataset.

- In accordance with the suggestion of the reviewer, the following is changed:

[304] This is done from 2018 due to the length of the WS dataset.

Line 241) Remove "(start of timeframe WS dataset)".

- As suggested by the reviewer, the "start of timeframe WS dataset" has been removed.

Figure 3) The same for all figures: Enlarge the text and labels in the plots. And I suggest putting the enumeration and the subtitles of the plots at the top and not at the bottom of the figures, as they are mixed with the captions.

- We thank the reviewer for the accurate remark. We agree and have enlarged the text and labels in the plot to increase readability. The enumeration and the subtitles of the plot are placed on top. The enumeration and the subtitles of the plot are below the plot still.

Fig.3a) Suggest replacing “amount of months” with "number of months in a year"

- As suggested, the y-axis title is changed from “amount of months” to “number of months in a year”.

General comment about the section) How is water scarcity associated with drought conditions? Have the authors made some comparisons between water scarcity and drought indices? Whether water scarcity always occurs during droughts or not should be analyzed and included in the section.

I am asking this as in 2020, there is no drought in Masarabit and Nyeri (Fig. 2), and many regions are under no water stress during 2018-2020 (Fig 3.b). Can it be that this no-water stress situation occurs mainly in 2020 (during no-drought condition)?

- We thank the reviewer for this suggestion. However, this is out of the scope of this research since the focus of the paper is to understand the link between drought impacts and drought indices and the analysis of water scarcity conditions have been carried out only within this context, to better understand the severity of the drought impacts and their spatial extent. For instance, the drought impact ‘trekking distance of water for households’ can be used as a proxy for water scarcity conditions while the origin of the impacts is different. This is elaborated in section 4.3 of the Discussion (lines 396-405). Therefore, the occurrence of water scarcity during different drought types has not been analyzed. As shown in Figure 3.a., 2020 is mostly characterized by ‘no stress’ conditions, except for Nyeri which has some months with ‘stress’ conditions. This implies that water scarcity does not necessarily have to occur during drought conditions.

And do all the impacts during 2018-2020 occur under droughts, or are these the total impacts during 2018-2020 (including non-drought period)?

- The impacts during 2018-2020 are selected based on the warning stages of the NDMA early warning bulletins, as explained above. Therefore, the drought impacts do not necessarily have to occur during drought conditions.

Line 246) The same comment about the "drought impact” in line 181.

- The definition of drought impacts has been elaborated and discussed in section *2.3.2 Drought impact data* as suggested earlier by the reviewer.

3.3. Drought Indices and Random Forest

General comment about the section) For this section, the question is again, why do the authors show and discuss the results for only some regions? In addition, in Fig.4 only some specific impacts are discussed for a few regions and not all. Why is that? If some regions do not show significant connections with impacts, they should still be mentioned and included in the plot. If including everything enlarge the plots, the authors can think about putting dots in different colors for different impact categories.

- Thank you for this observation. We have only discussed the models that performed relatively good and are in line with the outcomes of the point-biserial correlation. To make this clear, an extra column has been added to Table 4 to show which models will be discussed in terms of variable importance and which not, and we added the following to the text:

  [327] We will only discuss the variable importance metrics of the RF models (1) that performed relatively well on the test dataset because it is important that the model can predict drought impacts based on unseen data (not used for training of the model) and (2) whereby the results were similar with the results of the point-biserial correlation (Appendix C). The last column of Table 4 indicates if the model is discussed in terms of variable importance.

Line 250) I suggest changing the title to something similar to "association between drought impacts and drought indices" or "Random Forest to link drought impacts and drought indices" instead of the name of the method the authors use.

- As suggested by the reviewer, we changed the title of section 3.3. from "The Random Forest model" to "Random Forest to link drought impacts and drought indices".

Line 251) Mention here again that the regions are aggregated by their aridity levels or indices to make clear that the analysis in this part also considers the aridity levels of different counties.

- As suggested by the reviewer, we mention that the regions are aggregated by their aridity levels, the following has been added:

  [315] The performance of the Random forest (RF) models per impact category is shown in Table 4. The regions are aggregated by their aridity levels: Marsabit is classified as arid while Baringo/Kitui/Kwale are semi-arid and Narok/Nyeri are sub-humid regions.

Table 4) How is this performance measured? Clarify it in the caption.

- After conducting the additional 10 fold cross validation to increase the reliability of the RF results, we have decided to include multiple performance metrices. Namely, a performance metric for the train dataset (75%) whereby class imbalances has been dealt with by minority oversampling (SMOTE) and randomized under-sampling (RUS) techniques. Therefore, the OOB error rate (already in the RF model and explained at section 2.3.3.) and the accuracy are suitable performance metrices. However, the test data set has class imbalances whereby precision, recall and the F1 score are better suited as performance metrics beside the AUC. The caption of Table 4 is changed to:

(p. 14, Table 4) RF performance metrics: the performance of the RF model is tested by looking at the OOB and the accuracy while precision, recall, the F1-score and the AUC are computed as performance metrices for the performance of the RF model on the test data set (25%).

Line 262) What do the authors mean by "best linked"? How is the threshold for the best and worst defined? Please clarify.

- With “best linked”, we mean that those drought indices are very important for the RF model to make accurate predictions about the (non-)occurrence of drought impacts. This top 5 best linked drought indices are computed with the varImp function of the Caret package in Rstudio.

Line 265) The authors never mentioned what the MDA and Ginni Index are. Add a brief description of these scores and how they are estimated in the Method section. Some explanation of these scores is in the appendix, but this information should also be included in the Method.

- After consideration, we used the ‘variable importance’ (varImp) function of the caret package in Rstudio (version 6.0-93). This has the advantage that only one plot is needed (not two with the MDA and Ginni Index) which increase the clarity of the manuscript. We explain the ‘varImp’ function in section 2.3.3., to read:

[253] The ‘variable importance’ function (varImp) of this package was used to determine the importance of a predictor variable for the model to make accurate predictions. Specifically, the prediction accuracy on the OOB data is recorded for each tree, which is also done after permuting each predictor variable. The difference in accuracy between the two models is then averaged over all trees, and normalized by the standard error (Kuhn, 2008; Liaw et al., 2002).

Fig 3 and 4) Enumerate each of the plots to distinguish the Ginni and MDA plots.

- This is no longer needed as the importance plots are now one plot per drought impact category and region instead of two (see p. 15/16).

Line 270) SPEI is clearly not a meteorological index. The authors also mentioned in the introduction that meteorological drought considers only a precipitation deficit.

- See our explanation at the top of page 13.

4. Discussion

4.1. Data sources and methods

Line 290) How many monthly gaps does this report have (how many percentages of the total years)? This should be mentioned in the data section.

- The NDMA bulletins are structured per county whereby all the data available has been used until 2020: the earliest data that was available is July 2013, so that is the starting date of analysis. This is changed at several places in the text, namely line [202] and [210]. The drought impact data had some missing months, namely between 4 and 8 months, different per county, and mainly at the beginning of the period (between July 2013 and 2014). Table 1 shows the percentage of months with drought impact data of the NDMA for the period July 2013 and December 2020. For every county, more than 90% of data is available. For the first part of the analysis, only the years 2016 to 2020 are visualized in which there are almost no missing months. For the random forest analysis applies 'the more data available the better' as the RF model is not working with a timeline (the dates) of drought impacts. We added the following at the Discussion section:

> *[358] This data source had some missing months, namely between 4 and 8 months, different per county, and mainly at the beginning of the period (between July 2013 and December 2014). Despite these missing months, there is still a robust and reliable timeline of drought impact data available for more than 90% of the considered months.*

Table 1. Percentage of months with drought impact data of the NDMA for the period July 2013 and December 2020.

County	*Percentage of months with drought impact data*
Marsabit	*92.2%*
Baringo	*91.1%*
Kitui	*94.4%*
Kwale	*95.6%*
Narok	*91.1%*
Nyeri	*92.2%*

Line 292) What do the authors mean by iterative processes? Which kind of iterative process is applied to deal with the missing months? Please include this in the Method section.

- We thank the reviewer for this comment and we agree that the sentence might be confusing. We meant that NDMA bulletins are checked multiple times to ensure a reliable list of drought impacts. Therefore, we rephrased the sentence to the following:

> *[361] In addition, the bulletins were iteratively checked by several NDMA employees to ensure a reliable list of drought impact data.*

After all, how would these missing gaps affect the result of this analysis?

- For the first part of the analysis, only the years 2016 to 2020 are visualized in which there are almost no missing months. For the random forest analysis applies 'the more data available the better' as the RF model is not working with a timeline (the dates) of drought impacts. Therefore, as well as the low number of missing months in comparison with the timeline, we do not expect that the missing months would have substantial effect on the results of this analysis.

Line 295) What does DIR stand for? Also, add a citation for EDII and DIR.

- We thank the reviewer for this comment. We included the abbreviation of the DIR already in the Introduction section [47], same as for EDII [47]. Citations have been added to both abbreviations, namely EDC (2013) and NDMC (2005).

Line 299) What does EM-DAT stand for? Add a citation for EM-DAT and DesInventar.

- EM-DAT stands for the Emergency Events Database (Guha-Sapir et al., 2016) and DesInventar for the Disaster Inventory System (https//www.desinventar.net/) which is added to the text at lines [367-369].

Line 305- 306) I am not sure whether the AUC supports this statement. The authors are only using one method here. How can one confirm the output of the model if this is not compared to other methods? Also, how sensitive is an RF model to the number of samples? See my comment for Section 2.3.3.

- We thank the reviewer for this observation. In accordance with the comment of the reviewers, an additional correlation analysis has been conducted, namely the point-biserial correlation. This method shows similar results as the RF analysis. The following have been added to the discussion:

  > *[375] This study confirms this, as the performance metrics were good for several drought impact categories and all the discussed models in terms of 'variable importance' showed similarities with the point-biserial correlation (Appendix C).*

4.2. Relations with aridity

Line 310) I'm not sure if the result partially supports the first sentence. All the activities included here are mainly pastoral and not other activities, right? Then how do the authors know the impacts on other industrial or economic sectors?

- We thank the reviewer for this remark. The considered drought impact categories are indeed mainly related to pastoral and agricultural activities. Therefore, we cannot conclude what the highest share of all the drought impact categories are. However, we can say that the considered drought impact categories are linked to the main livelihood activity of the county as this is known (FEWS NET, 2013). We decided to phrase this part differently, to read:

  > *[366] The majority of the drought impact data are livestock- and pasture-related which is in accordance with the main livelihood activity of the considered counties and with the information provided by the NDMA.*

Line 320) What do the authors mean by the onset? The authors never defined when an actual drought starts and when is the onset of these droughts. See my comments for line 179.

- We have added our definition of drought at line 188, see our answer on the comment for line 179.

Line 322) After evapotranspiration, add some citations referring to this argument (for example, Seneviratne et al., 2021).

- We have added the following citations to the text, namely Seneviratne et al., 2021 and Wang et al., 2022. [391].

Line 322) "However, this study cannot link drought occurrence to aridity.". This sentence is not very clear to me what the authors want to say. Aridity is a constant arid condition. Do you mean the aridity level of the regions, constant desertification due to climate warming, or just the impacts of drought frequencies related to global warming? Elaborate on the sentence, please.

- We thank the reviewer for this notification. We mean that this study cannot link the frequency and intensity of drought occurrence with different aridity levels. The following has been added to improve the clarity:

  [378] However, this study cannot link the frequency and intensity of drought events with different aridity levels because of the short timeframe (approx. 7 years) analyzed.

Line 322) "short time period (10 years)". Why 10 years, if based on what I understood, the authors are taking the study period from 2014 to 2020? (7 years, including 2020).

- We thank the reviewer for this accurate remark. We changed the amount of years to approximately 7 years [392].

Line 323) "interannual trend": Do the authors want to say "long-term" trend? The interannual trend does not make sense for a longer time scale analysis that the authors claim to be necessary.

- The answer on the comment below elaborates on this comment.

Line 323-325) Overall, these two sentences are pretty difficult to understand what the authors want to say. How would a longer time series help to understand drought events and drying climate? Do you mean a long-term drying trend and a change in the climate? How is this connected to aridity and droughts? Please clarify this part.

- We thank the reviewer for these comments. Due to (anthropogenic) climate change, it is projected that there will be an increase in drought events, especially the ones related with increased PET (Seneviratne, 2020; Wang et al., 2022). However, long-term changes in PET and precipitation could indicate a drying climate, resulting in more arid regions (Xu et al., 2021). We have rephrased the sentence, to read:

  [392] The analysis of longer time series could not only indicate if there are changes in drought severity, area, and frequency but also if there is a long-term shift to a more arid climate (Xu et al., 2021).

4.3. Water scarcity and drought impacts

Line 329-331) “Increased distance from water sources was reported in arid and semi-arid regions during most of the months when meteorological and hydrological drought conditions occurred (Figure 3b)”: I don't see this result in Fig. 3b.

- We thank the reviewer for this accurate remark. Indeed, this is not visible in Figure 3.b. as it is about the reported drought impacts from the NDMA, so visible in Figure 2 (Trekking distance of water for households). We changed ‘(Figure 3b)’ to ‘(Figure 2)’. In addition, the end of this paragraph is slightly changed to elaborate on the discrepancies between water scarcity and the drought impact category ‘trekking distance of water for households’, to read:

    [399] Increased distance from water sources was reported in the arid (Marsabit) and sub-humid (Nyeri) region during most of the months when meteorological and hydrological drought conditions occurred (Figure 2). Noticeable is that Marsabit has more reported drought impacts on trekking distance for water (16.9%) than Nyeri (9.1 %) while Nyeri has more months with water scarcity than Marsabit which has zero months with water scarcity (Figure 3a.)

Line 346-347) I cannot see any statistical relationship between WS and droughts as the authors haven’t compared the two datasets in the Result section (see the comment for Section 3.2). In addition, for me, 2020 looks like an almost drought-free period in Masarabit and Nyeri (Fig 2.)

- We thank the reviewer for this remark. However, as Marsabit (the arid region) did not have any water scarcity issues during the period 2018-2020 (Figure 3a.) while drought occurred, it can be said that the WS dataset suggests that water resources were sufficient to meet the water demand. The same applies for Kitui which is a semi-arid region. In addition, our analysis indicates that the water scarcity dataset in general does not show simultaneous occurrence of water scarcity with any drought impact category (Figure 3b.), except for Nyeri. This suggest that water scarcity (based on our WS dataset) is not necessarily related to drought impacts. However, this is of course based on the data and assumptions behind the WS dataset which maybe should be improved, as we suggest at the Discussion in lines 483-485. We slightly changed the sentence to:

    [418] The WS dataset suggests …

4.4. Drought indices and the Random Forest model

Line 358) See my comment about the drought indices with a 24-month time scale in Section 3.1. SPEI-24 is not appropriate for this analysis with a 7-year-long time series to derive a robust statistical association.

- We thank the reviewer for this comment. See for our response our answer on p. 19.

Line 361) “These differences between the best match between drought impacts and drought indices implies therefore a link with human activities”: The argument here is not very convincing for me. How are the authors’ results on the best and worst indices related to human activities and the propagation of droughts? Please elaborate on this argument better.

- The difference in best match in relation to the accumulation periods and the drought index can indicate the presence of for instance water buffers or other related water management techniques, lagging the occurrence of the drought impacts. We added a part to the sentence, to read:

> *[433] These differences in best match between drought impacts and drought indices imply a link with human activities as they can lag the moment of impact occurrence.*

Line 363) Again, SPEI is not a meteorological drought. It is an agricultural or soil moisture drought, and when the time scale becomes longer, it also shows memories that are comparable with the soil moisture anomalies.

- We thank the reviewer for this comment. However, we consider SPEI as a meteorological indicator related to meteorological drought as well, since it is purely based on meteorological measurements and not on actual soil moisture status. We have given more explanation at p. 13.

Line 367) "Noticeable is that SSI gives a possible link with water-dependent activities while SSMI shows a possible link with agricultural practices.": Isn't this expected as SSI represents hydrological drought and SSMI the soil moisture drought? Clarify the division between different drought indices in the Method section (See my comment for Section 2.3.1).

- We thank the reviewer for this valuable remark. As suggested earlier by the reviewer, we elaborated about the drought indices and drought types at the end of section 2.3.1.

Line 369) Again reiterating. SPEI is clearly not a meteorological drought.

- See our answer above.

Line 371-372) Note that Dai et al. 2020 use only SPI which is a meteorological drought index. They do not include SPEI in their analysis.

- SPEI computed with long accumulation periods (e.g., SPEI-12) is often used as proxy for soil moisture or hydrological drought (Seneviratne et al., 2012; Wanders et al., 2017). Therefore, the time length and duration of both SPI and SPEI can be used to express soil moisture and hydrological drought.

Line 375) Again, for this short dataset of 7 years, I would not include an index with a time scale of 2 years for statistical analysis.

- We thank the reviewer for this comment. However, we elaborate about our decision to include an index with an accumulation period of 2 years, but a monthly time scale for the analysis, at p. 19.

Line 378-381) Could this difference in the result between the authors' work and Bachmair et al. (2016) in Europe be related to each country's economic and resilience capacity to droughts? It would be worth discussing how the economic condition of each country affects drought impacts and the result of this study.

- We thank the reviewer for this suggestion. It would be a very interesting analysis to discuss how the economic condition and resilience capacity of each country affects drought impacts. However, for now, this is too broad for the scope of this research, but it would be another very interesting follow-up research. We have added the following sentence to the text:

  [471] This could as well explain the differences in the optimal drought indices found among the researched areas as the level of preparedness can lag or mitigate the occurrence of drought impacts, resulting in drought indices with higher accumulation periods.

Line 386) The same comments about the indices with a 24-month time scale (line 375).

- See the answer above.

Line 387) If the authors claim that long SPI and SPEI are the hydrological droughts, the hydrological index (SSI) at various time scales must also show a significant association to those impacts associated with long SPI and SPEI. The association of SSI should be discussed in more detail here and in the Result section.

- We thank the reviewer for this accurate remark. We find it important to note that SPI and SPEI with large accumulation periods (>12) represent a lag and are therefore could be related to the hydrological index SSI. This is the same for soil moisture drought whereby accumulation periods of 3-6 months of SPI or SPEI are often found to correspond to a similar lag as found in soil moisture. This time lag and duration of SPI and SPEI is elaborated in lines [442-446]. We will not discuss this in more detail in the Result section as the result section only describes and not interprets the results. More information about this relationship is already given in the Discussion section [446-447].

Line 396) I can imagine that the adaptation measures will increase the resilience of smallholders. However, the sentence does not clearly explain how these measures will affect the result of this study.

- We thank the reviewer for this comment and we agree that more clarity is needed. This comment is in line with an answer on p. 25. In response to that question we suggested that differences in resilience and adaptation levels per country could be related to differences in optimal drought indices found. Therefore, the following sentence was added:

  [471] This could as well explain the differences in the optimal drought indices found among the researched areas as the level of preparedness can lag or mitigate the occurrence of drought impacts, resulting in a better fit with drought indices with higher accumulation periods.

Line 396-398) Difficult to understand. I suggest splitting the sentence into two.

- As suggested by the reviewer, the sentence is split into two, to read:

  [478] Our results show the best drought index for a given impact. This 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).

Overall comments) Also, the effects and drawbacks of short time series and the RF method should be discussed briefly in this section.

- We thank the reviewer for this suggestion. The RF model is sensitive to data availability which could explain the deviations in performance rates. We added the following on the Discussion – Data sources and methods (section 4.1):

  [377] However, there were differences in the predictive power of the RF model among the drought impact categories and the regions. This could be related to (1) data availability (e.g., data on impacts related to malnutrition) as the RF model is sensitive to data availability (Bachmair et al., 2016) and (2) deviations in the link between drought impacts and indices among the counties, making the model perform worse when the counties are combined together (i.e., the models in relation to the semi-arid regions).

5. Conclusion

Line 406) The same as the comment for line 22. Where and when?

- After consideration, we decided to remove this sentence.

Line 412) Again, why is the discussion mostly about Masarabit and Nyeri?

- Marsabit and Nyeri are highlighted because of their contrasting aridity levels and climate.

Line 418) What do the authors mean by “unreliable water condition”? Please clarify.

- The unreliable water conditions are present in the arid regions because of its dry climate. The following has been added to the text:

  [497] Water scarcity as derived by the WS dataset was not found to be related with aridity while this was expected because arid regions are often facing limited water resources.

Line 419) I can see that in 2020, not many drought months occurred. Again, are all the impacts counted during the drought conditions or for the entire period?

- The drought impacts do not need to happen simultaneously with drought months (from our analysis) as the drought impacts are gathered from the NDMA early warning bulletins which work with warning stages and are based on more variables rather than only drought indices.

Line 426) Same comments for the drought indices with a 24-month time scale.

- More elaboration about our choice to include the 24-month SPEI with a monthly time scale can be found on p. 19.

Line 432-433) Do the authors mean over the entire globe or in Africa? Clarify it.

- We mean ‘around the world’, the following has been added:

[512] Therefore, this study stresses the need of systematic drought impact data collection around the world following the example of the NDMA in Kenya.

Line 433-434) What about the temporal resolution of the dataset?

- We thank the reviewer for this thoughtful remark. However, monthly impact data is already a useful temporal resolution to work with. Therefore, we want to stress that smaller spatial scales, as regions differ very much, would contribute more to capture the differences in human influences.

Line 434) What do the authors mean by “other research areas”?

- We mean to compare the results of our study with other similar research areas . This will expand the knowledge base on drought and impacts and forms a stronger baseline to find relations with adaptation measures and resilience. We have changed the word ‘*other*’ to ‘*similar*’ [515].

Technical corrections:

Line 6) include “that” between “indices” and “best”.

Line 7) Remove “:”.

Line 10) “aridity” to “aridity level of the regions”.

Line 15) "impactful" instead of "impact-full".

Line 52) Replace "setting" with "regions".

Line 98) Remove "However".

Line 103) "reanalysis" instead of "re-analysis".

Line 104) Remove "extracting".

Line 109) the complementary nature of "the" highest.

Line 115) Remove (v3.5a).

Line 122) Add e.g., before the papers.

Line 140) "has" instead of "have".

Line 141) "from March 2018 to the present." instead of "between".

Line 146) Replace "However" with "For this" or something similar. “However” here does not make sense.

Table 1) Add a horizontal line between the first row and the second row.

Line 176) Instead of "between ... and ... " use "from ... to ...".

Line 236 and 247) Replace "a bit" with "slightly".

Line 317) Add "Also" after the first sentence. "Also, the interference...".

Line 320) Replace "a kind of" with "some time" lag.

Line 322) "cannot " instead of "can not".

Line 338) Remove ";" after Kenya.

Line 340) in "the" ASAL counties.

Line 362) "imply" instead of "implies".

Line 397) Add "Our" in front of the Result.

Line 399) "the first step".

Line 400) I'd use "exploring the link" instead of "analysing".

Line 409) Remove "In addition".

Plots in general) Enlarge the fonts and dots.

Response to RC2

Manuscript Number: egusphere-2022-458

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

Authors: Marleen R. Lam, Alessia Matanó, Anne F. Van Loon, Rhoda Odongo, Aklilu D. Teklesadik, Charles N. Wamucii, Marc J. C. van den Homberg, Shamton Waruru, and Adriaan J. Teuling

The work analyzes the applicability of drought indices in explaining drought impacts using Kenya as a case study. A Random Forest (RF) model was used to identify which drought indices best explains drought impacts on pasture, livestock deaths, milk production, crop losses, food insecurity, trekking distance for water, and malnutrition.

Major Comments:

- The results presented in lines 230-235 state that the drought impacts better overlaps with the accumulation time of 12 months in comparison with the others tested (1, 3, 6 and 24). However, isn't expected that the effects of meteorological droughts on pasture, milk production and etc. be lagged either than happening simultaneously? How is this lag effect considered? Also, the SPEI indices of all accumulation times displayed in Figure 2 show good overlap with the impacts listed.

We thank the reviewer for this remark. Indeed, it is expected that drought impacts are lagged in relation to the onset of a drought. The considered accumulation periods can be used to account for time lags and memory encountered in hydrological stores, but they cannot directly account for lags between drought impacts. The lag between drought impacts fall outside the scope of this research as we only look at which drought indices best match with drought impacts under different aridity levels. However, it would be a very interesting follow up research to study the actual lags between drought impacts.

- The study does not account for drought resilience actions that could have been applied in the region (e.g., reservoirs, irrigation, etc.) and can affect the correlation of meteorological droughts and its impacts.

We thank the reviewer for this valuable notification. This study does not account for drought resilience actions that could have been applied in the region which can affect the correlation of drought and its impacts. Therefore, we have added a piece of text, mentioning this together with the possible drought resilience actions that could have been applied:

> *[465] This study did not directly account for short and/or long term drought resilience actions applied in the ASAL regions and their link to the reported drought impacts in the NDMA monthly bulletins while this could be of influence on the drought impacts-indices relationship. Drought resilience actions can be related to (1) structural interventions for increasing the water availability (e.g. construction of reservoirs), (2) sustainable land management practices (e.g. inter-cropping, agroforestry and drought resistant crops), (3) pasture and livestock management (e.g. livestock restocking and improved varieties of grass) and (4) livelihood diversification (Kenya, 2016; Mude et al., 2007; Njarui et al., 2020; Opiyo et al., 2015; Parry, 2016; weADAPT, 2023). Such adaptation measures can increase the resilience of the communities (Nyberg et al., 2020). This could as well explain the differences in the optimal drought indices found among the researched areas as the level of preparedness can lag or mitigate the occurrence of drought impacts, resulting in a better fit with drought indices with higher accumulation periods.*

- How do the current results compare with the one cited in line 301-305, using other correlation metrics?

We thank the reviewer for this suggestion. As a similar suggestion was made by the other reviewer, we added another correlation analysis, namely the point-biserial cross correlation. The results are presented in Appendix C. In general, the results of the point-biserial cross correlation are comparable with the results of the RF model for the models considered for discussion (only relatively good performing models were discussed in terms of variable importance). For instance, SSI-06 is mostly correlated to the occurrence of impacts related to Pasture both with the point-biserial cross correlation and in the RF analysis (Figure 4). Same for Livestock deaths, for which SPEI-12 and SPI-12 are most relevant indices which is similar to the results of the RF analysis. These similarities confirm the reliability of the RF method.

Minor Comments:

It should be mentioned in the abstract which drought indices were utilized in the analysis and which ones better performed in predicting the drought impacts.

- We thank the reviewer for this valuable remark. Indeed, as the other reviewer noticed as well, the abstract was a bit too general. We have changed it as follows:

  [4-16] The relation between drought severity and drought impacts is complex and so far relatively unexplored in the African continent. This study assesses the relation between reported drought impacts, drought indices, water scarcity, and aridity across several counties in Kenya. The monthly bulletins of the National Drought Management Authority in Kenya provided drought impact data. A Random Forest (RF) model was used to explore which set of drought indices (Standardized Precipitation Index, Standardized Precipitation Evapotranspiration Index, Standardized Soil Moisture Index and Standardized Streamflow Index) 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 positive relation with aridity. A relation between water scarcity and aridity was not found. The RF model revealed that every region, aggregated by aridity, had their own set of predictors for every impact category. Longer timescales (≥ 12 months) and the Standardized Streamflow Index were mostly present, indicating the importance of hydrological drought to predict drought impact occurrences. 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.

In Figure 2, what do the circles in the upper panel represent? Is it only if the impact occurred or does it reflect the number of occurrences?

- The circles in the upper panel in Figure 2 represents the impact occurrence, the colors indicate which impact occurred, so it does not reflect the number of impact occurrences. We added the following to the caption of Figure 2: '*The colored dots indicate the type of impact occurrence.*' (p. 11).

Response to CC1

General comment:

This study discusses the possible linking of drought indices with text-based documents by a drought management authority in Kenya. An assessment is made how drought impacts relate with drought indices and water scarcity under various circumstances, by using a Random Forest Model. In doing so, the authors aim to contribute to ongoing debates about operational needs for drought monitoring. The importance of improving early warning systems to mitigate adverse consequences of drought is corroborated by the study and its results. In general, I believe the authors present a very well-written manuscript, with clear and complete (sub)sections. The quantity and quality of references to relevant state-of-the-art studies is spot on. I can only suggest some minor and a few moderate revisions require attention prior to publication. These are summarised below. Well done!

Minor and moderate suggestions for revision.

<table>
<tr><th>Linenumber</th><th>Comment</th></tr>
<tr><td>25</td><td>Consider changing ‘society’ to ‘societies’</td></tr>
<tr><td colspan="2">We thank the reviewer for this remark. However, we decided to stick to the word ‘society’ because we see this better fit.</td></tr>
<tr><td>30-37</td><td>I think the terms meteorological and hydrological drought do not need such an elaborate explanation. If you think the readership does need this explanation, then consider adding a short definition of ‘soil moisture/agricultural drought’ as well.</td></tr>
<tr><td colspan="2">We thank the reviewer for this comment. We agree that this terminology does not need such an elaboration. Therefore, this paragraph is shortened and more explanation about drought types in relation to drought indicators have been added which was suggested by the other reviewer as well [29-34].</td></tr>
<tr><td>83-85</td><td>Consider including the Köppen classification to describe climatological setting in Kenya, to highlight the diversity between the relatively wet southwest and dry/arid/desert north and east.</td></tr>
<tr><td colspan="2">We thank the reviewer for this comment. However, later on we introduce the difference between the regions in terms of aridity levels which is the focus of this study. Therefore, we have decided not to include the Köppen classification.</td></tr>
<tr><td>117</td><td>How was the difference in spatial resolution (0.25 degree) with the other datasets (0.1 degree) dealt with in by the authors?</td></tr>
<tr><td colspan="2">We thank the reviewer for this remark. The gridded values of the drought indices were spatially aggregated to the county resolution by averaging the values of all the grid cells per county. This is explained at lines [186-187]. The monthly averaged values per county were used for this study.</td></tr>
<tr><td>145</td><td>Consider changing ‘for each person’ to ‘per capita’</td></tr>
</table>

We consider that ‘for each person’ fits better in this case as otherwise there will be two times ‘per’ in the sentence.	
148	Replace ‘have been’ by ‘was’
As suggested by the reviewer, we have changed ‘have been’ to ‘was’ [173].	
159-160	How are the grid cells spread over the different counties? Did you only compute the means of grid cells completely within a county border? Please specify this, also the resolution that was used for the grids.
Information about the datasets (among others, their spatial resolution) used for the computation of the drought indices are given at section 2.2.1. The spatial resolutions of those datasets are slightly different but high in general (0.1 and 0.25). The gridded values are spatially aggregated to country resolution (so based on the administrative units of the counties), therefore averaged as explained at lines [186-187].	
173-174	So the indices were calculated for 1980-2020, but only used for 2014-2020. What is the added value of calculating 1980-2014, if it is not being used?
We thank the reviewer for this remark. The indices are calculated for 1980-2020 as historical data is needed for the computation of the drought indices and their accumulation periods, see the handbook of drought indices (WMO & GWP, 2016) whereby it is suggested to have a minimum of 30 years of data to determine the ‘normal period’. *World Meteorological Organization (WMO) and Global Water Partnership (GWP). "Handbook of drought indicators and indices." Integrated Drought Management Programme (IDMP), Integrated Drought Management Tools and Guidelines Series 2 (2016).*	
194	‘have’ = ‘has’
We thank the reviewer for this accurate remark. The word has been changed from ‘have’ to ‘’has’.	
202-203	does this aggregation have any drawbacks?
With aggregation of the counties per aridity level, we mean that the drought impact datasets are combined. Perhaps this can result into less detailed information on county level. However, the advantage is that the regions can be studied according to aridity level which correspond with the aim of this research.	
Table 3	Could you indicate p-values in this table? i.e. $p<0.05$ = *, $p<0.01$ = **, $p<0.001$ = ***
We thank the reviewer for this suggestion. We added asterisks indicating the statistical significance according to Chung *et al*., 2019. When the p-value is below 0.05, this is indicated with a *. *Chung, N. C., Miasojedow, B., Startek, M., & Gambin, A. (2019). Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data. BMC bioinformatics, 20(15), 1-11.*	
243-244	It is a bit unclear what the authors mean by 2, 4, 2 months out of 10/4,0,4 months out of 12. Consider rewriting this sentence.
We thank the reviewer for this valuable remark. The sentence could be a bit confusing as ‘(WS dataset starts at March 2018)’ is as well included at already a sentence in parentheses. Therefore, the following sentence is included at the beginning of section 3.2: *‘This is done from 2018 due to the length of the WS dataset’ [304]*. We have decided to include the rest of the part in parentheses because it is additional information which could be of value to the reader.	

<table>
<tr><td>251</td><td>The AUC abbreviation was not previously introduced and should be written out fully.</td></tr>
<tr><td colspan="2">We thank the reviewer for this accurate remark. The AUC is now written out fully at the Method section 2.3.3. and properly introduced [265].</td></tr>
<tr><td>Figure 4 & 5</td><td>Font size should be increased. In addition, the range of the X-axis is different for each figure. If relevant for comparison between the different categories (which I think is the case), consider using the same range for each figure. You might even want to plot the points of multiple categories in one figure with different colours, instead of having 14 individual subplots.</td></tr>
<tr><td colspan="2">We thank the reviewer for this remark. After consideration, we found that the ‘variable importance’ function (varImp) of the caret package in Rstudio (version 6.0-93) is very well suited for our study. This importance function determines how much accuracy the RF model would lose if the variable would be permuted of the model. The most important variables are the ones where the RF model depends on the most to make accurate predictions. The results are similar to the MDA and Gini index (as used in the first manuscripts) but captures the results in less plots, making it easier for the reader. As suggested by the reviewer, the font size of the new plots are increased and bar plots are used instead of dots which increase the readability (Figure 4, 5 and 6).</td></tr>
<tr><td>293</td><td>Please avoid using the term ‘reliable’ when talking about an objective appraisal, as this is a subjective judgement without qualitative or quantitative data to support the statement. It would improve the statement to shortly explain how the iterative processes and the focus on abnormal conditions took place, as I do not recall reading about this in the methods section. I believe this is important, because the bulletins form an integral part of your analyses.</td></tr>
<tr><td colspan="2">We thank the reviewer for this thoughtful comment. As suggested by the reviewer, more information about the focus on ‘abnormal conditions’ should be given. This is related to the selection of bulletins included for this analysis as only the bulletins with phase classification ‘alert’, ‘alarm’ or ‘emergency’ where included. In the revised manuscript, more information about this is given at section 2.3.2. Drought impact data. This is also important information in relation to our definition of drought impacts as explained at lines 223-224. As we explained these ‘abnormal conditions’ in more detail at the method section, it is not necessary to use this word anymore. In addition, the word ‘iterative process’ might be confusing in this context as we mean that the county early warning bulletins of the NDMA are checked multiple times. Therefore, we rephrased it as follows:

[361] In addition, the bulletins were iteratively checked by several NDMA employees to ensure a reliable list of drought impact data.</td></tr>
<tr><td>298-300</td><td>Great to see the authors mention multiple outlets to complement NDMA bulletins!</td></tr>
</table>

We thank the reviewer very much for this compliment!	
306	Good': please quantify this by for instance mentioning the average AUC value, or repeat the AUC value of the top 3 categories.
We thank the reviewer for this suggestion. However, for the revised manuscript, we have included multiple performance metrics beside the AUC, namely a different set of performance metrics for the test and training data set (Table 4). For our study, we only discuss the 'variable importance' of the relatively good performing models: to indicate this, we added a ✓ at Table 4. In addition, we conducted an additional correlation analysis (the point-biserial cross correlation) between the drought impacts and drought indices. The results of this correlation analysis are more or less similar as the results of the RF model which also confirms the potential of the RF model for drought. Therefore the sentence is changed as follows: *[375] This study confirms this, as the performance metrics were good for several drought impact categories and all the discussed models in terms of 'variable importance' showed similarities with the results of the point-biserial correlation (Appendix C).*	
320	'a kind of lag' – please specify in more detail.
As drought impacts mostly happen after the start of a drought event, those impacts are lagged in time in relation to the drought event itself. After consideration, we decided to remove this sentence.	
339	'low population density does not imply low water stress' – this calls for a reference.
We thank the reviewer for this comment. We agree and have added the following reference: *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, ac725, accessed: 2022-17-05, 2013.*	
350	'should have suffered from water scarcity during periods of drought due to the high population density' – this reads a bit dark. Consider rephrasing the sentence so it does not read as if you want these counties to suffer.
We thank the reviewer for this accurate remark. We have decided to change the word 'should' into 'could' and 'suffered' into 'experienced', to read: *[421] The sub-humid central-western counties, on the other hand, could have experienced water scarcity during periods of drought due to the high population density and hence the high pressure on available water resources.*	

<table>
<tr><td>371-373</td><td>this was already mentioned earlier in the manuscript and does not need repetition here.</td></tr>
<tr><td colspan="2">We thank the reviewer for this comment. However, we do not mention specific accumulation periods in relation to the drought types earlier in the text. Therefore, we decided to still include this sentence as it is an important division that should be clearly mentioned.</td></tr>
<tr><td>376-378</td><td>The authors mention comparisons with these studies are difficult due to different socioeconomic and climatic circumstances. That’s a fair point, but how do they compare in terms of resolution? Is it practically possible to compare these quantitatively, or does this also not work?</td></tr>
<tr><td colspan="2">We thank the reviewer for this comment. The main reason that it is difficult to compare those studies in relation to linking drought impacts with drought indices is the difference in socioeconomic and climatic circumstances. The occurrence of a drought impacts is determined by (1) the intensity, frequency and area of drought which differ per region and (2) by the amount of resilience and vulnerability of the system (related to drought preparedness) which is mostly determined by the socioeconomic circumstances of a country. This makes the relation between drought impacts and drought indices very country and even region specific.</td></tr>
<tr><td>400-401</td><td>Could you please include a reference to this existing database (or name it)?</td></tr>
<tr><td colspan="2">We thank the reviewer for this suggestion. However, as we do not want to highlight some studies with this statement but just make a ‘general’’ statement (adding to the existing literature database on drought and its impacts), we have decided to not include specific references as we mention a lot already in the Introduction and Discussion sections (e.g., Bachmair et al., 2015, 2016, 2018).</td></tr>
<tr><td>432-433</td><td>Is there any indication of the spatial resolution required to capture the regional differences?</td></tr>
<tr><td colspan="2">As drought can happen very locally, finer spatial scales of drought impact data is needed to capture the details. There is not an unequivocal answer which spatial resolutions of drought impact data are required to capture regional differences as this depend on the aim of the research (e.g. looking at drought impacts outside the affected region) and the drought event itself (e.g., the area of the drought event).</td></tr>
<tr><td>431-437</td><td>This last paragraph of the conclusion reads well and contains and sensible information, but I think it undermines the results of the study. It reads as if the work explored in this study is disqualified a bit, since focus is put on the need for finer resolutions to contribute to the development of early warning systems. I recommend the authors to ‘praise’ their own work a bit more in this</td></tr>
</table>

	last paragraph, instead of talking it down.

We thank the reviewer for this remark. The first part of the last paragraph of the conclusion is based on recommendations for further research. This should be stated very clear without talking down our own research, we have changed the following sentence to emphasize that it is a recommendation:

[514] In addition, we recommend to look at finer spatial resolutions to capture the regional differences in human influences on water scarcity and drought impacts.

In addition, we have added the following to make sure that we emphasize the innovation of this research:

[517] This study analyzed the link between drought indices and text-based impact reports with a focus on the African continent which has never been studied before.

References

Atzberger, C. and Eilers, P. H.: Evaluating the effectiveness of smoothing algorithms in the absence of ground reference measurements, International Journal of Remote Sensing, 32, 3689–3709, https://doi.org/10.1080/01431161003762405, 2011.

Atzberger, C., Formaggio, A., Shimabukuro, Y., Udelhoven, T., Mattiuzzi, M., Sanchez, G., and Arai, E.: Obtaining crop-specific time profiles of NDVI: the use of unmixing approaches for serving the continuity between SPOT-VGT and PROBA-V time series, International Journal of Remote Sensing, 35, 2615–2638, https://doi.org/10.1080/01431161.2014.883106, 2014.

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.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System, Journal of hydrometeorology, 10, 623–643, https://doi.org/10.1175/2008JHM1068.1, 2009.

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, 2017a.

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A. I., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., and Wood, E. F.: Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling, Hydrology and Earth System Sciences, 21, 6201–6217, https://doi.org/10.5194/hess-21-6201-2017, 2017b.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I., McVicar, T. R., and Adler, R. F.: MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment, Bulletin of the American Meteorological Society, 100, 473–500, https://doi.org/10.1175/BAMS-D-17-0138.1, 2019.

Beguería, S., Vicente-Serrano, S. M., Reig, F., and Latorre, B.: Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring, International journal of climatology, 34, 3001–3023, https://doi.org/10.1002/joc.3887, 2014.

Cattani, E., Ferguglia, O., Merino, A., and Levizzani, V.: Precipitation Products' Inter–Comparison over East and Southern Africa 1983–2017, Remote Sensing, 13, 4419, https://doi.org/10.3390/rs13214419, 2021.

Chung, N. C., Miasojedow, B., Startek, M., and Gambin, A.: Jaccard/Tanimoto similarity test and estimation methods for biological presenceabsence data, BMC bioinformatics, 20, 1–11, https://doi.org/10.1186/s12859-019-3118-5, 2019.

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.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the royal meteorological society, 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

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

Erian, W., Pulwarty, R., Vogt, J., AbuZeid, K., Bert, F., Bruntrup, M., El-Askary, H., de Estrada, M., Gaupp, F., Grundy, M., et al.: GAR Special Report on Drought 2021, 2021.

Falkenmark, M., Lundqvist, J., and Widstrand, C.: Macro-scale water scarcity requires micro-scale approaches: Aspects of vulnerability in semi-arid development, in: Natural resources forum, vol. 13, pp. 258–267, Wiley Online Library, https://doi.org/10.1111/j.1477-8947.1989.tb00348.x, 1989.

Hanley, J. A. and McNeil, B. J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve., Radiology, 143, 29–36, https://doi.org/10.1148/radiology.143.1.7063747, 1982.

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.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hirpa, F. A., Salamon, P., Beck, H. E., Lorini, V., Alfieri, L., Zsoter, E., and Dadson, S. J.: Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data, Journal of Hydrology, 566, 595–606, https://doi.org/10.1016/j.jhydrol.2018.09.052, 2018.

Jaccard, P.: The distribution of the flora in the alpine zone. 1, New phytologist, 11, 37–50, https://doi.org/10.1111/j.1469-8137.1912.tb05611.x, 1912.

Kenya: Kenya National Adaptation Plan: 2015–2030, 2016.

Klisch, A. and Atzberger, C.: Operational drought monitoring in Kenya using MODIS NDVI time series, Remote Sensing, 8, 267, https://doi.org/10.3390/rs8040267, 2016.

Mishra, A. K. and Singh, V. P.: A review of drought concepts, Journal of hydrology, 391, 202–216, https://doi.org/10.1016/j.jhydrol.2010.07.012, 2010.

Mude, A. G., Ouma, R., Steeg, J. v. d., Kaiuki, J., Opiyo, D., and Tipilda, A.: Kenya adaptation to climate change in the arid lands: Anticipating, adapting to and coping with climate risks in Kenya-Operational recommendations for KACCAL, ILRI Research Report, 2007.

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

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.

Njarui, D., Gatheru, M., and Ghimire, S. R.: Brachiaria grass for climate resilient and sustainable livestock production in Kenya, African Handbook of Climate Change Adaptation, pp. 1–22, https://doi.org/10.1007/978-3-030-42091-8_146-1, 2020.

Opiyo, F., Wasonga, O., Nyangito, M., Schilling, J., and Munang, R.: Drought adaptation and coping strategies among the Turkana pastoralists of northern Kenya, International Journal of Disaster Risk Science, 6, 295–309, https://doi.org/10.1007/s13753-015-0063-4, 2015.

Parry, J.-E.: Review of current and planned adaptation action in Kenya, 2016.

Priestley, C. H. B. and Taylor, R. J.: On the assessment of surface heat flux and evaporation using large-scale parameters, Monthly weather review, 100, 81–92, https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Vicente-Serrano, S. M., Wehner, M., and Zhou, B.: 11 Chapter 11: Weather and climate extreme events in a changing climate, 2021.

Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., et al.: Changes in climate extremes and their impacts on the natural physical environment, 2012.

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.

Van Der Knijff, J., Younis, J., and De Roo, A.: LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation, International Journal of Geographical Information Science, 24, 189–212, https://doi.org/10.1080/13658810802549154, 2010.

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.

Wanders, N., Van Loon, A. F., and Van Lanen, H. A.: Frequently used drought indices reflect different drought conditions on global scale, Hydrology and Earth System Sciences Discussions, pp. 1–16, https://doi.org/10.5194/hess-2017-512, 2017.

weADAPT: Enhancing resiliency to drought in Kenya's arid and semi-arid lands, https://www.weadapt.org/, 2023.

WFP, V.: Food consumption analysis: calculation and use of the food consumption score in food security analysis, WFP: Rome, Italy, 2008.

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.