Manual on how to use the codes:

The codes which are used for this analysis are written using python version 3.8 and R-4.1.2.

In each folder you will find a code named “packages and dists.R” which will install packages which are required for this analysis, if they are not already installed in your personal computer. Also, this code contains the distribution functions (Generalized extreme value, Generalized logistics and Pearson type 3 distribution). Gauging station name, latitude and longitude with length of precipitation data is given under the excel file titled “125 stations names.csv”.

To reproduce the results which are presented in the paper, you will have to run multiple codes/ or jump straight to the last part (Additional code) where only a single code can produce most of the results where the procedure is presented here under.

**Step 1:** First fill missing values of the raw precipitation data by bias corrected Chirps data. In order to do this, you will have to run the codes inside the folder “Precipitation and PET data”.

Inside this folder, you will have to run the code “Chirps data and bias correction.R”. This will extract the Chirps data for each gauging station (saved inside the “Chirps global dataset” folder) and then bias correct them (saved inside the “Chirps bias corrected”) using the raw precipitation data (found inside the folder “Raw precipitation in situ data”).

In addition, PET values are calculated using the “PET estimation.R” code which uses the monthly average temperature acquired from the ERA5 (located inside the ERA5 temperature data). After that this similar code will calculate the water balance (precipitation (missing values filled in) – PET). This water balance will be used for most of the tasks, except for the SPAEI analysis. Hence, “WB” folder is copied to each of the other folders (not to the SPAEI folder).

**Step 2:** Using the water balance (WB folder) values, the goodness of fit (GOF) will be assessed. In the GOF test, critical values are needed which are computed using the “critical\_value\_MC.R” code. This code will result in critical values for 10,15,20,25 and 32 sample sizes for the three distributions. For the GOF, you will have to use codes inside the folder: “GOF” where codes for testing using the methods of Anderson-Darling, Kolmogorov-Smirnov and Shapiro Wilk’s test are provided.

N.B: the Shapiro Wilk’s test is done on the estimated SPEI values (found inside the folder: “SPEI”), hence, first, estimation of the SPEI values has to be done before using the SW test code (step 4 has to be done ahead for obtaining the SW test result).

The result of the different tests are exported as .csv file (“AD test result all together.csv”, “KS test result all together.csv” and SW test result all together.csv) which gives the rejection frequency for each distribution per accumulation period.

**Step 3:** The analysis on the acceptance frequency (1-rejection frequency) at a month level is done using a code inside the folder “Acceptance frequency each month”. This will result in a acceptance frequency for each month per distribution function.

**Step 4:** Next, you can estimate the SPEI values using the three distributions and this can be achieved by using the code “SPEI.R” inside the “SPEI” folder. This will save the SPEI values estimated for all accumulation periods into different folders attributed to the three distributions.

In this folder, there exists also python codes (“Extreme drought index analysis and plot.py” and “NSE estimation.py”) to produce the plots of the result from extreme drought and similarity analysis (which are presented in Sec. 3.3.2 and 3.3.3 of the main article).

**Step 5:** Next, SPAEI analysis codes can be found inside the folder : “SPAEI”. First, use the “GLEAM.R” code to extract the actual evapotranspiration (AET) values from the GLEAM dataset (ethiopia\_low\_resgleam.Rdata) which is stored in AET folder. The same code will also estimate water balance values (Precipitaiton – AET) and store them in WB SPAEI folder. Then, goodness of fit is assessed using the “GOF fitdist AD SPAEI.R” to assess which distribution is appropriate to fit these water balance values. Finally, SPAEI values can be estimated using the “SPAEI.R” code and stored on the different distribution attributed folders.

**Step 6:** The sort timeseries analysis codes are available inside the “Shorter timeseries fitting” folder. The codes include; goodness of fit (GOF) analysis codes for each timeseries length (e.g “GOF 10 MW.R”) where the moving window (MW) timeseries selection is done inside this codes and then ultimately assess the GOF.

The codes with the name “SPEI**n**.R” where n is the number of years of timeseries (10,15,20 and 25) are used to estimate the SPEI values using the three distribution functions. The codes with the name “For comparing SPEI values **n**.R” where **n** is the timeseries length are used to compare the SPEI values estimated from the shorter timeseries length against those estimated using the benchmark data (folder labeled Genlog as it is the distribution that is found to be best among the three distributions). The shorter timeseries length starting and ending year is selected using the python code “Drought starting and ending selection.py”. The results will be saved in the “Statistics” folder. The results can be plotted for NSE using the “plot.py” code and the plots are enumerated as 1 NSE, 3 NSE, etc.

Finally, it is required to run the codes with a name of “NSE\_estimation**n**.R” where **n** is the timeseries length. This code will perform moving window and select timeseries with **n** length and estimate the Nash-Sutcliffe efficiency (NSE), Pearson Correlation coefficient and root mean square error (RMSE). This code will take several hours of simulation, hence, you will have to wait up to 3.5 days to obtain the results (as the simulation is for each moving window timeseries and each accumulation period and each distribution function).

**Additional code:** The codes for all tasks except for shorter timeseries analysis and critical value estimation, can be run once by running the “master\_code.R” that you will find in the main folder. This is to avoid confusion and also make the process easy.

Thank you!

Estifanos Addisu Yimer

21-12-2021