

University of Augsburg

In cooperation with Karlsruher Institut für Technologie (KIT)

Summer Term 2021

Supervision: Prof. Dr. Harald Kunstmann

Submission Deadline: 30.09.2021

Classification of weather patterns for the Greater Horn of Africa using cluster analysis

Master Thesis

Jan Niklas Weber

Rießerseestraße 40

82467 Garmisch-Partenkirchen

jan.niklas.weber@student.uni-augsburg.de / jan.weber9@kit.edu

Matr. No.: 1601235

Master of Science: Climate and Environmental Sciences

I. Table of Contents

I. Table of Contents	II
II. List of Figures	IV
III. List of Tables	V
IV. Abstract.....	VI
1. Introduction	1
2. Methods and Datasets	3
2.1. Clustering.....	3
2.1.1. Cost733Class	3
2.1.2. SANDRA.....	4
2.2. Explained Cluster Variance and Adjusted Rand Index.....	5
2.3. Statistical tests	6
2.3.1. Pearson-Correlation and p-Value	6
2.3.2. Mann-Whitney-U-Test.....	6
2.3.3. Mann-Kendall-Trend-Test	7
2.4. Drought indices	8
2.4.1. Standardized Precipitation Index	8
2.4.2. Standardized Precipitation Evapotranspiration Index.....	8
2.4.3. Daily Wetness Index and Daily Temperature Index.....	10
2.5. ERA5	11
2.6. Teleconnection data.....	12
3. Study Region	13
3.1. Overview	13
3.2. Topography	14
3.3. Climatology	16
3.3.1. Rainy seasons	17
3.3.2. Teleconnections and their indices.....	19

3.4. Preselection of variables	21
4. Results.....	22
4.1. Underlying trends of the variables.....	22
4.2. Selection of the number of clusters	28
4.3. Selection of the variables	30
4.4. Characterization of the clusters	33
4.4.1. tp-Clusters	33
4.4.2. SST-Clusters	41
4.4.3. wd900-Clusters.....	46
4.4.4. Drying influence of the clusters.....	50
4.5. Interactions of the clusters with each other	51
4.6. Influence of teleconnections.....	54
5. Discussion.....	62
5.1. Outlook.....	66
5.2. Acknowledgements	67
V. Bibliography	VII
VI. Appendix.....	XV

II. List of Figures

Figure 1:	Study Area	14
Figure 2:	Topographic elevation map.....	15
Figure 3:	Climate diagram of four selected cities	17
Figure 4:	Standardized Precipitation Evaporation Index (SPEI).....	25
Figure 5:	Dry Days of every country from 1980 to 2018	26
Figure 6:	Shift in the start and end of the rainy season per country.....	27
Figure 7:	Explained Cluster Variance overview	28
Figure 8:	Adjusted Rand Index.....	30
Figure 9:	ARI-Plot of the studied variables with the rainfall clusters	32
Figure 10:	Deviation of tp cluster centroids from mean precipitation.....	34
Figure 11:	Cluster centroids of precipitation.....	35
Figure 12:	Deviation of SST of the cluster centroids from the mean	41
Figure 13:	Daily Wetness Index and average wind direction	42
Figure 14:	Clusters of wind direction at 900 hPa.....	47
Figure 15:	Wetness and heat of all clusters using DWI and DTI	50
Figure 16:	Multi-Directed Graph depicting the transition probabilities.....	52
Figure 17:	Correlation of the frequency of the individual clusters	56

III. List of Tables

Table 1:	Temperature and precipitation per Country.....	18
Table 2:	Trends of selected variables and precipitation	23
Table 3:	Cluster days more than one standard deviation away from the mean.....	40
Table 4:	Correlation of wind clusters with the precipitation clusters	54
Table 5:	Correlation between the clusters and the teleconnections	61

IV. Abstract

The Long Rains, the dominant rainy season of the Greater Horn of Africa, are so far only poorly understood. The dominant precipitation occurring in the March-April-May (MAM) season responds weakly, if at all, to teleconnection influences. In this thesis, a new approach to study the MAM Season using a cluster analysis is tested. The cluster analysis groups together days of similar characteristics of a specific variable so that main circulation patterns can be more clearly identified, which in turn can be examined for their properties. The variables determined to be most appropriate for clustering are the Sea Surface Temperature (SST) and the wind direction at the 900 hPa level. The performed cluster analysis confirms the continuing trend towards less precipitation since 1998 with an increase in dry rainfall clusters and a significant drought trend in the individual clusters of predictor variables. The root cause of this trend is a shortening of the MAM rainy season in the study region, especially in Tanzania. The end of the season depends mainly on the El Niño Southern Oscillation (ENSO). A particularly dry SST pattern was found that clearly correlates with the ENSO3.4 index and causes the rainy season to end earlier under La Niña conditions, so that the total precipitation of the Long Rains decreases in these years. This connection between the pressure patterns in the Pacific Ocean and the end of the rainy season is already established in December of the previous year, so there is a possibility that this relationship could be used to predict the date of the end of the rainy season several months in advance. Furthermore, precipitation in Kenya is co-determined by the characteristics of several wind direction clusters. The precipitation amount correlates with the frequency of the clusters, the precipitation patterns produced by the wind clusters are clearly contributing to a dry/wet MAM season in Kenya. In conclusion, cluster analyses can improve the understanding of the climate system on how and to what extent Sea Surface Temperatures and teleconnections can influence precipitation in specific regions of the Greater Horn of Africa.

1. Introduction

The Greater Horn of Africa (GHA) region has experienced a number of devastating droughts, storms, and other extreme weather events in the preceding decade, leading to famine and the displacement of millions of people (ROWELL et al., 2015). For example, a severe drought occurred in East Africa in 2017, leading to extreme food shortages (FUNK et al., 2018). This is contrasted with extreme rainfall during the rainy season in recent years, which caused severe damage to agriculture and infrastructure in Kenya in 2018 and Tanzania in 2019 (WMO 2020). Such damage forced resident people to migrate; Ethiopia had with 390,000 people, Ethiopia had the third highest number of internally displaced people in the world in 2019, due to drought and a simultaneous plague of locusts, surpassed only by Afghanistan and India (WMO, 2020). In 2020, food shortages forced as many as 2.45 million residents of the Greater Horn of Africa to migrate. Such extreme events are becoming more frequent due to climate change (GUDOSHAVA et al., 2020). Additionally, a desert locust outbreak developed in late 2019 and continued through mid-2021, covering all of Ethiopia and Somalia and parts of Kenya, Uganda and Eritrea. Such outbreaks result from extreme precipitation events, which are also becoming more frequent due to climate change. In this case, the aforementioned heavy rainfall in 2018 caused desert locusts to multiply unnoticed in several locations, and subsequent generations began migrating in 2019, destroying much of the farmland in Ethiopia and Somalia (MEYNARD et al., 2020). While not all disasters can be prevented, strategies that start several weeks in advance can effectively mitigate impacts. For example, the breeding grounds of desert locusts can be sprayed early with targeted pesticides, which stops the exponential reproduction of insects (MEYNARD et al., 2020), or drought-resistant crops can be planted.

Such information on the course of the rainy seasons is offered by seasonal forecasts. To improve them, the mechanisms and influences on the different rainy seasons have to be analyzed first. In the southern and central areas of the Greater Horn of Africa, there are two rainy seasons per year, from March to May (MAM Season, "Long Rains") and from October to December (OND Season, "Short Rains"). More precipitation falls during Long Rains; in terms of replenishing groundwater resources and restoring soil moisture, they are the key factor because precipitation falls more regularly, allowing for better soil infiltration (YANG et al., 2015). But climate change is transforming the rainfall patterns and the intensity of the Long Rains. While no significant change was observed in Short Rains from 1998 to 2018, precipitation decreased in the MAM Season (BAHAGA et al., 2019; ENDRIS et al., 2019). In addition, temperatures and thus evaporation have been steadily increasing since 1980 (OMONDI et al., 2014). This exacerbates the water situation, especially in northern Somalia, southern Ethiopia, and northern Kenya;

already existing water availability problems become more severe due to increased moisture evaporation.

Several authors have already examined the causes of variability in Long Rains to gain a better understanding of the atmospheric processes during the MAM season. FUNK et al. (2018) highlights the influence of ocean temperatures on the severity and frequency of droughts in East Africa. MUTAI et al. (1998) and HASTENRATH et al. (2004) analyze that not only the spatially close Indian Ocean SSTs influence precipitation in East Africa and the East African coast, but also the Pacific SSTs and, to a lesser extent, the East Atlantic SSTs. GISSILA et al. (2004) differentiates, using Ethiopia as an example, that the influence of SSTs varies significantly by region and uses the findings to improve seasonal forecasts. KORECHA and BARNSTON (2007) found that by looking closely at the evolution of ENSO rainfall anomaly patterns can be predicted. PHILIPPON et al. (2002), on the other hand, focuses on the extent to which the eastern Indian Ocean affects the atmosphere in terms of zonal wind, vertical velocity, and moist static energy, which in turn influence precipitation in Kenya and Somalia. BAHAGA et al. (2015) reached similar conclusions. However, those studies deal with the Kiremt Rains (June–September in Kenya) or the Short Rains, the dynamical causes and mechanisms of those seasons are hence relatively well understood, knowledge of the Long Rains is much poorer to date (POHL and CAMBERLIN, 2006; MACLEOD, 2019; WALKER et al., 2019). Moreover, as described at the beginning, climate change brings additional dynamics to the system, which are difficult to assess so far. GLEIXNER et al. (2017) compares several seasonal prediction models and concludes that all perform worse in terms of predicting precipitation during the Long Rains than a simple linear prediction model based on ENSO3.4, which supports the WMO's statement that for MAM season, climate services are therefore still weak despite improved funding opportunities (WMO 2020). CAMBERLIN and PHILIPPON (2002) focused the Long Rains, using a Principal Components Analysis to examine the influences on the rainy season. They indicate a link between MAM season precipitation and local zonal wind as well as 500-hPa geopotential dynamics. ENSO also has a small influence at this time of the year, with La Niña leading to below-average precipitation in Kenya and Uganda through a weakening of the near-surface Easterlies. However, these relationships are considered rather weak and the atmospheric variables are inferior to SST in predictive power. VIGAUD et al. (2017) uses a cluster analysis of the outgoing longwave radiation to capture the convection variations during the MAM season and to investigate their influence on the long rains. A cluster analysis is a method that divides a data set into groups within which the individual group members are as similar as possible. VIGAUD et al. (2017) thus classifies the atmospheric state as a circulation pattern and can therefore present an improved analysis of the relationship between outgoing longwave radiation and precipitation. Building on this approach, in order to contribute to a better understanding of the Long Rains, the present thesis will use cluster analyses to explore correlations between teleconnections, wind systems and precipitation. In addition, climate trends of precipitation, temperature and

several drought indicators will be analyzed to draw conclusions about the influence of climate change. The aim is to improve the understanding of the climate system in the hydrologically vulnerable region of the Greater Horn of Africa, in order to improve the performance of seasonal forecasts in subsequent studies.

2. Methods and Datasets

2.1. Clustering

In order to process large-scale synoptic information well, it is useful to classify the atmospheric state into circulation patterns. This task is commonly performed by cluster algorithms (e.g., HOFFMANN and SCHLÜNZEN, 2013; LAUX et al., 2020) that reveal the main circulation patterns. In addition, cluster algorithms can detect and represent SST patterns. This allows to draw connections between different variables when a particular cluster of one variable interacts with another variable. Cluster algorithms group data according to their similarity to each other. Within the clusters, the individual cluster members should have the smallest possible distance to each other with a simultaneous maximum distance to the members of other clusters. There is a variety of clustering methods, in 2.2.2. the algorithm used will be discussed in more detail. Centroid-based clustering methods assign a point to each cluster based on the minimum possible distances between the data points and the centroid. After an iteration, the centroid moves to the center of the data points assigned to it and performs the assignment again. These iterations are repeated until no more change occurs or the pre-programmed iteration limit is reached. (UPPADA, 2014). Optimization Methods work similarly, randomly assigning clusters in the first iteration, then removing and reassigning one data point at a time in the second iteration and so forth.

2.1.1. Cost733Class

For clustering the variable fields the software Cost733Class Version 1.2 (PHILIPP et al., 2014) was used (hereafter referred to as COST). COST is software that focuses on creating and evaluating weather and circulation classifications using a variety of methods. It includes both predefined and undefined approaches. Thereby, the predefined approaches show higher persistence, interannual variation and long-term trends (PHILIPP et al., 2010). The software is also flexible with respect to the selection of input

data such as number of clusters, time period used or file format. But COST has the disadvantage that it does not adapt to the incoming data sets, but that the data sets have to be adapted to COST. So, the data sets fed in must have a certain format. This is done by means of the Climate Data Operators software (CDO). This in the climate modelling community widely established file manipulation tool (KASPAR et al., 2010) allows to convert the data fields downloaded from ERA5 into the required format.

2.1.2. SANDRA

Centroid-based or optimization methods are particularly suitable for finding basic patterns such as those in this work. These hard clustering methods work well with gridded data and have the tendency to form relatively evenly sized clusters. This prevents extreme events with a return period of several decades from being assigned to a cluster of their own. Thus, such an algorithm is well suited for the classification of weather situations (HOFFMANN and SCHLÜNZEN, 2013). However, the number of clusters must be determined before starting the calculation. The applied cluster method uses Simulated AN-nealing and Diversified Randomization, hence the name SANDRA. Compared to standard clustering methods like k-means, SANDRA has the advantage that the algorithm allows "wrong" assignments temporarily and thus, unlike k-means, comes out of local optima quite reliably and converges to the global optimum (PHILIPP et al., 2010). Freeing from the local optimum is reliably achieved by a control parameter T , which is stepwise reduced by a cooling factor C in each iteration.

$$T = T * C$$

This determines a probability P with which incorrect assignments are allowed.

$$P = \exp [(D_{cur} - D_{new})/T]$$

D_{cur} is the Euclidean distance to its current cluster and D_{new} the distance to the new cluster. If a random number is less than P , the data point is wrongly assessed. The cooling factor decreases relatively fast, so that quickly no more changes occur (PHILIPP et al., 2010). To counteract, SANDRA is performing one thousand runs. Therefore, in most cases, SANDRA finds better results than k-means (PHILIPP et al., 2014). In addition, SANDRA can process spatial data, as required in this work; and is also capable of clustering two variables at the same time. However, due to the one thousand runs, the computational cost is significantly higher. SANDRA is implemented in COST. More technical information about the SANDRA classification can be found in PHILIPP et al. (2010, 2014) and HOFFMANN and SCHLÜNZEN (2013).

2.2. Explained Cluster Variance and Adjusted Rand Index

To evaluate the cluster distributions, the Explained Cluster Variance (ECV) is calculated. It denotes the ratio of the squared distances of the points of a single cluster to the corresponding centroid (WSS) by the total sum of squares (TSS) (PHILLIP et al., 2014).

$$ECV = 1 - \frac{WSS}{TSS}$$

An ECV of 1 means a distribution perfectly explained by the clusters. If the ECV tends towards 0, the clusters represent the distribution poorly. With an increase in the number of clusters, the ECV increases irregularly, but mostly relatively logarithmically.

To compare the different clusters, the Adjusted Rand Index (ARI) is applied. The index compares the agreement between two series of numbers, in this case two series of cluster assignments. According to HUBERT and ARABIE (1985), the ARI is derived from the Rand Index according to RAND (1971) and supplemented by the general random correction. Because of the random numbering of the clusters by SANDRA, this calculation is particularly suitable, because for the ARI the numbers are considered only as a random cluster title, not as important numbering (SANTOS and EMBRECHTS 2009). Thus, the ARI is calculated from the quotient of the difference of the Index and the Expected Index with the difference of the Maximum Index and the Expected Index.

$$ARI = \frac{\sum_{i,j} \binom{n_{i,j}}{2} - \sum_i \binom{n_i}{2} * \sum_j \binom{n_j}{2} / \binom{n}{2}}{\frac{1}{2} [\sum_i \binom{n_i}{2} + \sum_j \binom{n_j}{2}] - \sum_i \binom{n_i}{2} * \sum_j \binom{n_j}{2} / \binom{n}{2}}$$

The notation $\binom{n}{2}$ corresponds to the calculation $n(n - 1)/2$. i denotes the numbering of the first cluster from one to the maximum cluster number, j accordingly that of the second cluster. n_i is the number of days that have been assigned to cluster i of the first cluster calculation, $n_{i,j}$ the number of days that have been assigned to both cluster i of the first cluster calculation and cluster j of the second cluster calculation. n corresponds to the total number of days.

An Adjusted Rand Index of 1 means a perfect match of the cluster assignments. If the index approaches 0, the cluster assignments are not influenced by each other. The larger the sample, the more likely values are close to 0 are, since a near-perfect match is increasingly unlikely.

2.3. Statistical tests

2.3.1. Pearson-Correlation and p-Value

The Pearson correlation, in the following referred to as correlation, compares two data series with the same number of data points and calculates a correlation coefficient for this. In contrast to the Spearman correlation, which only calculates with ranks of the variables, the absolute values are considered with the Pearson correlation. The correlation coefficient can take values between -1 and +1. A coefficient of -1 means a perfect linear negative correlation, +1 a perfect linear positive correlation. The closer the coefficient is to zero, the less similar the behavior of the correlated series. As the length of the data series under study increases, the coefficient approaches zero as the small differences between the curves add up. Therefore, the correlation coefficient alone is not meaningful. To overcome this problem, the coefficient is offset against the number of data points:

$$p = T - Distr \left(\frac{|k| * \sqrt{n-2}}{\sqrt{1-k^2}}, n-2 \right)$$

$T - Distr$ denotes the calculation of the Student-t-distribution for the two end faces in the bracket. p corresponds to the significance value p , also called p-value. The null hypothesis is that the data series do not correlate significantly with each other. If p falls below the specified significance level, the null hypothesis is rejected; the correlation of the data series is therefore significant (FISHER, 1922).

In this work, the correlations between different properties of clusters, teleconnections, and rainy seasons are calculated. The significance level is $\alpha = 0.05$, unless otherwise specified. Note that mostly correlation coefficients are argued without mentioning the corresponding p-value. Since all correlations are made over 39 years each and thus the series always have 39 data points, the correlation coefficient which remain beneath a p-value of 0.05 is constant. It is 0.317 and -0.317, respectively. All correlation coefficients above 0.32 and below -0.32 are therefore statistically significant.

2.3.2. Mann-Whitney-U-Test

The Mann-Whitney U test is a non-parametric statistical test to check for a bias between the values of a variable assigned to two categories, e.g., the assignment of days to two different clusters, tested for their daily mean temperature. The test is preferable to simply comparing the median or arithmetic mean because a statement of significance can be made. The null hypothesis is that there is no

significant bias between the two categories (NACHAR 2008). For example, if the days assigned to one cluster are significantly warmer than those assigned to the other cluster, the null hypothesis is rejected. The significance level in this work is set at $\alpha = 0.05$. The calculation consists of two parts. First, the U-statistic is calculated for the two categories:

$$U - Statistic_x = \left(\sum_{i=0}^n Rank_x \right) - \frac{n * (n + 1)}{2}$$

$Rank_x$ denotes the individual rank of the individual values of the variables that are assigned to the category x . For example, the days are ordered by temperature, the warmest day gets the rank 1, the second warmest the rank 2, etc. and the rank sum of the two clusters is calculated in each case. n is the number of x assigned values, the days in the respective cluster. Furthermore, a critical value is calculated. As this is not trivial, the values for different significance levels can be looked up in MILTON (1964). If one of the values of the U-statistics of the two categories is smaller than the critical value, the null hypothesis is rejected, i.e. a significant bias exists (MANN und WHITNEY, 1947).

2.3.3. Mann-Kendall-Trend-Test

The Mann-Kendall trend test looks at the correlation between the rank order of observed values and their sequence over time. Because the test is nonparametric, it works for non-normally distributed data. The null hypothesis for the Mann-Kendall test is that the data are independent and randomly ordered, i.e., there is no trend or serial correlation structure between the observations (HAMED and RAO 1998). The test compares each data point in the series with all data points following it and analyzes whether the difference is greater or less than zero in each case. If the point occurring later in the series is larger, +1 is assigned to the data pair, if it is smaller, -1. The assignment of all data pairs is summed (S). If $S > 0$, the series has a positive trend, if $S < 0$, a negative trend. The trend is significant with $\alpha = 0.05$ if:

$$S < 0: \frac{S + 1}{\sqrt{\frac{n(n - 1)(2n + 5)}{18}}} > 0,975$$

$$S > 0: \frac{S - 1}{\sqrt{\frac{n(n - 1)(2n + 5)}{18}}} > 0,975 ,$$

n is the number of data points (WANG et al., 2020). A disadvantage of the test is that only linear trends can be detected, polynomial trends like a decrease followed by an increase within a data set are not detected.

2.4. Drought indices

2.4.1. Standardized Precipitation Index

The Standardized Precipitation Index (SPI), introduced by MCKEE et al. (1993), describes the deviation of precipitation from the arithmetic mean for a given period, divided by the standard deviation. The mean and standard deviation are determined over the entire period. A normal distribution of the variables is assumed. For precipitation data, this is approximately the case, but not exactly. Therefore, the time series is first fitted to a gamma distribution, which is converted into a normal distribution in the next step. With this, the calculation can be performed for each time point t as described above:

$$SPI_t = \frac{tp_t - (\sum_{i=1}^t tp_i)/n}{std(tp)}$$

tp_t is the precipitation sum of the time point under study, tp is a list of the precipitation sums of all time points, std is the standard deviation, and n is the number of time points. The time unit is irrelevant for the calculation, whether monthly or yearly; a monthly analysis is common. The time steps, on the other hand, must be consistent and missing data should be avoided. The SPI has the advantage of being comparable in space and time because droughts indicated by it are relative to the mean precipitation of the area and are not tied to specific precipitation values. Theoretically, the SPI reaches values from $-\infty$ to $+\infty$, but in reality the values mostly range between -3 and +3. Higher values are usually only reached in deserts where precipitation falls in multi-year cycles. An SPI of 0 describes normal conditions, a SPI of below -1 denote a drought, values below -2 indicate an extreme drought. Similarly, values of +1 are indicative of very humid conditions (MCKEE et al., 1993).

2.4.2. Standardized Precipitation Evapotranspiration Index

The Standardized Precipitation Evapotranspiration Index (SPEI) calculates the potential evapotranspiration for each unit of time and subtracts it from precipitation. This provides a better representation

of the actual amount of water available than precipitation-only data such as the SPI. The potential evapotranspiration (PEV) is calculated according to VICENTE-SERRANO et al. (2010) as follows:

$$PEV = 10K \left(\frac{10T}{I} \right)^m$$

$$I = \sum \left(\frac{T_i}{5} \right)^{1.514}$$

T_i is the mean monthly temperature of the twelve months.

$$m = 6.75 \times 10^{-7} * I^3 - 7.71 \times 10^{-5} * I^2 + 1.79 \times 10^{-2} * I + 0.492$$

$$K = \left(\frac{N}{12} \right) \left(\frac{NDM}{30} \right)$$

NDM is the number of days of the month to be calculated, N is the maximum number of sunshine hours, depending on the latitude, which is calculated as follows:

$$N = \left(\frac{24}{\pi} \right) \omega_s$$

ω_s is the hourly sunrise angle.

$$\omega_s = \arccos (-\tan(\varphi) * \tan(\delta))$$

φ is the latitude in radians, δ is the declination of the sun in radians, which is calculated:

$$\delta = 0.4093 \operatorname{sen} \left(\frac{2\pi * J}{365} - 1.405 \right)$$

J is the average Julian day of the month.

Once the potential evaporation is calculated, it is subtracted from the monthly precipitation total to get tp_res , a simple measure of the water surplus/deficit:

$$tp_res = tp - PEV$$

In the next steps, the SPEI is calculated analogously to the SPI: The time series of tp_res values is fitted to a gamma distribution and then converted to a normal distribution. The arithmetic mean of the resulting variables over the time series is subtracted from the variable of the time step to be calculated, and this difference is divided by the standard deviation.

SPEI has the disadvantage that PEV cannot be calculated at temperatures below 0 °C, it is assumed that there is no evaporation. However, this is irrelevant in the study area, since the temperature values

are always above 7 °C. The advantage of the SPEI is that global warming is included in the consideration (VICENTE-SERRANO et al., 2010). A compound event of simultaneous heat and lack of precipitation significantly intensifies the dryness of droughts, especially vegetation is severely affected as the upper soil layer loses its moisture (ADAMS et al., 2009).

2.4.3. Daily Wetness Index and Daily Temperature Index

SPEI and SPI have the disadvantage that they are usually only used for monthly data. This limits the possibilities to analyze individual cluster days. To achieve this, on the one hand the days without precipitation (< 0.2 mm, in the following called dry days) can be worked out for the individual regions. On the other hand, the Wetness Index can be calculated. This calculates the average daily rainfall of a station/region, once regardless of the cluster assignment and once per cluster. The average rainfall of the cluster under consideration $\bar{z}_\alpha(x)$ is divided by the total mean precipitation $z(x)$.

$$I_{wet}(x) = \frac{\bar{z}(x)}{z(x)}$$

The Wetness Index can thus take values from 0 to ∞ . A value of 1 means that the cluster has normal wet conditions. If the Wetness Indices are higher, the cluster is drier than normal. If the Wetness Index reaches 2, it means twice as dry conditions of the cluster as normal. If the index reaches 0.5, it means that the cluster has twice as much precipitation as normal (BÁRDOSSY, 2010).

One problem of the Wetness Index is high seasonal variations of precipitation. When clustering, this usually results in two cluster groups, a dry season group and a wet season group, each with several clusters. By averaging the precipitation, one group will have values close to 0 and the other group will have values above 2. Thus, on the one hand, the difference between the clusters within the two groups is very small and difficult to interpret, and on the other hand, the two cluster groups are difficult to compare. To overcome these problems, the Daily Wetness Index (DWI) is developed. The DWI first calculates the average precipitation for each of the 92 days in the MAM season over the 39 years studied, either per country, grid point, or for the complete study area ($\bar{z}_d(x)$). The algorithm then goes through the clustered days, from 03/01/1980 to 05/31/2018, looking for days that are associated with the cluster of interest. Of these days, it notes the average amount of rain calculated for that day $\bar{z}_d(x)$ and separately the actual amount of rain that fell $z_d(x)$. Days that are assigned to another cluster are skipped. If all corresponding days are noted, the sum of the average and the actual precipitation is calculated. Then the sum of the average precipitation is divided by the sum of the actual precipitation.

$$DWI(cl) = \frac{\sum_d \bar{z}_{d,cl}(x)}{\sum_d z_{d,cl}(x)}$$

The meaning of the DWI values is identical to I_{wet} ; a DWI of 2 means twice as dry, 0.5 means twice as wet. By observing the seasonality, the clusters of the dry season become comparable with those of the rainy season.

Following the same procedure, the Daily Temperature Index (DTI) was calculated, only the variable differs. The DTI is intended to highlight clusters that are unusually warm or unusually cold. However, the DTI values are much closer to 1 than the DWI, because the relative deviations in temperature are not as large as in precipitation due to the use of the Kelvin scale. Therefore, a relative classification of the DTI values of the individual clusters is important.

2.5. ERA5

Most of the studies mentioned before are based on station data, especially with respect to precipitation. A re-analysis model is used less frequently, e.g. YANG et al. (2015) investigated precipitation in East Africa with ERA-Interim. An analysis using the successor model, ERA5, is not known. ERA5 is a re-analysis of global climatic conditions provided by the *European Centre for Medium-Range Weather Forecasts* (ECMWF) and made freely available to the research community (HERSBACH et al., 2020). In hourly resolution, ERA5 provides a wide range of meteorological variables for a period from 1979 to three months before the current date. This data set is continuously quality corrected by ECMWF. In addition, data are available from five days after the current date, but these are not quality checked. The grid resolution of ERA5 is $0.25^\circ \times 0.25^\circ$. The data set incorporates observational data from satellites, ground stations, radiosondes, ship and aircraft measurements. Among other things, this prevents drift in the model data and ties the analysis runs to fixed points (HERSBACH et al., 2020). Compared to its predecessor ERA-Interim, ERA5 has the advantage of a significantly higher spatial and temporal resolution (BALSAMO et al., 2015). For this purpose, the performance of ERA5 is significantly better than that of ERA-Interim in most regions of the world, so that the ERA5 values are now very close to the observed data (GLEIXNER et al., 2020; HERSBACH et al., 2020; TAREK et al., 2020). In particular, great progress has been made in surface solar radiation and in ozone distribution in the atmosphere.

The study period is set from 1980 to 2018. ERA5 data are hourly resolved, but for the cluster analysis daily data was used, because otherwise the clusters would reproduce the diurnal cycle and not the actually interesting recurring circulation patterns. Therefore, daily data were used. The rainfall corresponds to the cumulative daily precipitation, the temperature at 2 m height, the wind variables, the

surface pressure and the SST to the respective value at 11:30 am. In some cases, the gridded data are spatially averaged/aggregated, to calculate, e.g., the average precipitation of a country. The narrowing of the longitudes towards the poles was taken into account by a weighting grid which, when calculating the mean values, maluses the grid points located farther from the equator according to the factor of compression of the area. The rainy season of a country was defined as the time between the first and last occurrence of a rainfall above the median seven days in a row. The median is calculated from the 92 days of the MAM season over the span of the study period. To account for the early/late onset of the rainy season in some countries, a second median was calculated using year-round data.

2.6. Teleconnection data

In addition to the ERA5 data, the indices of various teleconnections are needed for the analysis of the results. Teleconnections describe the “contemporaneous correlations between geopotential heights on a given pressure surface at widely separated points on Earth” (WALLACE and GUTZLER, 1981). These pressure systems have a direct effect on climatic processes, so that the precipitation sum of the rainy seasons in many regions depends on a teleconnection. There are stationary teleconnections such as the El Niño Southern Oscillation (ENSO) and non-stationary ones such as the Madden-Julian Oscillation (MJO) (POHL and CAMBERLIN, 2006). The teleconnections are usually described by indices, which indicate the strength and direction of the deviation from the normal state. These indices are calculated for the stationary pressure systems via the SSTs, since the air pressure is linearly dependent on them (DAVIS, 1976). For this master thesis, ENSO, MJO, the Indian Ocean Dipole (IOD), and the western and eastern poles of the IOD, the Western and Eastern IOD (W-IOD and E-IOD) were considered. Effects of ENSO are best described by the Niño 3.4 index; thus, it was used. The IOD, W- and E-IOD data are from NOAA (2021a), the ENSO 3.4 data are from NOAA (2021b), and the Madden-Julian Oscillation data are from BOM (2021). The characteristics of those teleconnections and their possible effects on the study area are described in more detail in chapter 3.3.2.

3. Study Region

3.1. Overview

This master's thesis considers the area of the Greater Horn of Africa (hereafter GHA). More precisely, the area between 10.5° S and 14.25° N as well as 30° and 51.75° E is examined, shown in Fig. 1. This region comprises the African countries Ethiopia, Burundi, Democratic Republic of Congo, Djibouti, Eritrea, Kenya, Malawi, Mozambique, Rwanda, Zambia, Somalia, Sudan, South Sudan, Tanzania as well as Uganda. Completely within this area are Djibouti, Kenya, and Somalia, and Ethiopia, Tanzania, and Uganda are missing only 0.1; 5.9; and 3.0 percent of the country's area, respectively. In addition, small areas of Yemen, Dem. Rep. Congo, Zambia, Malawi, and Seychelles are located within the area. Since these countries are not counted as part of the Greater Horn of Africa, and since their share of the study area is in some cases negligible, they are not analyzed individually.

The focus on the Greater Horn of Africa is based on the need to improve the understanding of the Long Rains which cover most of this area. The region is hydrologically very vulnerable and millions of people are yearly affected by variations in the precipitation of the Long Rains. It coincides with the participation in the CONFER project, which aims to create and improve Climate Services for East Africa (for details see <https://confer-h2020.eu>). The study region includes parts of all countries involved in the project, namely Sudan, South Sudan, Uganda, Somalia, Kenya, Tanzania, Ethiopia, Djibouti, Rwanda, Burundi and Eritrea. For large-scale pressure patterns to be included in the analysis, the area must be appropriately large. However, the size is limited by the focus on Kenya and southern Ethiopia. If the study area is chosen too large, clusters may emerge that are determined primarily by conditions far off that have little effect on precipitation over the considered countries. In addition, the area is limited by the number of gridpoints, the computational effort increases substantially for larger fields. The northwest corner was chosen to accommodate Ethiopia almost entirely. If the boundary was pushed further north, larger parts of the Sudanese Sahara would affect the area, which should be avoided if possible because the extreme dry conditions can disrupt the cluster centroids. The southern and western boundaries were placed accordingly so that large parts of Tanzania also lie within the study area. The eastern boundary extends several hundred kilometers into the Indian Ocean. This is necessary due to the large scale pressure systems building up over the ocean such as the Indian Ocean Dipole (IOD) as well as the wind system caused by the ITCZ. In addition, clusters were formed based on Sea Surface Temperatures (SST), a larger water area is needed for this. However, the water area should not be dominant, so the boundary of 51.75° E was chosen. Overall, the area roughly corresponds to the size

used by YANG et al. (2015) from 12° N to 10° S and 30° to 52° E, who, however, did not perform a cluster analysis but a principal components analysis.

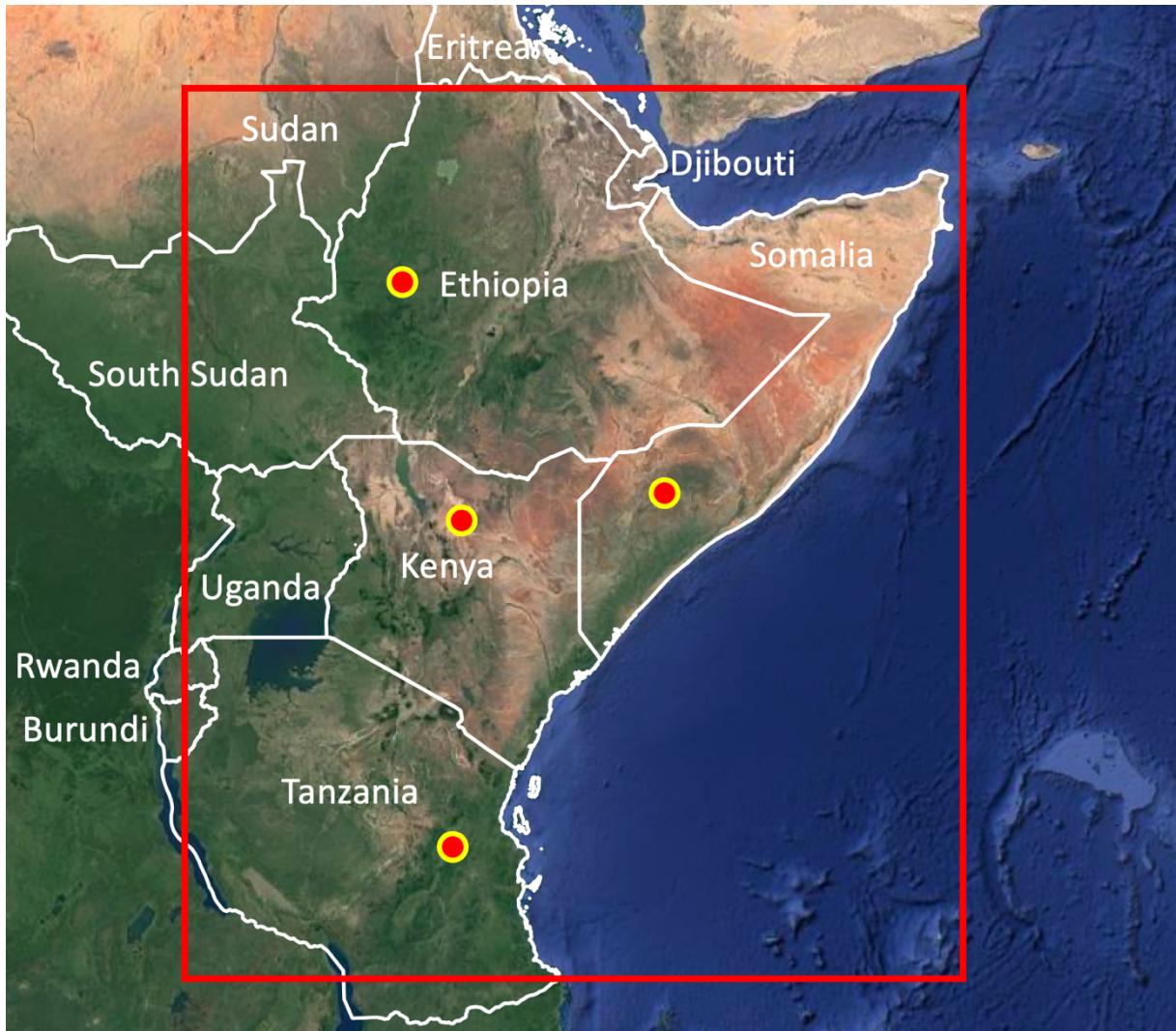


Figure 1: Study Area, from 14.25° N to 10.5° S and 30° to 51.75° E. The points indicate the cities of Nekemte (Ethiopia), Baidoa (Somalia), Marsabit (Kenya) and Morogoro (Tanzania) which are further examined in Fig. 3.

3.2. Topography

The heterogeneity of the study area creates some challenges. For example, elevations range from sea level to 5895 meters above sea level with Mount Kilimanjaro in Tanzania. In contrast, the other topographic feature, the Ethiopian Highlands, spreads across much of the north (see Fig. 2). With a maximum altitude of 4550 m a.s.l., the 750-hPa level is clearly exceeded here. In addition, the Highlands are very extensive, covering a large part of Ethiopia, their area roughly corresponding to the

dimensions of the Alps. Half of Ethiopia's land area is therefore higher than 1200 m a.s.l., a quarter is above 1800 m a.s.l., and more than 5 % of the country reaches altitudes above 3500 m a.s.l. The edges of the mountain range are very steep, like a plateau. This leads to convective rain especially on the south and east side of the Highlands. For example, the town of Nekemte, located in this region, has an annual precipitation of 5044 mm, which significantly exceeds that of the surrounding area (see Fig. 3 a). The variation in absolute rainfall is therefore very large here compared to the rest of the study area, which could significantly influence the clustering of rainfall, which needs to be verified. Without the Ethiopian Highlands, East Africa would be wetter overall because Westerlies would dominate, not the East African low level jet (YANG et al., 2014). But in the highlands themselves the intensity of precipitation would decrease as less convective rain falls (ENYEW and STEENEVELD, 2014).

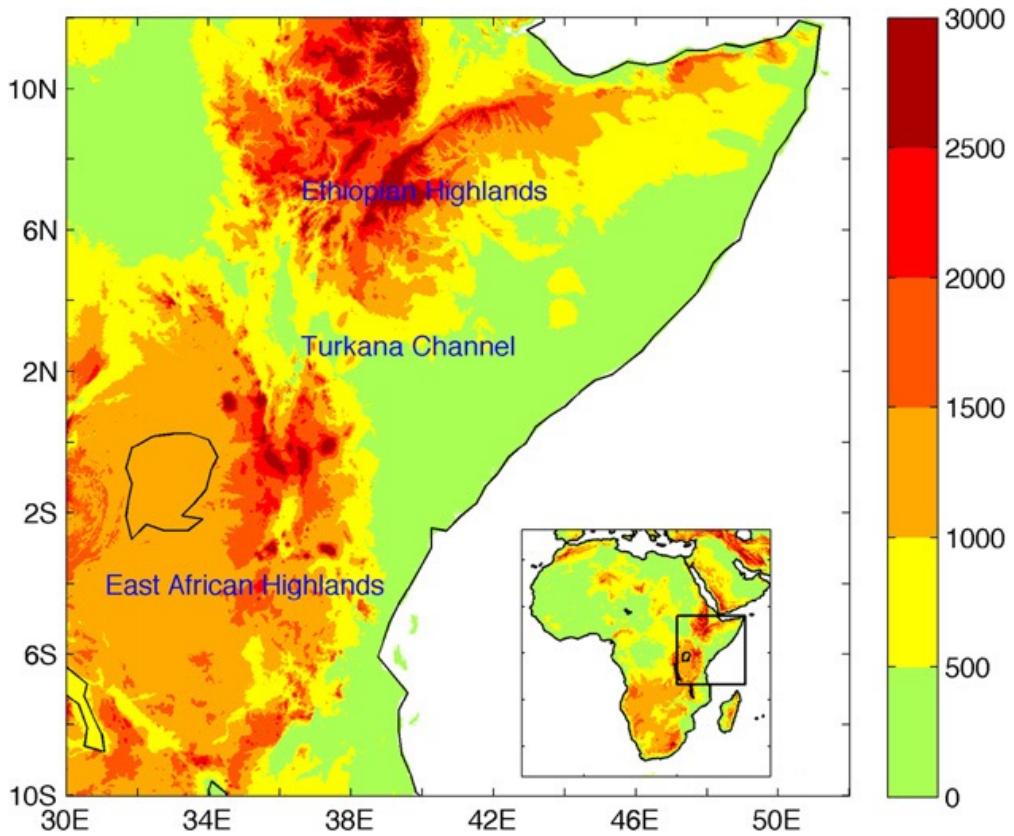


Figure 2: Topographic elevation map [m] of the study area. Taken from YANG et al., 2015.

The highest rainfall amounts are found in the Cherang'any Hills in Kenya, northeast of Lake Victoria, which are part of the East African Highlands. This mountain range, which runs in a north-south direction with a sharp break-off edge, is hit by south-easterly winds, especially in April and May. Due to the elevation differences of more than two thousand meters vertically within about three kilometers horizontally along about one hundred kilometers width of the mountain range, the humid air masses of the Indian Ocean are pushed up, so that large amounts of topographically enhanced precipitation fall. As a result, 4386 mm of rain falls there in the MAM season alone. Thus, this mountain range potentially

also has a strong influence on the clustering, but it is to be considered smaller than that of the Ethiopian Highlands because of the smaller area of the mountains. The highest mountains in the study area are located in the East African Highlands, the Mount Kenya massif at 5199 m a.s.l. and Mount Kilimanjaro at 5895 m a.s.l. Both mountain massifs have a great influence on the local precipitation (CHAN et al., 2007). However, since Kilimanjaro is rather prominent, the impact on the analysis is considered to be relatively small. Between the East African and Ethiopian Highlands the Turkana Channel is located, with slopes down to 350 meters above sea level. The Turkana Channel causes a strong jet effect of wind from east to west. In the MAM season, the mean wind speed in the valley is 6.5 m/s, more than twice as high as in the surrounding area (ERA5 data).

3.3. Climatology

The climate of the GHA is determined by the annual passage of the Inter-Tropical Convergence Zone (ITCZ) (DIRO et al., 2008; BAHAGA et al., 2015). The regions south of about 10° N are crossed by the ITCZ twice a year, once in about April and once in about October. This results in two rainy seasons in these areas due to convective precipitation near the ITCZ (see Fig. 3 b and c). The farther north the regions are, the more the two rainy seasons merge into one single rainy season in June/July (see Fig. 3 a). In the study area, this is true for the represented parts of Eritrea and Djibouti as well as northern Ethiopia. Due to the location of the study area in the tropics and partly in the subtropics, diurnal climates prevail. Temperature differs significantly due to elevation differences (cf. Fig. 3 a and b). Maximum temperatures are reached between March and July, depending on latitude. While northern areas such as northern Ethiopia have a clear maximum (see Fig. 3 a), regions near the equator have two months of local maxima, one around March and one in October (see 3 c, d).

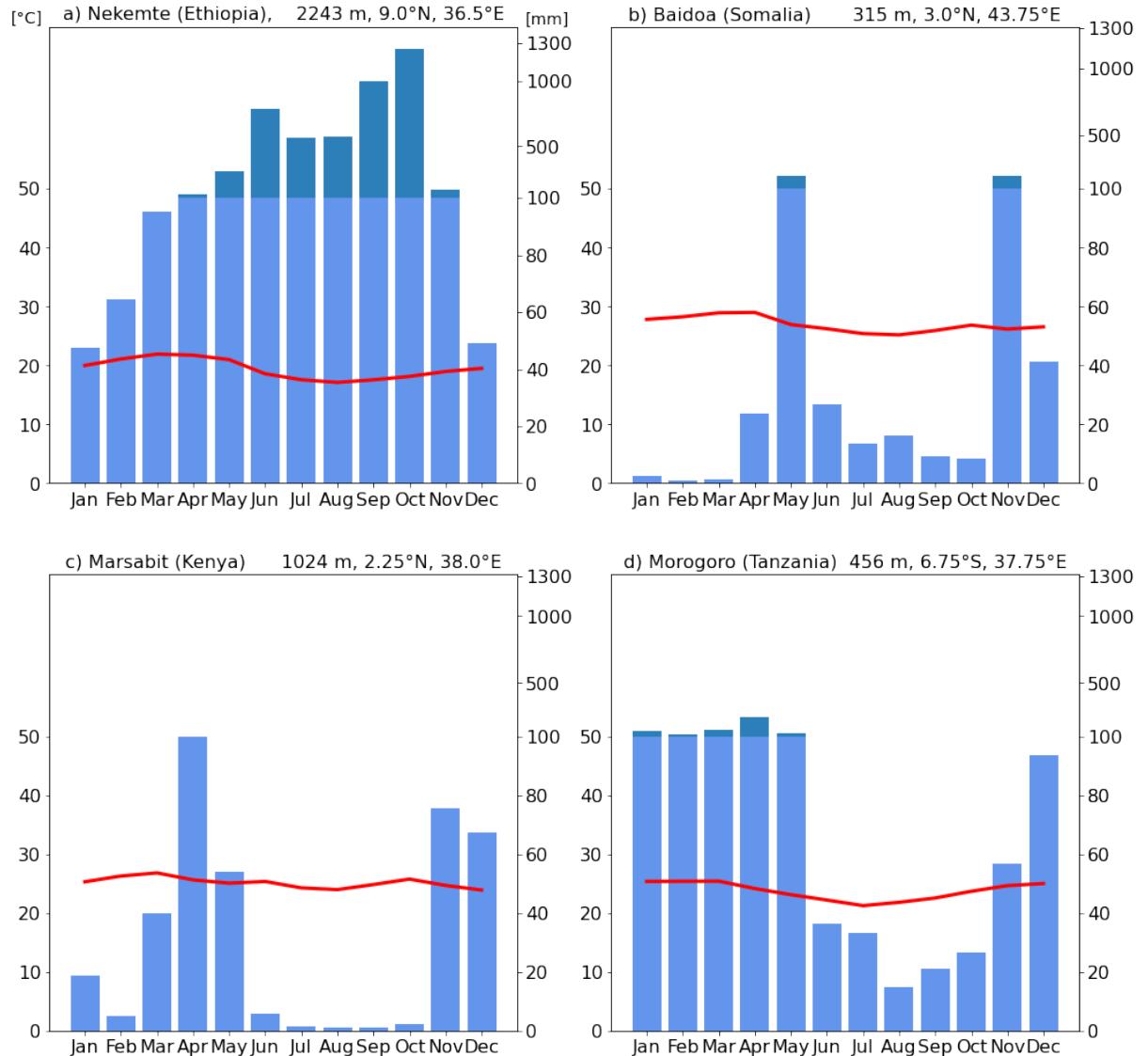


Figure 3: Climate diagram of four selected cities of the GHA. a) is located in the Ethiopian Highlands, b) in relative coastal proximity in Somalia, c) in the plains of Northern Kenya and d) in Eastern Tanzania, thus four different climate zones. The location of the cities is depicted in Fig. 1. The blue bars show the precipitation, from 100 mm/month on the bar is compressed by a factor of 20 for clarity, recognizable by the darker blue color. The red line shows the monthly average temperature. The data basis is ERA5 data, the reference period is 1980–2018.

3.3.1. Rainy seasons

The focus of the work is on the investigation of the so-called “Long Rains”, which describe the rainy season during the months of March to May. At this time, the ITCZ lies over East Africa, so that convective precipitation significantly increases the absolute amount of rainfall (CAMBERLIN and PHILIPPON,

2002). Here, the wind comes from the east and brings the moist air of the Indian Ocean to East Africa. The Long Rains are the more dominant rainy season in Kenya, Somalia and parts of southern Ethiopia (see Fig. 3 b and c). It is locally known as “Belg” and ends the dry period lasting from October onwards (“Bega”) (GEBREMICHAEL et al., 2014). Moreover, in Tanzania, the local rainy season is most pronounced in the MAM season (see Fig. 3 d). Consequently, the countries of focus for this study experience heavy rainfall during these three months. By staking out the study area based on latitude and longitude, other countries lie partly or entirely within the region under analysis, such as Eritrea. Their rainy seasons only partially coincide with the MAM season. Nevertheless, they are included in the analysis because it is beneficial for understanding the climate system of the region as a whole. For example, the countries of Djibouti and Eritrea do not have a strong rainy season; they are desert climates with about 225 mm of rainfall per year. April is the second rainiest month in Djibouti with 25.8 mm and the third rainiest in Eritrea with 27 mm; August is the maximum in each case. Sudan, with 500.7 mm of precipitation per year, is exactly on the border of classification as semi-desert. It has a rainy season that begins at the end of May and reaches its peak in August. Thus, only the start of this rainy season can be analyzed in this work due to the determination of the MAM season.

Table 1: Values of the individual countries, the mean of the Indian Ocean and the entire Study Area with respect to the variables total precipitation (tp), average temperature at 2 m (T2m), maximum temperature (Tmax) and minimum temperature (Tmin). The variables marked with MAM are the average of the years 1980–2018 over the months March to May, ann stands for the calculation over the whole year.

	<i>Burundi</i>	<i>Djibouti</i>	<i>Eritrea</i>	<i>Ethiopia</i>	<i>Kenya</i>	<i>Rwanda</i>	<i>Somalia</i>
<i>tp ann [mm]</i>	1294,3	223,7	205,8	1006,3	717,3	1054,0	297,9
<i>tp MAM [mm]</i>	461,6	64,9	59,1	290,9	271,7	398,9	137,2
<i>T2m MAM [°C]</i>	20,2	28,2	29,0	24,5	25,9	20,5	27,4
<i>Tmax MAM [°C]</i>	24,9	33,4	33,7	30,1	31,1	25,1	32,8
<i>Tmin MAM [°C]</i>	16,2	23,4	24,9	19,4	21,6	16,5	22,9

	<i>Tanzania</i>	<i>South Sudan</i>	<i>Sudan</i>	<i>Uganda</i>	<i>Indian Ocean</i>	<i>Study Area</i>
<i>tp ann [mm]</i>	1077,9	1042,1	500,7	1591,6	681,8	810,3
<i>tp MAM [mm]</i>	394,8	244,6	57,1	533,8	270,2	304,5
<i>T2m MAM [°C]</i>	22,6	29,5	31,3	23,2	27,4	26,1
<i>Tmax MAM [°C]</i>	27,2	35,0	38,3	27,7	28,1	29,8
<i>Tmin MAM [°C]</i>	18,7	25,0	25,0	19,5	26,5	22,9

During the three months, the ITCZ moves from south to north through the study area. In the period of the Long Rains, about 40 % of the annual precipitation falls in Kenya, Tanzania and Somalia (see Tab. 1). Averaged over the entire area, the highest amount of precipitation falls in late April/early May. This is because the rainy season is strongest pronounced at this time in Ethiopia and the difference between the wet and dry seasons is greater at these latitudes than at the equator, so it is more severe. Compared to the second rainy season, the so-called "short rains" from October to December (see Fig. 3 b and c), the long rains tend to last somewhat longer and are more intense. The passage of the ITCZ also shifts the main wind direction at the 750-hPa level, from north-easterly before to south-easterly after. This occurs particularly strongly in Ethiopia, Somalia and over the Indian Ocean, in Kenya the effect is still noticeable but weaker (ERA5 data).

3.3.2. Teleconnections and their indices

The climate of the Greater Horn of Africa is partly determined by different teleconnections, therefore the most important ones and their associated indices are briefly explained. The teleconnections mentioned in this section are tested for different correlations with the clusters in chapter 4.6.

The most prominent teleconnection is the El Niño Southern Oscillation (ENSO) (WOLTER and TIMLIN, 2011). For indicating the occurrence of El Niño and La Niña phenomena, the Niño-3.4 index is commonly used, which describes the SST anomalies in the region 5° N - 5° S, 170° W - 120° W using a five-month running mean (BAMSTON et al., 1997). When interpreting the Kiremt Rainfalls (June to September) in the north of the study region, the Niño 3.4 already achieves good results (GLEIXNER et al., 2017), and possibly this is transferable to the MAM season. An El Niño event is defined as exceeding the Niño 3.4 index by 0.4 K for six or more consecutive months, La Niña events at -0.4 K. ENSO events have an impact on the climate of the GHA in that El Niños are associated with a more pronounced rainy season during the OND season ("short rains") and La Niña is associated with reduced precipitation during this time (DIRO et al., 2008). However, this connection is rather weak and not directly transferable to the MAM season. The effect on the Long Rains will be investigated in more detail in this paper. As an example of an influence on the Long Rains, the strong El Niño in 1998/99 can be mentioned, which led to a significant drop in MAM precipitation in East Africa (LYON and DEWITT, 2012). On a decadal scale, variations in short rains are better described by the Pacific Decadal Oscillation (PDO). In addition, the decline in MAM precipitation after 1998 is also attributed to the PDO (BAHAGA et al., 2019). Thus, this index, which describes the deviation from normal SST in the North Pacific ($> 20^{\circ}$ N), is also included in this study.

More strongly than with the Pacific pressure systems, precipitation interacts with the Indian Ocean Dipole (IOD), especially during the Short Rains (Cook and Vizy, 2013). The IOD describes the anomaly of the SST between the Western Indian Ocean Dipole (W-IOD) and the Eastern Indian Ocean Dipole (E-IOD), it is formed from their difference:

$$IOD = \text{Western IOD} - \text{Eastern IOD}$$

The W-IOD describes the temperature deviation from the mean state off the coast of East Africa and the E-IOD the temperature deviation off Indonesia. A strong IOD event occurs when the W-IOD is warmer and the E-IOD is cooler than normal, resulting in a difference of at least 0.4 K from normal over a period of more than four months. In addition to the IOD, the W-IOD and E-IOD themselves may also have an influence on precipitation in East Africa. When SST is high in the W-IOD region, more ocean water evaporates and is driven toward the African mainland by easterly winds. Another influence may be the resulting more pronounced or smaller differences between the land and sea temperatures, especially on the wind systems (BAHAGA et al., 2015). Because of its proximity to the study area, W-IOD is potentially more important to the climate of the GHA than E-IOD. IOD and ENSO interact with each other only slightly. However, when an ENSO as well as an IOD event occurs in a year, the periodicity and strength of the IOD is clearly influenced by ENSO (BEHERA et al., 2006). Another pressure system over the Indian Ocean is expressed by the Somali Jet. This strong wind band at about 800 hPa level is especially important for precipitation during the Short Rains, as it brings larger masses of humidity to East Africa (BAHAGA et al., 2019). Since the Somali Jet blows mainly from June to October, its influence on the precipitation of the Long Rains is, however, rather small.

Another influence on MAM precipitation is the Atlantic Ocean off West Africa's coast. Water temperatures there may reflect the role of SST in regulating the strength and extent of the northward migration of the ITCZ. However, the correlation is rather weak (KORECHA and BARNSTON, 2007). Nevertheless, since the influence should not be neglected, an index was created to represent the gradient between the SST off the South American coast and that of the Gulf of Guinea. The index, abbreviated ATL in the following, describes the deviation of the temperature difference between 5° N, 0° E and 5° S, 25° W from the mean difference of the years 1979–2018, based on ERA5 data. In addition, to determine the absolute deviation of ocean temperatures from the mean of this region, the eastern point was considered in more detail. The index abbreviated E-ATL describes the monthly deviation of SSTs at 0° N, 0° E from the long-term (1979–2018) means of the corresponding months, analogous to E-IOD.

While the teleconnections mentioned so far are stationary, the Madden-Julian Oscillation (MJO) is a pressure variability that moves around the Earth along the equator. For this purpose, the MJO does not occur seasonally like e.g. El Niño phenomena, but has a 30–60 day cycle with a wet and a dry phase.

The wet phase in particular has an influence on the Long Rains of East Africa; it can prolong wet periods and make heavy rainfall events more extreme. Due to the mountains in the region, the humid air masses transported by the MJO are forced to rise and rain down (POHL und CAMBERLIN, 2006).

In summary, while the impact on Short Rains has been well studied and clear relationships can be drawn between teleconnections and precipitation distribution, this is not the case for Long Rains, the relationships are more ambiguous (WALKER et al., 2019). Thus, IOD and ENSO in particular have a significantly weaker impact on precipitation during the Long Rains than during the OND season.

3.4. Preselection of variables

ERA5 provides a large number of variables on 137 pressure levels as well as surface parameters. Testing all available options is not practical with such a large number of options. Therefore, a literature search was conducted to find the variables and pressure levels that could have an impact on precipitation in the region of the GHA ant that could be useful for the cluster analysis.

Sea surface temperature (SST), through its influence on pressure patterns, has a significant impact on the dynamics of climate events globally (PAI and RAJEEVAN, 2005). Furthermore, the teleconnections are formed by the fluctuations of the SSTs. The SSTs off the coast of Somalia and Kenya have a clear influence on the short rains (BAHAGA et al., 2015) and on the entire year (YANG et al., 2014), an influence on the long rains can therefore not be excluded. Thus, it makes sense to include this variable in the study.

To cover the wind conditions near the ground, the 900-hPa level is included in the study. Due to wind rotation, there is no clear better wind variable, so both u- and v-wind are considered. Since the influence of the Somali Jet on the May precipitation has not yet been clarified (VIGAUD et al., 2017), the v-wind is considered as a predictor variable, since it can be well described by it. The maximum of the Somali jet is found at about 800 hPa. In addition, Easterlies at 700 hPa have an influence on the countries near the equator (OKOOLA 1999). The 750-hPa level is considered as a compromise. Since strong westerlies over the western Indian Ocean suppress atmospheric convection (BAHAGA et al., 2015), the u-wind component should be included in the preselection. A level favorable for describing the westerlies is 350 hPa, since they blow strongly at this height over East Africa (BAHAGA et al., 2015). In their analysis, CAMBERLIN and PHILIPPON (2002) used the 500 hPa level, which is why this height is also checked in this investigation. Due to the topography of the study area, some of the grid points of the lower wind levels are only virtual. Therefore, wind near the ground (10 m above the ground) is used as an additional variable to investigate a wind variable that follows the topography. The u and v wind

components can be used to calculate the wind direction. This could be important because it includes both aspects in one value. Thus, phenomena such as a for the season atypical coast-parallel orientation of the wind could be better detected than with only u- and v-wind components. Consequently, the wind variables are wind direction, u-wind, and v-wind at the 900, 750, 500, and 350 hPa pressure levels, respectively, and at ground level.

Pressure systems such as a pressure bridge or trough have a significant influence on the large-scale precipitation of an area (z.B. LAUX et al., 2020), consequently, the inclusion of pressure in the cluster analysis is reasonable. In addition, the W-IOD is located at the edge of the study area, so its influence by the pressure systems may be better understood with the cluster analysis. Because of the large elevation differences, Surface Pressure (sp) should be preferred over Mean Sea Level Pressure (MSLP). Since the clustering only shows the differences from normal, it is not significant that the absolute sp values of the mountains and the sea are so significantly different. Moreover, the magnitude of the differences in the bottom air pressure could also be measured in the wind strength. Further, since it interacts with or consists of u and v winds, wind strength at 10 m above ground level (ws10) is considered as a linking variable.

In summary, the following variables were considered: u-wind and v-wind at 900, 750, 500, and 350 hPa levels and at 10 m above the ground surface, wind direction at 900 hPa, 750 hPa levels, and 10 m above the ground surface, sea surface temperature (SST) and atmospheric pressure at NN, and wind strength at 10 m, plus, of course, total precipitation.

4. Results

The Python codes and excel sheets used to calculate the results can be viewed at https://github.com/JanNWeber/Master_Thesis_Data.

4.1. Underlying trends of the variables

Before the clustering is carried out, a first trend analysis of the variables as well as of the drought indices is performed. Thus, the picture of the climatology of the study region from chapter 3.3 is extended. It also contributes to the understanding of the climate system. The analyses were carried out with the help of the ERA5 dataset, and the mentioned significances are confirmed by the Mann-Kendall trend test. If the significance level α is not defined separately, it is $\alpha = 0.05$.

Global warming is well reflected in the GHA data. In all countries, average temperatures are increasing significantly. Tab. 2 shows this for the MAM season. While over the sea, the warming of 0.48 K was slightly weaker than the global warming over the same period (about 0.5 K). When considered over the whole year, the warming of the land masses is stronger with 1.06 K. The warming of 0.85 K over the whole year is also stronger than the global warming over the same period (about 0.5 K). The MAM season warms less strongly (0.39 K over the sea, 0.80 K overall). The warming is also reflected in an increase in particularly hot days (see Tab. 2).

Table 2: Overview of the trends of selected variables and precipitation indices for the countries. If the arrows point up or down the trend is significant at $\alpha = 0.05$, if they point to 45° or 135° the trend is significant at $\alpha = 0.10$. The colors are used for clarification.

MAM-Season	Burundi	Djibouti	Eritrea	Ethiopia	Kenya	Rwanda
Temp.	↑	↑	↑	↑	↑	↑
Prec.	↓	↓	↓	↓	↓	↓
SPEI	↓	↓	↓	↓	↓	↓
Drought Days	↑	↑	↑	↑	↗	↑
Days above 80% Prec	↓	↘	↓	↓	↗	↓
Days above 95% Temp	↑	↗	↑	↗	↑	↑
MAM-Season	Somalia	Tanzania	S. Sudan	Sudan	Uganda	
Temp.	↑	↑	↑	↑	↑	
Prec.	↓	↓	↓	↓	↓	
SPEI	↓	↓	↓	↓	↓	
Drought Days	↑	↑	↑	↑	↑	
Days above 80% Prec	↓	↗	↓	↓	↓	
Days above 95% Temp	↑	↑	↑	↑	↑	

In all countries except Djibouti and Ethiopia, there are significantly more days per year above the 95 % percentile by the end of the study period. This trend is almost congruent when the 90 % or 75 % percentile is used instead of the 95 % percentile. The increase in days per season above the average temperature is even more clear, for all countries significant at a level of $\alpha = 0.015$. The higher temperatures lead to more convective upwelling, which in turn causes the air masses at ground level to flow more strongly. This is noticeable in a significant 10 % increase in wind speed near the ground. With respect to the days below the 5th percentile, a significant trend towards fewer cold days can be detected in all countries except Djibouti and Somalia, and below the 10th percentile as well. Consequently, the range between extremely warm and extremely cold days does not open up significantly. This agrees with the results of OMONDI et al. (2014). The warming trend can also be seen in the SST. The SST of the Indian Ocean increases continuously, by about 0.47 K during the study period.

In addition to warming, particularly critical for the hydrologically vulnerable region of the GHA is the trend of decreasing precipitation. On average, there is 3.6 mm/yr less rainfall in the study area. The decrease over the whole year is significant in all countries except Kenya, Tanzania and Ethiopia, thus the already vulnerable countries such as Djibouti, Eritrea, Sudan and Somalia are particularly affected by lack of rainfall, also confirmed by YANG et al. (2015). The precipitation brought by the Long Rains in the MAM season also decreases significantly with 1.4 mm/yr. All regions are affected (see Tab. 2), and the significance of the trend reaches values below $\alpha = 10^{-7}$ in all countries. This is also manifested in a significant decrease in the frequency of extreme precipitation, most strongly observed in the 80th percentile, but also in the 90th and 95th percentiles. Eritrea, Sudan, and South Sudan are most affected in this regard. The intensity of the season's rainiest precipitation day varies over the years, however, and no significant trend can be derived in any of the countries. In addition, the number of days without precipitation and those with particularly hot days increases significantly (see Tab. 2). This cannot be considered in isolation, as the number of dry heat days above the 90th percentile of temperature with less than 0.2 mm of precipitation at the same time increases significantly in Djibouti and Eritrea, and the driest regions consequently become even drier and hotter. Accordingly, the SPEI decreases by several percent in each country as both variables move in a negative direction (see Fig. 4), particularly strong in countries already affected by precipitation scarcity. In all countries, the trend of decrease in SPEI is significant, especially strong in South Sudan. Overall in the study area, SPEI decreases on average by 5.1 % per decade in March, 3.8 % per decade in April, and 7.2 % in May. To this end, positive values are achieved much less frequently from 1992 onward, and outliers to an SPEI of -1 become more frequent. While the five-year mean in Fig. 4 tends toward 1 in the 1980s, from 2000 to 2017 all but one country is in the significantly dry range. The years 2000–2010 were severely affected by drought. The last year of the curve in Fig. 4 rises due to extreme precipitation with flooding in Kenya and surrounding countries in April 2018 (KILAVI et al., 2018).

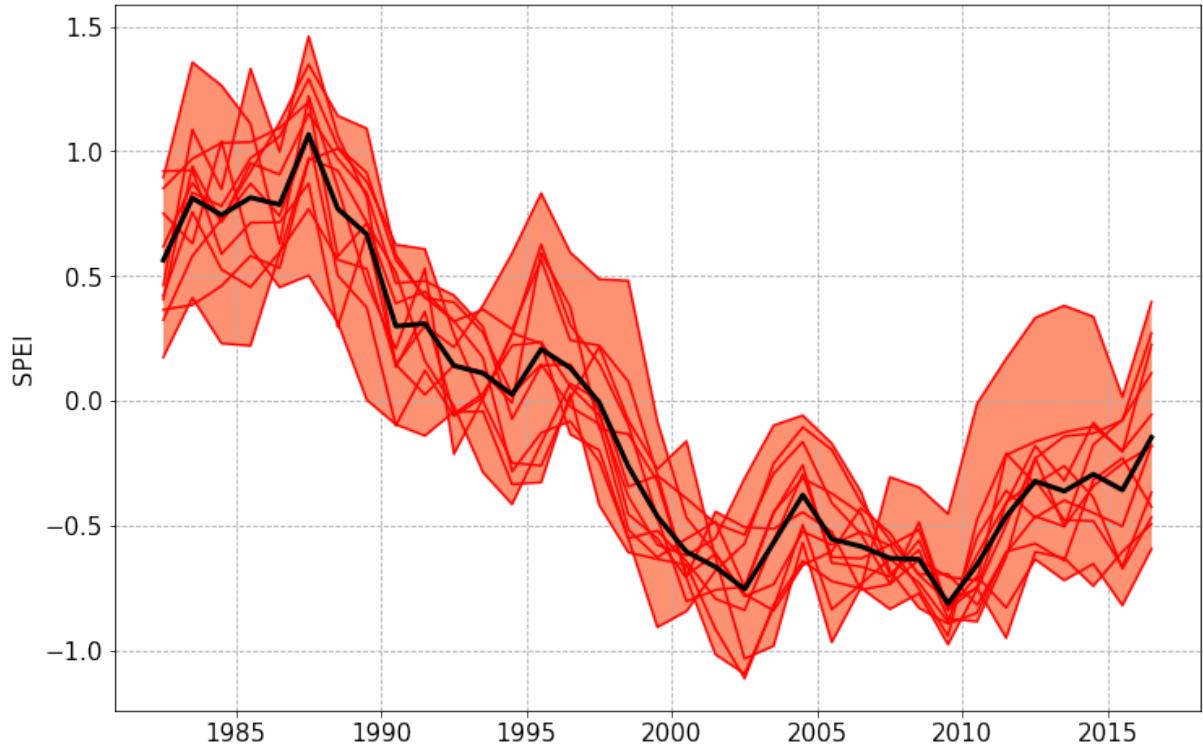


Figure 4: The Standardized Precipitation Evaporation Index (SPEI) of the countries for the MAM-Season. The SPEI is calculated using a five-year moving window. The black line is the mean of the countries. Values of zero indicate the average SPEI of the respective country of the 1980–2018 period.

More precisely analyzed, the reduced precipitation has several reasons. On the one hand, the dry days (< 0.2 mm precipitation) become significantly more frequent in each region (see Tab. 2 and Fig. 5). In all countries except Kenya, the decrease is significant according to the Mann-Kendall test. The already dry regions of Sudan, Eritrea, and Djibouti are also particularly affected. Each has more than 12 more dry days in the MAM season in 2018 than 39 years ago (see Fig. 5 the top lines). Another reason is the decline in the frequency of heavy precipitation. Whereas in the 1980s there were still an average of 14 days with precipitation averaged over 5 mm across the entire region, since 1992 there have only been four years in which more than 12 such events occurred. 2018 is a clear exception with the most of these events. In contrast, there is no general trend in the intensity of heavy rainfall, with only Eritrea and Sudan showing a significant trend towards less extreme rainfall.

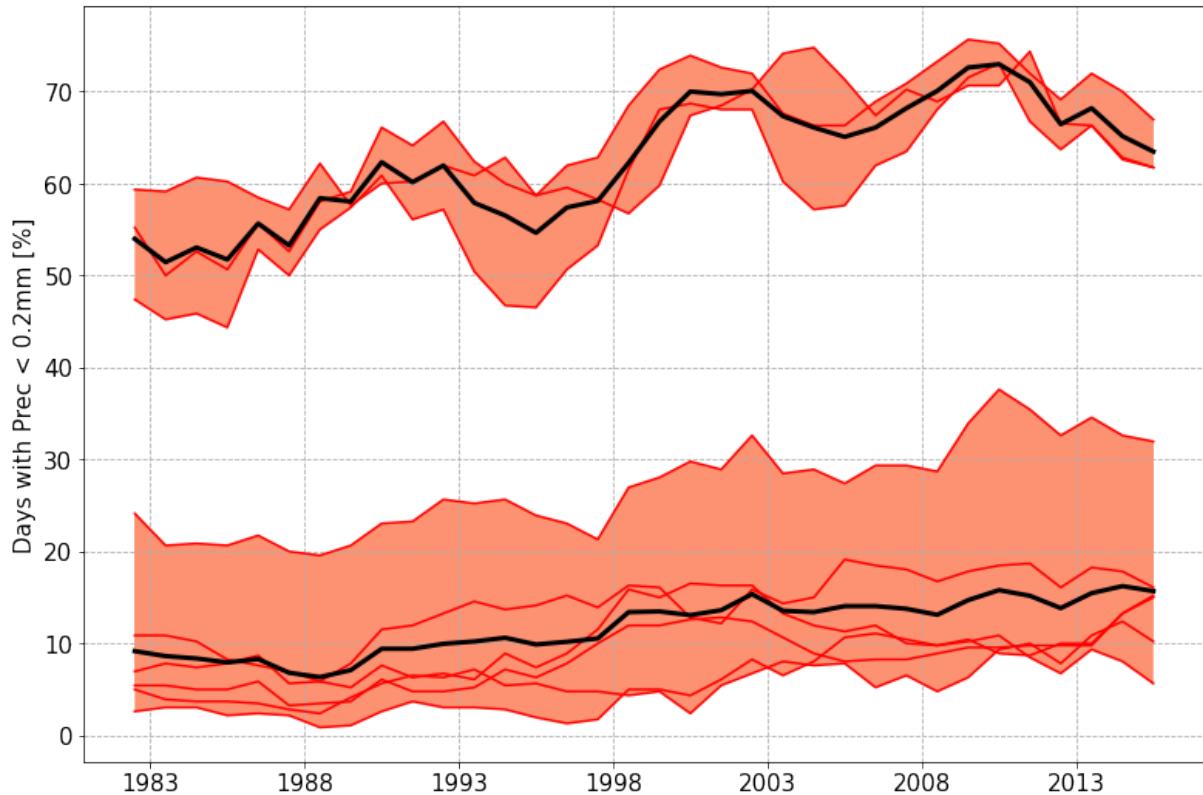


Figure 5: The percentage of days per year per country with a precipitation of less than 0.2 mm in the MAM Season, calculated using a five-year window. The upper countries are Djibouti, Eritrea and Sudan. They are separated because their rainy season does not begin until mid-May, so that there is already a base load on dry days. The black lines are the average of the respective country groups.

The third factor in the reduction of precipitation in the MAM season is that the onset of the rainy season occurs later and later in most countries (see Fig. 6). The onset was defined as precipitation above the median seven days in a row. The percentile of precipitation of the MAM season was used and only during these three months the start and end of the rainy season was examined. To avoid errors arising because of this boundary, the same analysis was performed with year-round data. While the trend toward a later onset of the rainy season is evident in all countries except Tanzania (see Fig. 6), it is significant in Djibouti, Eritrea, South Sudan, and Sudan. This corresponds to the four northern countries. In addition, the annual trend of Uganda and Rwanda is significant, their rainy season tends to start before March 1st, which is why for these countries the calculation included the months of January and February. Contrary to expectations, the end of the rainy season does not shift backwards in terms of the calendar, but rather forwards. This applies to all countries except Sudan (see Fig. 6), and the trend is significant in Burundi, Kenya, Rwanda, Tanzania and Uganda.

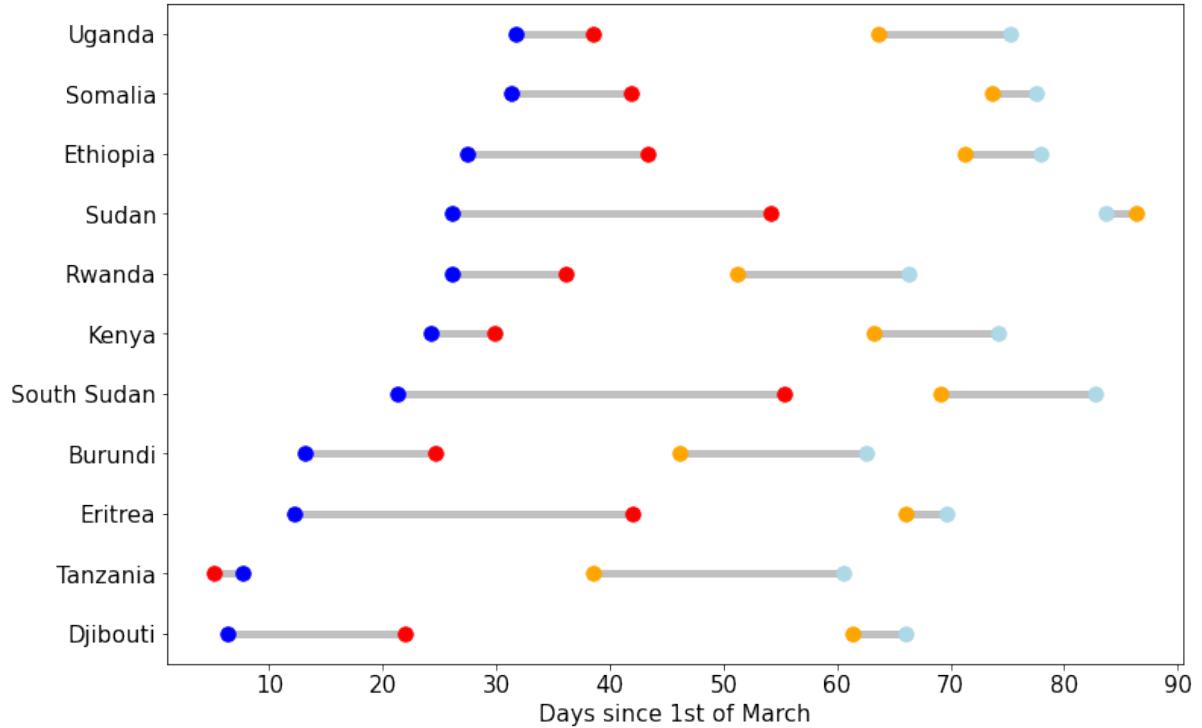


Figure 6: The shift in the start and end of the rainy season per country. The date of the start is determined by a country-average precipitation of more than the median from MAM-precipitation for at least seven days in a row. The dark blue dots show the date of the start of the rainy season at the beginning of the study period, determined by the starting point of a trend line. The red dots show the date of the start of the rainy season from the end of the study period, determined by the ending point of a trend line, to be read in days since the first of March. The light blue dots show the date of the end of the rainy season at the beginning of the study period and the orange dots the end at the end of the study period, determined in the same way. The dates are calculated with the start and end point of a linear trend line. The order of the countries on the y-axis has no deeper meaning.

Overall, the rainy season thus becomes shorter over the study period in many countries, as start and end move closer together. This is significant in the south (Burundi, Rwanda, Tanzania, Uganda) and in the northwest (Sudan, South Sudan); the annual data confirm this. For the remaining countries, the significance just does not reach $\alpha = 0.05$; they range from 0.05 to 0.15. In South Sudan and Sudan, the length of the rainy season correlates significantly with W-IOD (-0.60 and -0.40, respectively); consequently, the rainy season there tends to be shorter with warm water temperatures in the Indian Ocean. In Eritrea, on the other hand, ENSO correlates with length (0.44). There is no clear trend in the duration of dry periods within the rainy seasons (consecutive days below the 50th percentile between the beginning and end of the rainy season), except for South Sudan, whose dry periods become significantly shorter (correlation with ENSO: -0.42). The date of occurrence of the day with the highest amount of precipitation and that of the day with the maximum temperature also do not follow a significant trend in any country.

4.2. Selection of the number of clusters

Since SANDRA is a clustering method with cluster sizes to be defined in advance, the best cluster size must be determined first. To this end, it must be mentioned that there is no objective best cluster size, depending on the consideration of the results, different conclusions can be drawn. It is a good idea to look at the Explained Cluster Variance (ECV) to ensure that the cluster centers are as close as possible to the initial conditions of the days being calculated. The higher the ECV, the better. Here, the ECV increases as the number of clusters increases, first rapidly, then more slowly. The Elbow criterion can be used to estimate the point at which the slope of the curve flattens out significantly. It is important to find a balance between getting the ECV as large as possible and keeping the number of clusters as small as possible so that the clusters can still be easily distinguished from each other. The latter limits the maximum number of clusters to less than twenty, rather less than fifteen. In addition, a minimum of about eight clusters is necessary to prevent dissimilar patterns from being assigned to a single cluster. Only the days of the MAM season are clustered, not the complete years. This prevents that there are only 3–4 clusters per season, which would have made the interpretation very difficult, since mainly the rough course of the year would have been depicted, the details would have been blurred.

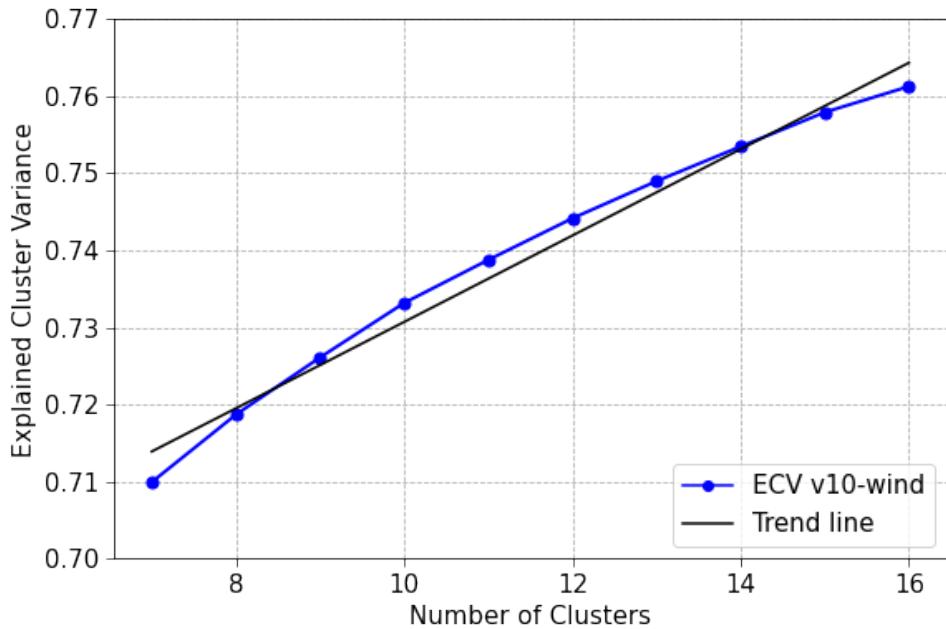


Figure 7: Explained Cluster Variance (ECV) overview for different number of clusters. The variable is v-wind at 10 m height. The clustering was done with the SANDRA algorithm. An ECV close to 1 corresponds to a very good clustering, close to 0 to a very bad one.

Figure 7 shows the ECV versus cluster count for the v-wind at 10 m height. The v-wind was used due to the pre-selection in chapter 4.1. It confirms the set frame between eight and fifteen, there is a relatively clear dropoff from eight to seven and the increase in ECV is only relatively weak from fifteen

on. A clear elbow criterion cannot be discerned; the curve is relatively regular in logarithmic form. In order to find the most favorable number of clusters, further investigation is therefore necessary. A suitable criterion is the Rand Index, since it can determine the similarity between the cluster calculations of different variables. One of the variables in this case is always precipitation, since this is the focus of the work. The more similar the clusters are to each other, the better air pressure or wind patterns can be related to rain patterns. Therefore, the Rand Index goes beyond just the performance of the cluster software and directly evaluates the impact of the clusters on precipitation. Because SANDRA randomizes cluster numbering, the Adjusted Rand Index (ARI) is used because it works independently of numbering (SANTOS und EMBRECHTS 2009). If the datasets are compared directly, the correlation between the precipitation and the variable on the day the data are collected by ERA5 becomes apparent. To capture the influence of the variable clusters on future precipitation, the precipitation clusters can be shifted by a specified number of days. The resulting ARI score then reflects how much of an impact the variable cluster has on precipitation for that particular day in the future. Since only three months of the year are examined in more detail, it makes sense to limit the shift to less than one month, because otherwise too much data, in this case all March precipitation data and all May variable data, would be dropped from the analysis. For atmospheric variables such as wind data that change on short-term time scales, shifts are initially analyzed primarily for two weeks because the influence is expected to be limited to that time period. In Fig. 8, an average of the ARI values of this two-week period is taken to get a first overview of which cluster number of precipitations with which cluster number of variables, in this case v10 wind, achieves the highest Adjusted Rand Index. The finer differences between the ARI of the unshifted clusters and that of the clusters shifted by 14 days will be discussed in more detail in chapter 4.3.

Through Fig. 8, it is clear that for precipitation, a number of eleven clusters works best. If fewer clusters are used, there is a significant drop in the ARI, which draws a clear line. In the direction of more clusters, the ARI again decreases, both in terms of maximum values and in terms of the ARI averaged over the different number of v10 clusters. In addition, eleven clusters should be easily distinguishable from each other in order to be able to interpret them. For v10 wind, the best ARI values are for seven clusters. This is at the minimum limit of what is reasonable. Since the ARI numbers of nine clusters are similarly good, this number is better for analysis and subsequent interpretation. The number found out in this way is used for all further variables. While it is possible, though unlikely, that the ideal number of clusters differs from variable to variable, the testing of each variable would require ten clusterings, which would demand far too much computational effort.

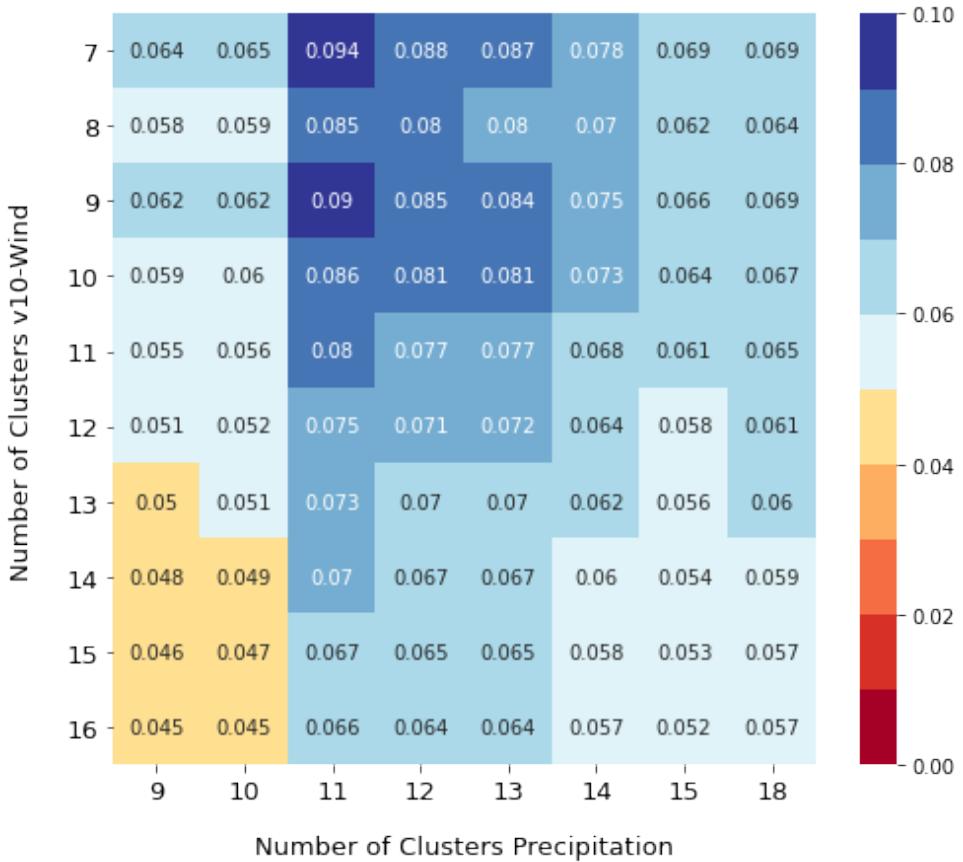


Figure 8: Adjusted Rand Index (ARI) for a different number of clusters per variable. Vertically the number of clusters for the v-wind at 10 m height is given, horizontally the number of precipitation clusters. The ARI in this case is an average of the calculations of the ARI for shifts from 0 to 14 days. If the ARI goes towards 1, the clusters match perfectly, if the ARI reaches 0, no correlation between the clusters can be found. It should be noted that with such a large amount of data of several thousand days, an ARI of 0.1 can certainly be considered good.

4.3. Selection of the variables

Using the now known number of clusters, the clustering of all variables was performed with SANDRA and then the ARI was calculated for the different variables. Not only the direct correspondence between the variable clusters and the tp clusters was investigated, but also how strongly the variable clusters influence the precipitation clusters in the following two weeks. This was calculated by comparing day 0 of the variable cluster with day x of the tp cluster using the ARI. This is shown for all variables in Figs. 9 a) and b). It should be noted that with such a large amount of data of several thousand days, an ARI of 0.1 can certainly be considered good.

First, the u- and v-wind parameters are considered at different heights. Regarding the pressure levels a clear trend can be seen, the higher in the atmosphere the investigated wind parameters are, the worse the variable clusters match the precipitation clusters, evident in Fig. 9 a). This is especially true for the v-wind component at the altitudes of 350 and 500 hPa. At this altitude, the u-wind has significantly more influence. This is probably because the strength of the trade wind can be inferred from the u-wind at this altitude, which in turn has an impact on precipitation (MANATSA et al., 2013). Nevertheless, the influence of the wind close to the ground is the largest of the different levels. This could be due in part to the profile of the study area. For example, only six of the eleven countries in the study area have an average ground pressure above 900 hPa (ERA5 data). Therefore, large parts of the u- and v-winds at 900 hPa level are only virtual there. This is also true to a lesser extent for the 750 hPa level, which cuts through mountains, especially in the Ethiopian Highlands.

As expected, the ARI drops constantly with increasing temporal distance. For most variables, the ARI is only slightly decreasing for about two days, but then drops significantly by the end of the first week before stabilizing at a lower level and dropping relatively regularly. For the different u and v wind levels in Fig. 9 a), this occurs in a similar manner in each case, and the curve shapes resemble each other. An exception is v10, which maintains its ARI skill best. From five days on, this makes v10 the u-/v-wind variable that best describes the tp clusters. Before that, the dominant u-/v-wind variable is u10.

Figure 9 b) shows the ARI values of the other variables mentioned in chapter 3.4. The surface pressure has a relatively low ARI, it is not included further in the analysis. Due to the prior experience of the u-/v-wind, the clustering of the wind direction was not performed for 350 and 500 hPa. The wind direction 10 m above ground level (wd10) drops significantly from the other variables. It does have the highest ECV of the wind direction levels at 0.50 (vs. 0.32 at wd900 and 0.30 at wd750), but the agreement with the tp clusters is substantially worse. At the 900 and 750 hPa levels, the wind direction has better performance than the pure u and v wind data from which it is composed. Here, wd900 is preferable to the wind direction at 750 hPa. Especially when looking at the first three days wd900 performs very well, best of all variables. ws10 has a similarly strong ARI on day 0, but drops off relatively quickly. Thus, wd900 is selected as the first predictor variable for precipitation. For the significance of the tp cluster days further in the future from the initial day, a second variable is needed. SST is a good choice for this purpose; its performance loss is very small compared to the other variables. On day 14, the ARI is still as high as for wd900 on day 5. Even with a shift by 30 days, the ARI of 0.062 is still acceptable. Thus, the SST and wd900 cover the ARI-ranges in the best possible way.

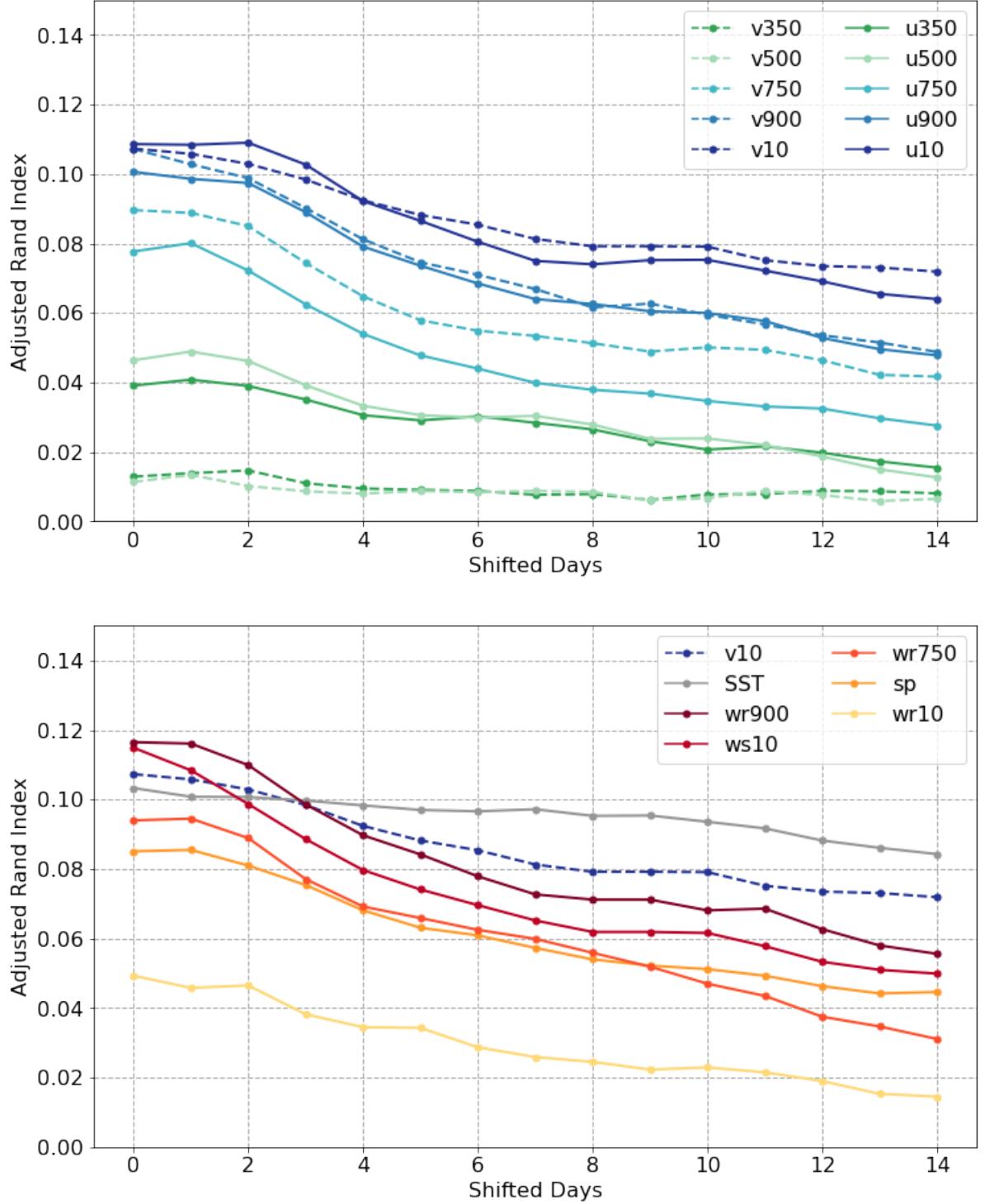


Figure 9: The Adjusted Rand Index of the studied variables with the rainfall clusters. The x-axis describes the days by which the rain is shifted, i.e., how well the clusters of the variable of day 0 match the clusters of the rain of day x. If the ARI tends towards 1, the clusters match perfectly, if the ARI reaches 0, no correlation between the clusters can be found. In this case, 1 cannot be reached because different numbers of clusters are compared. It should be noted that with such a large amount of data of several thousand days, an ARI of 0.1 can certainly be considered good. Due to the large number of variables examined, they have been split into two graphs for clarity. v10 is shown twice to improve visual comparability.

4.4. Characterization of the clusters

In this and the following sections individual clusters are discussed. They are abbreviated with the corresponding variable and its number, for example SST-1 for the 1st cluster of sea surface temperatures, wd-4 for the 4th cluster of wind direction at 900 hPa or tp-10 for the 10th cluster of precipitation. The numberings have been arbitrarily distributed by SANDRA and then retained.

4.4.1. tp-Clusters

First, the precipitation clusters are shown (Fig. 10). Here, it is directly noticeable how heterogeneous the precipitation is in the Ethiopian Highlands. While in some clusters the precipitation is uniformly higher or lower than the mean precipitation (e.g. tp-6), there is no uniform condition in the other clusters, the east of the plateau is drier while the west is wetter or the other way around (e.g. tp-1 or tp-2). Also visible are the effects of the mountains in southern Kenya, but their influence on precipitation is rather smaller, only at tp-11 their effect is stronger. By the representation as a map, it can be recognized at first sight, which areas are particularly strongly affected by which cluster. For example, clusters tp-3, -4 and -10 are clearly conspicuous, as they show significantly increased precipitation amounts compared to the normal condition off the coast of Tanzania and Kenya. Also striking is tp-6, as it is indicated as very dry across the area. This can be confirmed by the precipitation values broken down by country; in seven of eleven countries, tp-6 is the driest cluster, over the ocean as well, and in the other countries it is the second or third driest (see Tab. A2). tp-4 is the wettest of the clusters, this is mainly due to the heavy rainfall over the Indian Ocean, especially south of the equator. The other clusters are discussed individually in the course of this chapter. Figure 10 gives a good overview of the absolute rainfall of the cluster and its distribution, but it is susceptible to seasonality. For example, tp-11 occurs primarily in late March/early April. At this time, the rainy season is usually even weaker. In absolute terms, the cluster is expressed as tending to be somewhat wetter, but not significantly wetter than tp-4 or tp-5, for example. However, these two clusters occur at the time of most rainfall in the MAM season and are therefore significantly wetter than the other clusters in advance. When seasonality is factored out by calculating the deviation from the mean of the individual days of the season over 1980–2018 instead of the mean of the season, tp-11 emerges as a particularly wetter cluster. This is shown in Fig. 11 using the DWI. It is noticeable when comparing Figs. 10 and 11 that in the latter, the influence of the individual mountains is almost not seen. Among the individual clusters, tp-6 stands out, where the area almost consistently receives half as much precipitation as usual at this time.

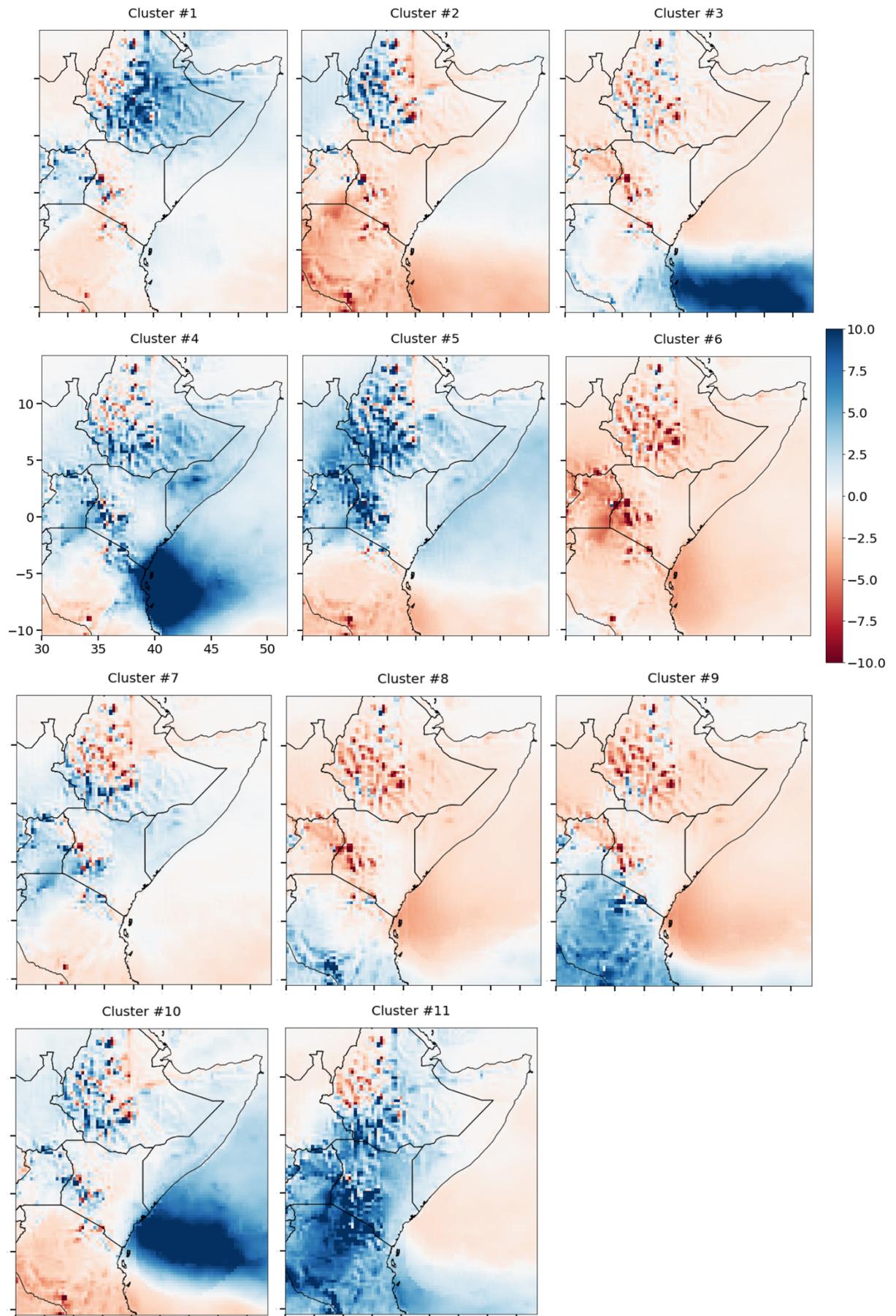


Figure 10: Deviation of tp cluster centroids from mean precipitation for March to May. The unit is mm/day.

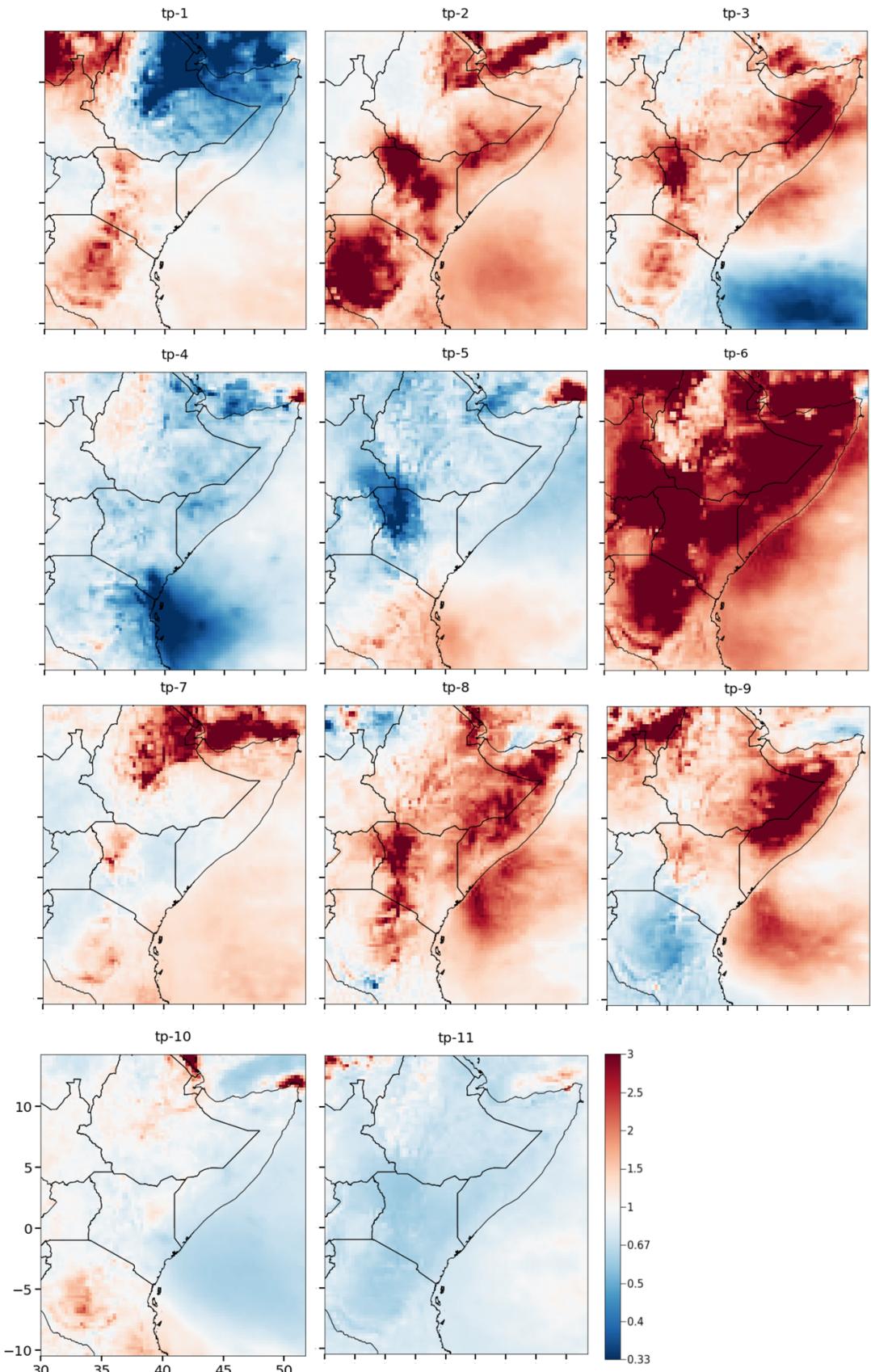


Figure 11: Daily Wetness Index (DWI) of cluster centroids of precipitation (tp). Values of 1 indicate that the cluster has a normal amount of precipitation in relation to the usual precipitation at the time of occurrence of the cluster. Values of 2 indicate regions that are twice as dry, values of 0.5 indicate regions that are twice as wet.

In the following, the clusters are discussed individually. In particular, the influence of the characteristics of the clusters on the precipitation distribution in the countries/regions is elaborated. The exact temporal classification of the clusters can be found in Tab. A1 in the appendix, and the local effects of precipitation are already shown in Figs. 10 and 11. The values in parentheses in the following text are correlation values, significant at $\alpha = 0.05$ from 0.32/-0.32, unless otherwise defined.

tp-1 is the cluster during which most of the Northeast's heavy precipitation events occur. This can be seen in Fig. 11. When tp-1 occurs early in the year, it significantly increases the number of extreme precipitation events in the MAM season, a correlation of -0.43 between the date of the first occurrence and the Number of extreme precipitation events is given. Over the course of the study period from 1980 to 2018, the occurrence of this cluster has significantly decreased, occurring 58.5 % less frequently toward the end than it did at the beginning (Fig. A2). Since it is a rather moist cluster overall (DWI of 0.89), this is consistent with the trend of decreasing MAM precipitation. To this end, tp-1 occurs later and later in the season, with the significant trend amounting to a delay of 16.8 days. This is in line with the general trend of a later onset of the rainy season (Fig. 6). A rare occurrence of tp-1 in a year leads to a later start of the rainy season in Ethiopia and Somalia (correlation of -0.51 and -0.49, respectively). In addition, this also indicates an earlier end of the rainy season in Djibouti (0.42), Ethiopia (0.34), Kenya (0.40), and Somalia (0.37); resulting in a reduction in the length of the rainy season in Djibouti (0.34), Ethiopia (0.56), Kenya (0.42), Somalia (0.64), and South Sudan (0.33). The frequency of tp-1 also correlates with rainfall, in Djibouti (0.67), Eritrea (0.56) and Ethiopia (0.63), in addition to the number of dry days per year of Djibouti (-0.65), Eritrea (-0.65) and Somalia (-0.51). Overall, tp-1 consequently characterizes mainly a strong precipitation surplus in the north-east of the study area and follows the drought trends.

tp-2 is a very hot cluster in the north, where it is 2 K warmer than during the other clusters (compare with Tab. A5). It occurs dominantly in May and negatively correlates in that month with precipitation in Kenya, Burundi, Rwanda, Tanzania, and Uganda (correlation with frequency stronger than -0.56 in each case; Fig. 11). Moreover, during tp-2, a clustering of dry days of the wettest (Burundi 0.49 and Rwanda 0.52) and driest regions (Djibouti 0.38 and Eritrea 0.33) happens and the frequency of extreme precipitation decreases (-0.52). If cluster 2 occurs later, the MAM precipitation in South Sudan (-0.32) and Sudan (-0.41) decreases. The frequency of tp-2 correlates with the end of the rainy season in Burundi (-0.56), Kenya (-0.66), and Rwanda (-0.62) and the length of Burundi (-0.40) and Rwanda (-0.43). The cluster is pushed back calendrically over the duration of the study period, tp-2 occurs 19.8 days later at the end than at the beginning. Overall, tp-2 is a warm cluster, particularly dry in the south, indicating the end of the rainy season around the equator.

tp-3 is a rather humid cluster, currently more frequent than at the beginning of the study period (increase of 39.8 % during 1980–2018, see Fig. A2). The date of last occurrence correlates with May precipitation of Tanzania (0.45), Rwanda (0.45), and Burundi (0.47), consequently brings rains to the south. If tp-3 occurs frequently, the number of dry days of Somalia (0.37) increases and the rainfall of Eritrea (-0.41) and Somalia (-0.36) decreases, implying a dry north. Figure 10 and 11 show a strong excess of precipitation in the Indian Ocean, which is mainly local. Overall, tp-3 is a cluster that used to be substantially wet but became significantly drier during the study period. tp-4 also has a precipitation surplus at about the same location. Overall, however, cluster 4 is substantially wetter than tp-3; it is the wettest cluster in absolute precipitation and the second wettest in DWI, with a value of 0.62. Thus, heavy precipitation events occur throughout the study area during tp-4, with the exception of Burundi and South Sudan, and the frequency of the cluster correlates with the frequency of these extreme precipitation events (0.59). In addition, the frequency negatively correlates with the dry days of Burundi (-0.41), Djibouti (-0.35), Eritrea (-0.41), and Rwanda (-0.37), and positively with the end of the rainy season in Burundi (0.45) and Tanzania (0.36). If tp-4 occurs early in the season, a heavy May rainfall occurs in Rwanda (0.46). tp-4 now occurs 9.9 days earlier than it did in the 1980s; because it is a cluster of the rainy season, it goes against the general trend of a later-onset rainy season. Overall, the occurrence of tp-4 is indicative of higher rainfall across the region during the MAM season (correlation of 0.61), especially in Kenya (0.63).

tp-5 is the coldest of all clusters compared to normal, especially in South Sudan and Ethiopia (see Tab. A5). The typical time of occurrence is when the rainy season of Ethiopia, Kenya and South Sudan is most pronounced in late April. Accordingly, the cluster frequency also correlates with the May precipitation of Kenya (0.43) and with the end of the rainy season in Kenya (0.43), as well as with the frequency of extreme precipitation (0.33). If tp-5 occurs early in the season, April precipitation increases from Kenya (-0.45), Ethiopia (-0.37), and Somalia (-0.38). If cluster 5 stops again early, May precipitation of all countries except Tanzania decreases significantly, especially markedly in Ethiopia (0.51). The more often tp-5 occurs, the longer the rainy season lasts in Ethiopia (0.43); but the relationship is also significant in Uganda (0.35) and Kenya (0.33). Overall, tp-5 is a wet cluster almost everywhere; it has a DWI below 0.84 in all countries except Tanzania.

tp-6 is the most frequently occurring cluster, which alone accounts for 20.3 % of all cluster assignments (see Tab. A1). Although it occurs most frequently in March, it can be assigned to a day at any time. Cluster 6 is special because it is by far the driest cluster, both in absolute terms and compared to normal (DWI of 2.01). The most severely affected areas cannot be discerned in Fig. 11 because it is consistently more severe than the maximum value on the scale. The individual country values indicate that the precipitation deficit is particularly severe in Kenya (DWI of 3.36), Somalia (DWI: 2.80), and South

Sudan (DWI: 3.01). Consequently, only one-third of the for this time usual precipitation falls in those countries during tp-6. The center of impact is therefore at about 7° N. Half of all days are more than one standard deviation away from the mean (see Tab. 3), which classifies them as particularly dry days. In addition, tp-6 is the cluster with the fewest extreme precipitation events (-0.49). The cluster frequency is significantly positively correlated with the dry day frequency of all countries except for Tanzania and Sudan (e.g., South Sudan 0.69; Ethiopia 0.52; Somalia 0.62) and negatively correlated with rainfall from most countries across all three months. The impact is strongest in March, particularly in Uganda (-0.69), Rwanda (-0.60), Kenya (-0.60), and Tanzania (-0.58). The last occurrence of tp-6 correlates with May precipitation for all countries except Tanzania and Rwanda, and is particularly strong for Somalia (-0.65), Ethiopia (-0.60), and Djibouti (-0.66). The length of the rainy season significantly shortens a frequent occurrence of cluster 6 in all countries except for the three northern ones, especially significantly in Kenya (-0.55). In the northern countries, this tends to be manifested by a delayed start of the rainy season, and in the southern ones by an early end. Cluster 6 occurs more often at the end of the study period than earlier (+38.9 %, see Fig. A2), following the trend towards a drier MAM season. Overall, tp-6 is a particularly dry cluster that should be looked at more closely during further analysis, as it indicates particularly strong impacts on the Greater Horn of Africa.

tp-7 is striking, as it is drier than normal in the east and wetter in the west. The driest impacts indicated by the cluster are in the already dry countries of Djibouti (DWI: 2.48) and Eritrea (DWI: 2.51). For the other countries, tp-7 tends to indicate an extension of the rainy season. The frequency of occurrence correlates with the rainy season start date of Somalia (-0.49) and with the end date of Kenya (0.34), Rwanda (0.36), and South Sudan (-0.34), and the length of the rainy season in Burundi (0.34), Kenya (0.42), Rwanda (0.54), and Uganda (0.36). With the cluster frequency significantly correlated with March rainfall for all regions except Sudan and South Sudan, particularly Ethiopia (0.55) and Somalia (0.54), rainfall has an influence on this. In Somalia, this is also reflected by an absence of dry days in many tp-7 events (-0.53). If the last occurrence of cluster 7 is late, May precipitation is high, especially from Burundi (0.68) and Tanzania (0.57), and there are more heavy rain events (0.39). tp-7 occurs much earlier at the end of the period than in the past (10.8 days). Overall, tp-7 tends to indicate more precipitation over land areas except for the northeast.

tp-8 has virtually no strong precipitation events. If the last occurrence of the cluster is early in the season, it means an earlier start of the rainy season; March precipitation for the entire region correlates with -0.53. Thus, a clustered occurrence of tp-8 changes the length of the rainy season in Burundi (-0.32), Rwanda (-0.50), and Uganda (-0.45), as well as Djibouti (0.37). The correlation is somewhat stronger in the south than in the north. The frequency of tp-8 is also more likely to correlate with the number of dry days for southern countries (Rwanda 0.36) and with April precipitation for Ethiopia (-

0.55) and Rwanda (-0.55). In addition, heavy precipitation days (> 5 mm a day per country) become less frequent (-0.33). Overall, tp-8 is considered dry (DWI: 1.23); it is also the second driest cluster in absolute terms. tp-9 has a dipole with respect to DWI (Fig. 11). In the northeast and in the southeast over the Indian Ocean, there is less precipitation than usual, but in the southwest there is much more. As a result, on the one hand, frequent occurrence of the cluster means more rainfall in Tanzania (0.51) and fewer days of drought in Uganda (-0.40), but at the same time it delays the start of the rainy season in Ethiopia and Somalia (0.39 and 0.38) and shortens the duration of the rainy season of the two countries (-0.35 and -0.34). Overall, this dipole structure shapes the effects of tp-9.

tp-10 has the fewest extremely hot days above the 95th percentile. Otherwise, it is a rather humid cluster, especially over the ocean (DWI there: 0.52; Fig. 11). If the cluster occurs frequently, the end of the rainy season over the Indian Ocean is delayed (correlation of 0.49). In addition, its frequency also affects May precipitation for the entire region (0.61) and extreme precipitation (0.40). The date of first occurrence of tp-10 correlates with April precipitation of the whole study area with -0.52 and the date of last occurrence correlates with May precipitation of Kenya (0.41) and Somalia (0.43). Overall tp-10 is a wet cluster (except for the northwest) representing the end of the rainy season.

tp-11 is particularly wet, having the lowest DWI of all clusters, averaging 0.59 (Fig. 11). The excess precipitation is particularly pronounced in Kenya (DWI of 0.39). tp-11 occurs in March and April, and the first cluster occurrence correlates strongly with the rainfall of countries located near the equator (Burundi -0.39; Ethiopia -0.41; Kenya -0.57; Tanzania -0.55; Uganda -0.63). The property of frequency of occurrence of tp-11 is even more strongly associated with rainfall. The frequency correlates very strongly with Kenya's MAM precipitation (0.75), and with Rwanda (0.61), Burundi (0.57), Ethiopia (0.58), Somalia (0.56), Tanzania (0.68), and Uganda (0.62) also particularly strongly. In addition, the cluster indicates the most heavy precipitation events south of 4° N (0.63). The number of dry days is particularly low in Somalia (-0.51) and South Sudan (-0.47) when tp-11 is frequent. tp-11 has a significant impact on the rainy season of several countries. The frequency of tp-11 negatively correlates with the start of the rainy season in Burundi (-0.58), Ethiopia (-0.52), Kenya (-0.49), Rwanda (-0.58), Somalia (-0.43), and Uganda (-0.53), and positively with the end of the rainy season in Tanzania (0.38). Thus, it characterizes the length of the rainy season of Burundi (0.54), Ethiopia (0.48), Kenya (0.46), Rwanda (0.63), and Uganda (0.56). Overall, tp-11 is a crucial cluster for the study region as it indicates an early or late start of the rainy season.

Table 3: Percentage of cluster days that are more than one standard deviation (std) away from the mean, which corresponds to a ranking below/above the 16th/84th percentile. In a) the days are listed that are below mean minus std, in b) those that are above mean plus std. Particularly high (low) percentage values in a (b) thus indicate dry clusters and are marked red for better recognition, particularly low (high) values indicate wet clusters and are marked blue. The tables are divided into two parts, the left columns calculate with the standard deviation and the mean value of the whole MAM season, the right columns calculate for each day a mean value and std and calculate the percent values thereby, equivalent to the DWI. This removes the influence of seasonality. The SST- and wd900-classification has been done with nine clusters, the tp-classification with eleven.

a)

Cluster	Mean			Daily		
	SST	wd900	tp	SST	wd900	tp
1	12.1 %	2.9 %	0.0 %	13.0 %	7.6 %	2.1 %
2	3.8 %	28.2 %	13.5 %	12.9 %	19.8 %	19.9 %
3	1.9 %	19.4 %	0.0 %	14.5 %	16.9 %	0.4 %
4	14.4 %	3.1 %	0.0 %	13.8 %	4.5 %	0.0 %
5	43.7 %	38.7 %	0.0 %	32.0 %	21.6 %	0.5 %
6	3.9 %	19.6 %	56.2 %	10.3 %	17.8 %	49.2 %
7	24.6 %	10.3 %	0.0 %	12.5 %	19.9 %	4.8 %
8	37.1 %	10.4 %	16.7 %	15.0 %	2.2 %	12.7 %
9	13.8 %	8.7 %	0.8 %	18.4 %	15.3 %	0.8 %
10			0.0 %			0.0 %
11			0.0 %			0.0 %

b)

Cluster	Mean			Daily		
	SST	wd900	tp	SST	wd900	tp
1	11.4 %	42.4 %	17.5 %	15.0 %	27.6 %	15.4 %
2	27.8 %	3.5 %	0.2 %	16.9 %	7.9 %	2.0 %
3	34.9 %	8.3 %	14.8 %	22.2 %	12.1 %	21.1 %
4	11.9 %	29.3 %	81.9 %	16.1 %	28.8 %	63.3 %
5	0.4 %	0.7 %	45.5 %	2.6 %	7.4 %	31.1 %
6	24.3 %	9.5 %	0.0 %	20.1 %	11.4 %	0.0 %
7	6.8 %	15.8 %	5.6 %	18.9 %	11.8 %	3.0 %
8	1.1 %	14.3 %	0.4 %	9.6 %	30.2 %	4.0 %
9	15.1 %	18.1 %	1.4 %	12.2 %	13.5 %	12.1 %
10			47.5 %			39.7 %
11			64.3 %			66.8 %

4.4.2. SST-Clusters

While the correlations for the tp clusters with precipitation are understandable, since they are the same variable, the clusters of the first predictor variable are now described. The SST clusters follow a fixed seasonal passage that changes little. The deviations from mean water temperatures shown in Fig. 12 move from south to north following the sun through the study area. Because of the high heat storage capacity of water, the SSTs are lagged with respect to the zenith of the sun by about one-half to one month (DONOHOE et al., 2013). In most cases, the starting cluster is SST-8, and the ending cluster is SST-5. These can also be identified as the most distinct deviation from the average of SSTs during the MAM season.

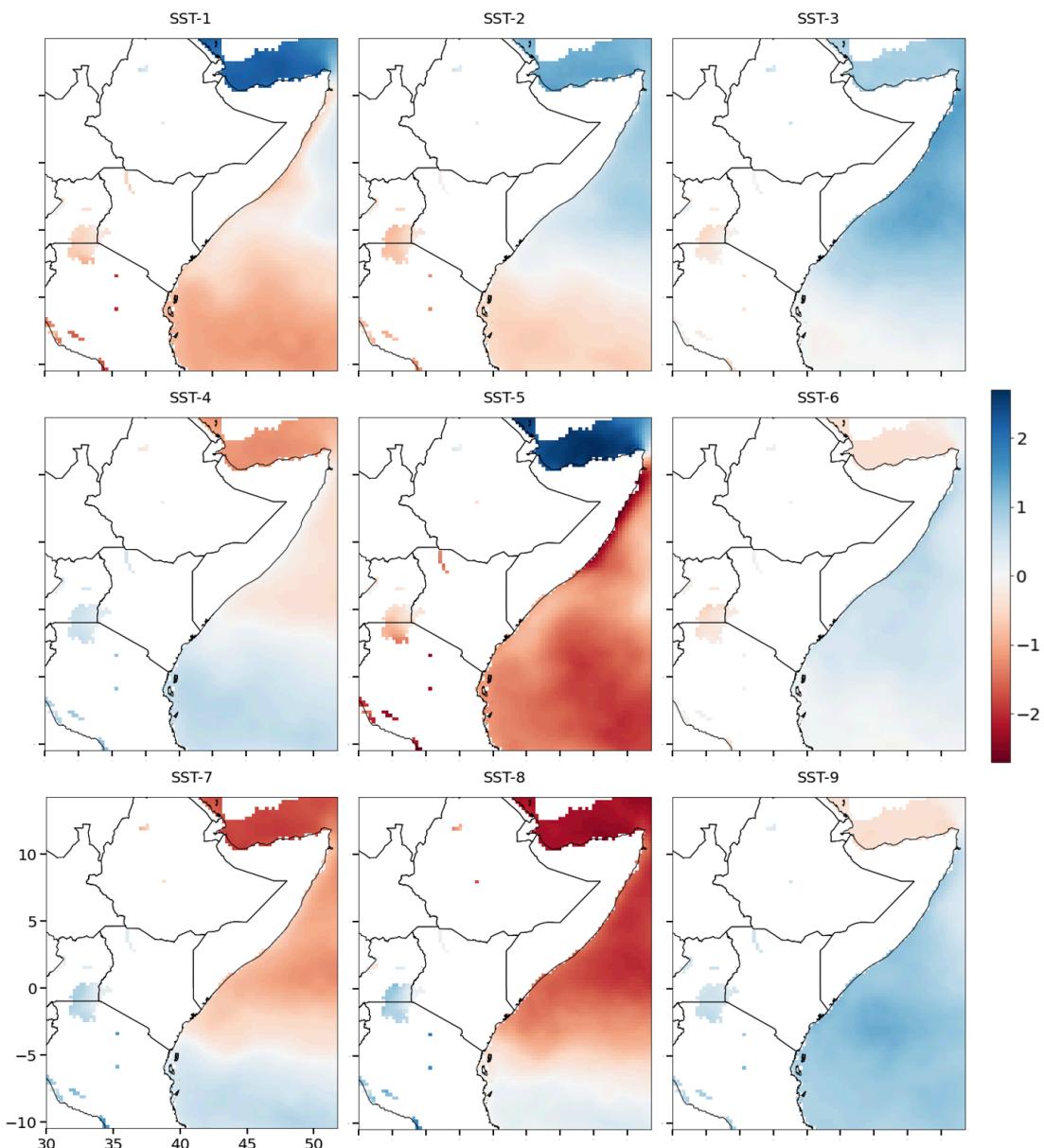


Figure 12: Deviation of SST of the individual SST cluster centroids from the mean of the MAM season in K. The units of the color bar are kelvin. The numbering of the clusters is done randomly by the SANDRA algorithm.

Compared to the tp clusters, which last only two days on average before being alternated by another cluster, the SST clusters are more stable and have an average duration of about nine days. Compared to the precipitation clusters in Fig. 10, the DWI fluctuations are much smaller for the SST. In order to be able to explain the precipitation deficits and surpluses somewhat, the wind system of the respective cluster is additionally shown in Fig. 13. Especially SST-5 and SST-8 stand out due to their high DWI in large areas.

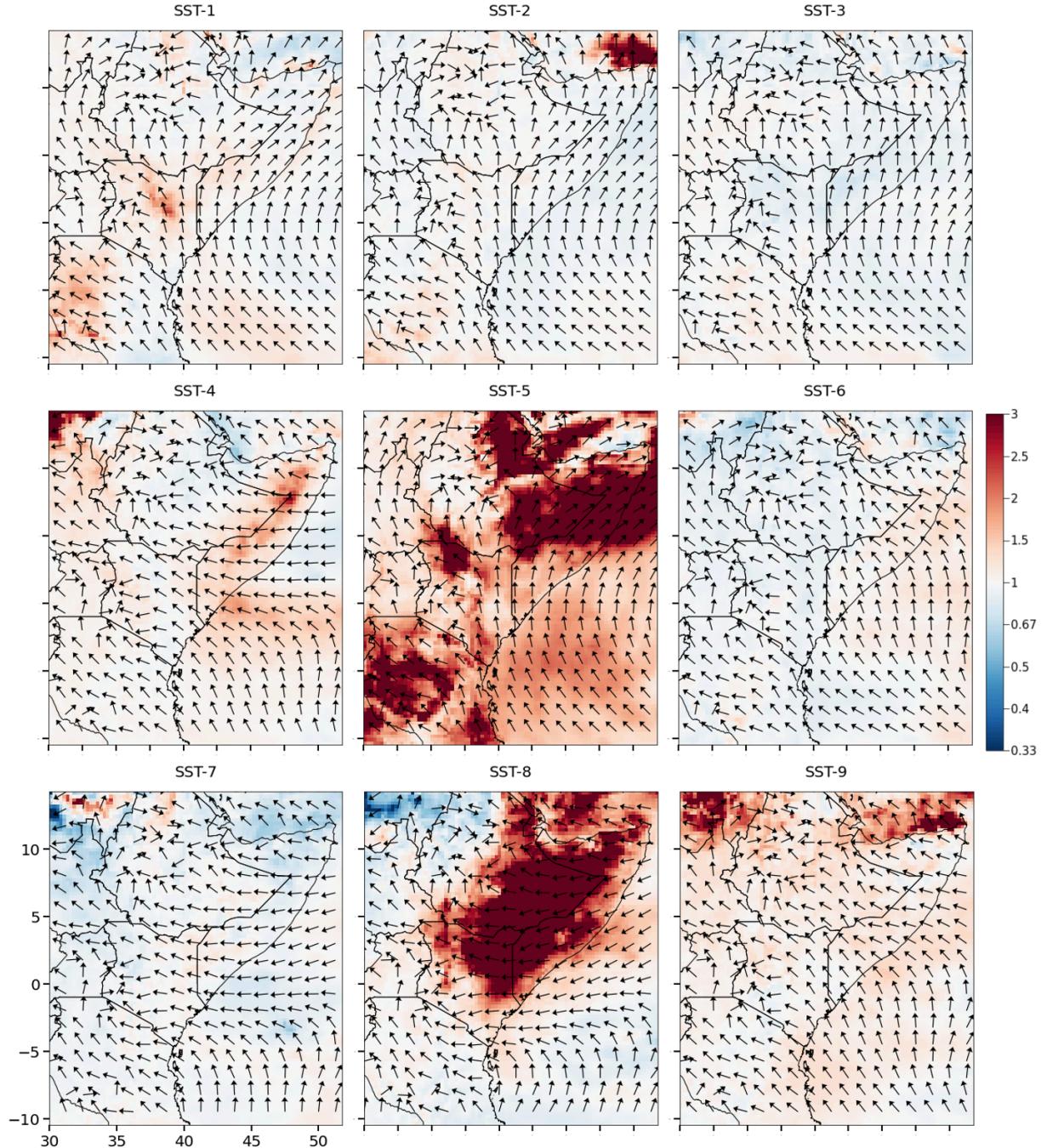


Figure 13: Daily Wetness Index (DWI) and average wind direction of the SST clusters. The wind strength is excluded, all arrows have the same length. A DWI > 1 indicates drier than usual conditions, < 1 wetter than usual conditions.

SST-1 brings dry and hot conditions to the south of the study region. When cluster 1 occurs early, May precipitation is lower than usual in all areas south of 11° N (correlation of 0.34 in Ethiopia and 0.51 in Kenya). The date of last occurrence correlates with May precipitation in southern countries, e.g., 0.45 for Tanzania and with the onset of the rainy season in Rwanda (0.45). The southern coastal areas are wetter ($DWI < 1$) because the wind there comes from the ocean (see Fig. 13). Over the Gulf of Aden, the wind comes from Somalia, then cools due to the low ocean temperatures and rains out, resulting in a slightly wetter DWI there. Overall, SST-1 is a transitional cluster that is the central cluster of the rainy season in the north, the last cluster of the rainy season in the center (Ethiopia, Somalia), and the first of the dry season in the south. SST-2 generally occurs before SST-1. In SST-2, the absolute rainfall in Burundi, Tanzania, and Rwanda is still twice that in SST-1. SST-2 is the cluster that precedes the rainy season for Sudan. Similar to SST-1, an early onset in May produces less rainfall, but this influence extends only from the southern boundary of the area to 3° N, not to Ethiopia. The date of last occurrence correlates significantly with MAM precipitation over the Indian Ocean at 0.52 and with May precipitation in southern countries (e.g., Kenya 0.57 and Uganda 0.55). This correlation occurs because a late ending means that subsequent drier clusters do not occur in May. Overall, SST-2 is the last rainy season cluster of the southern countries.

SST-3 is the cluster most often responsible for heavy rainfall. When SST-3 occurs, the rainy season is most pronounced in Somalia, Kenya, and Ethiopia. Compared to SST-9, which occurs at about the same time, the ocean is much warmer in the south. The south is the origin of the ocean winds leading to Kenya and Somalia, so the DWI there is significantly wetter than in SST-9 (see Fig. 13). Analogous to SST-2 and SST-1, when SST-3 occurs early, May tends to be dry (correlation of 0.43 across the region). This is probably due to a resulting earlier occurrence of SST-5, which is particularly dry. In addition, an early first occurrence leads to high April precipitation in Somalia (-0.61), Kenya (-0.45), and Ethiopia (-0.39), because the rainy season reaches its maximum in April instead of May. The last date of the cluster correlates positively with May precipitation in the southern countries, quite strongly over the whole study region at 0.57. Cluster 3 occurs almost twice as often at the end of the 1980–2018 period as at the beginning (+88.6 %; see Fig. A2). This is also manifested in an earlier occurrence by 7.6 days. This goes against the trend of the MAM season becoming drier on average. Overall, SST-3 is the wettest and warmest of all clusters, with the highest precipitation amounts and the lowest DWI of the SST clusters (0.93).

SST-4 now occurs 8.2 days earlier than in the 1980s. The date of occurrence correlates with the main precipitation in Somalia and Sudan with 0.44 and 0.43, respectively; the date of last occurrence with the April precipitation of Ethiopia (-0.38), Kenya (-0.37), and Somalia (-0.57). If SST-4 occurs frequently in a season, dry days in Tanzania (-0.44) are reduced on the one hand, but April precipitation (-0.47) is

also reduced. Since the p-values are 0.005 and 0.003, respectively, an error is very unlikely. Moreover, it is not the case that March and April precipitation compensate for this; over the entire MAM season, the correlation is still clearly significant (-0.39). Since the frequency of cluster 4 is negatively correlated with the frequency of extreme rainfall (-0.48), one explanation could be that it rains more often than usual in Tanzania, but less per day. This is consistent with the slightly drier area over Tanzania in Fig. 13, which has virtually no outliers up or down. Overall, SST-4 is a relatively average cluster that is drier than normal in Somalia.

SST-5 is the cluster most different from the other SST-Clusters. It is significantly drier than all other clusters (DWI of 1.45), and all countries have a significant rainfall deficit. Thus, SST-5 negatively correlates with May precipitation, especially those in Somalia (-0.46) and Tanzania (-0.44); and positively with the number of dry days in Burundi (0.47) and Rwanda (0.40). In addition, it brings the end of the rainy season in Eritrea (-0.45) and Somalia (-0.55) and shortens the duration of the rainy season in Djibouti (-0.45) and Eritrea (-0.41). A monsoon effect is evident off the coast of Kenya, with winds changing direction from southeast to southwest after crossing the equator. Over the Gulf of Aden, an isolated area can be seen that has a moist DWI in remarkable contrast to its surroundings. This is because the origin of the air masses that meet the relatively cold sea there (see Fig. 12) is in the Ethiopian Highlands. There, at this time, the rainy season has just ended, so the air masses are relatively moist. The air does not rain over the land as it descends along the profile from the Highlands to the coast. When the moist, warm air masses meet the relatively cold air masses from the sea, they slide up onto them, causing adiabatic cooling. The moisture condenses and a light drizzle falls. Since this region is normally very dry, the small amount of drizzle is enough to bring the DWI below 1. On a smaller scale, the moisture penetrates a few kilometers into the cold coastal air masses, creating coastal fog. The process is similar to that on the coast of Namibia, where hot air masses from the Kalahari meet the cold Benguela Current offshore, producing fog that provides moisture to the surrounding area (SEELY and HENSCHEL, 1998). This effect is likely to be even stronger at the Gulf of Aden, where the air masses are significantly more humid, although the SSTs are not as cold. Overall, SST-5, the last cluster in the series, brings a precipitation deficit that mainly affects countries south of the equator. The earlier SST-5 sets in, the drier the MAM season becomes.

SST-6, which virtually only occurs in April, brings extreme precipitation to Rwanda, Burundi, and Uganda. In these three countries, cluster 6 is also the cluster with the highest rainfall of the season. The first occurrence date correlates negatively with April precipitation for Ethiopia (-0.36), Kenya (-0.50), Uganda (-0.42), and Somalia (-0.54); i.e., the region slightly north of the equator. Consequently, if SST-6 occurs early in April, the month becomes wetter. The date of last occurrence correlates negatively with April precipitation for Kenya (-0.34) and Somalia (-0.47). Cluster 6 occurs less frequently at

the end of the 1980–2018 period than earlier (-49.9%, see Fig. A2). The SST anomalies are spatially very similar to the DWI. That is, where the ocean is warmer than normal, more precipitation falls and the other way around. This is expected, but occurs in a similar form only at SST-4 and -7. It fits that on the coast of Somalia, where SST is relatively cold, the DWI indicates drought (Fig. 13). Overall, SST-6 is a moist cluster responsible for the start of the rainy season around the equator.

SST-7, while a rather dry cluster in absolute terms, tends to be moist (0.96) according to the DWI, especially in the north (Fig. 13). This is because it is to be dated in the transition between dry and wet seasons in early March. The start (0.39) and length (0.37) of the rainy season in Sudan is co-dependent on SST-7. This is also reflected in the rainfall figures; the more frequently SST-7 occurs, the more rain falls in Sudan in March (0.52) (cf. Tab. A3). The first occurrence date correlates negatively with March precipitation in Kenya (-0.35) and the frequency of SST-7 correlates negatively with the number of dry days in Tanzania (-0.40). The cluster occurs significantly less frequently (-44.9 %, see Fig. A2) over the study period. The Horn of Somalia and the Gulf of Aden are substantially wet according to DWI, this is probably due to the particularly warm temperatures of the Gulf. In the south, the location of the ITCZ can be clearly seen from the wind patterns, it is indicated by a band of easterlies with converging northeast and southeast trade winds. It is slightly more northerly than usual at the time, as it overlaps with Cluster 8, which has a more southerly ITCZ. This creates a shift in rainfall that creates the two-striped pattern of DWI over southern Indian (see Fig. 13). Overall, SST-7 is a slightly moist cluster.

SST-8 is the second driest cluster of the season after SST-5. While SST-5 is the last cluster of the season, SST-8 is always at the beginning of the MAM season. It is the driest cluster especially for the northern countries as it is still in their dry season (Fig. 13). Especially dry are Kenya, Somalia and Ethiopia. In Tanzania and Uganda it is the first cluster of their rainy season. In Sudan it brings a very wet DWI, but since the average rainfall in Sudan at this time is 0.07 mm/d, a single rainfall event may have caused this and is thus negligible. The significantly dry DWI over Somalia and eastern Ethiopia likely comes from the fact that SST-7 may also occur at the peak occurrence time of SST-8. SST-7 is four times wetter over this region than SST-8, so the mean on the corresponding date is also modified by only a few SST-7 events. In years with dominant SST-8, March precipitation is significantly lower in the corresponding countries. Thus, the frequency correlates with the number of dry days of Somalia (0.36) and Ethiopia (0.45) and with March precipitation of Ethiopia (-0.33). If SST-8 extends well into March, the March precipitation of the northeastern countries, especially Ethiopia (-0.56), is significantly lower. Overall, SST-8 is a very dry cluster at the beginning of the MAM season that rarely brings extreme precipitation.

SST-9 is the cluster with the warmest anomaly relative to normal, according to the DTI. If the cluster occurs early in the season, it rains more in April (-0.51). The end date of SST-9 correlates with May precipitation for the countries south of Ethiopia, particularly Burundi (0.58) and Uganda (0.48). The

drier conditions over the sea and coasts are likely due to the substantially lower sea surface temperatures in SST-9 compared to the co-occurring SST-3 and SST-6. Cluster 9 ends up occurring much more frequently than it did in the 1980s, with an increase of 73.4 % (see Fig. A2). Overall, SST-9 is a dry, warm cluster that influences precipitation in April.

4.4.3. wd900-Clusters

The wind direction (wd) clusters are more dynamic than the SST clusters, resulting in higher DWI values in comparison (cf. Figs. 13 and 14). In the wd clusters, the influence on the regions is clearly more small-scale, there are less uniform areas of heavy precipitation as in the SST clusters. The mapped wind directions deviate more clearly from the average wind conditions of the respective occurrence time, so that an interpretation of different regions of high/low DWI values is possible. Because the main inflow direction of the region is mostly east, it is necessary to know whether wind from the sea has a cooling or warming influence. This is not trivial because there are no cold ocean currents offshore. The difference between the temperature above the water surface and the land surface is crucial. In early March, the sea is still 2 K warmer than the 2-m temperatures (Fig. A1). By the end of March, the sea and land are warming at about the same rate, so the 2 K spread remains. In April, the air temperature over the sea drops while the 2-m temperature over land continues to rise until by the end of April the 2-m temperature above land has caught up with the temperature above sea. During May, the air masses above land and sea have nearly the same temperature (Fig. A1). Consequently, the sea wind has a rather warming influence until the end of April, in May the differences are too small that it should have an effect.

wd-1 is a distinctly wet cluster, with a DWI below 0.86 in all countries. The wd-1 frequency and April precipitation are closely related, especially in Kenya (0.60), Ethiopia (0.51), and Somalia (0.54). This also increases the number of extreme precipitation events (0.51). The date of first occurrence correlates with May precipitation in Somalia (0.38), Uganda (0.38), and Tanzania (0.35); the last occurrence with MAM precipitation of countries south of Ethiopia (e.g., Kenya 0.51; Rwanda 0.57). wd-1 is the cluster that decreases temperatures the most compared to normal (Tab. A7). This could be due to the high precipitation, which prevent the region from warming more due to increased cloud cover and latent energy transport. Cluster 1 differs from the average wind conditions at this time in that the area's otherwise typical southerly flow, especially over Somalia, does not exist and the wind over the sea shifts to the east. Overall, wd-1 is a very wet, cold cluster compared to normal, occurring during the peak of the rainy season.

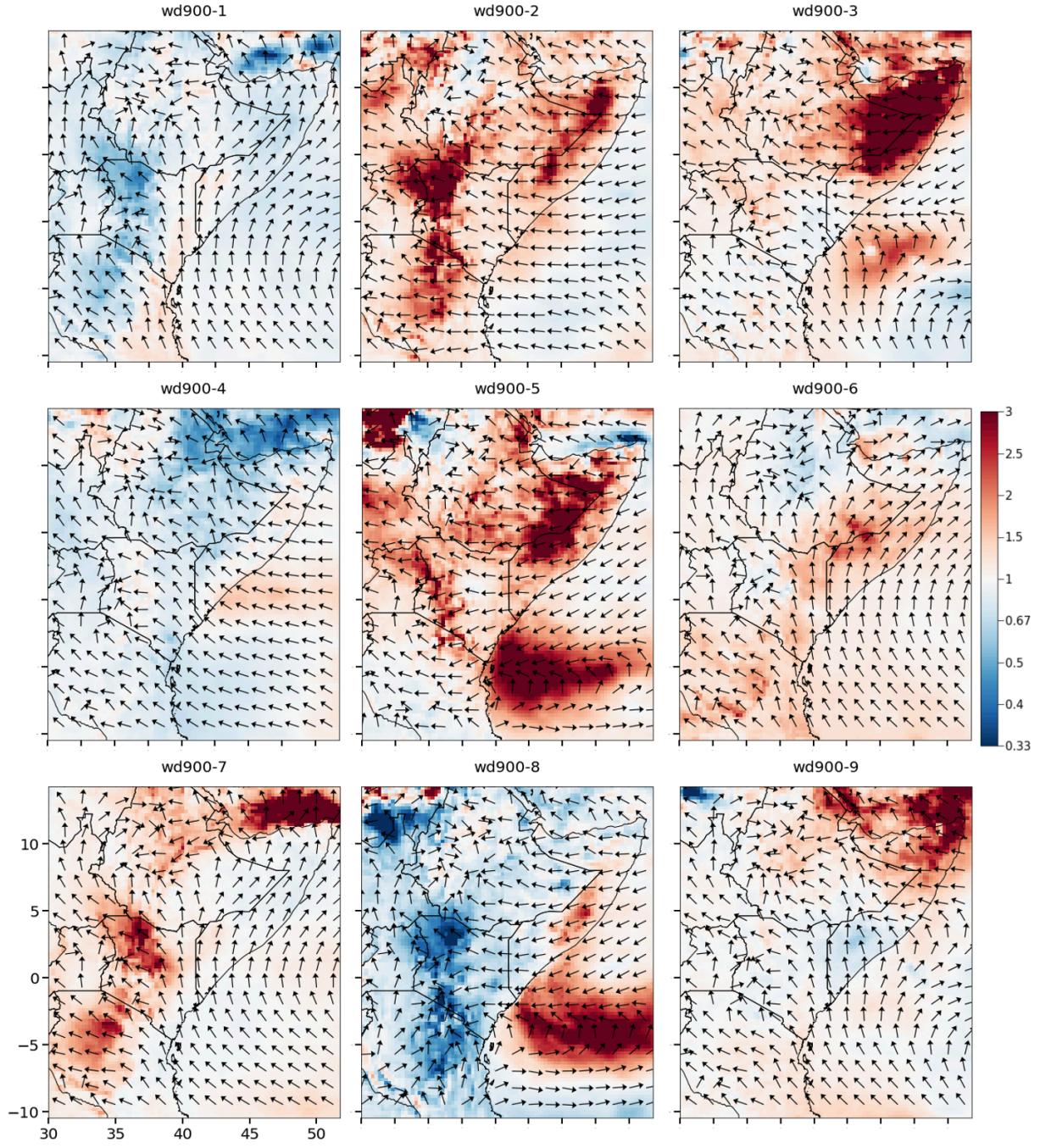


Figure 14: Illustration of the clusters of wind direction at 900 hPa. The arrows are the wind direction, underlined with the Daily Wetness Index of the respective cluster. A Wetness Index value of 1 means normal conditions, 0.5 means twice as much precipitation and 2 means half as much precipitation as usual.

wd-2 has many extremely hot days with daily mean temperatures 2 m above the ground beyond the 95th percentile in the region of the equator. Clusters 3, 5, and 8, which occur at about the same time, result in fewer such days. It is also the second driest cluster. This is especially evident over land (see Fig. 14). Thus, the frequency of wd-2 correlates negatively with Tanzania's March precipitation (-0.40) and positively with Kenya's dry days (0.39). If wd-2 occurs earlier in March than normal, less precipitation falls over the Indian Ocean. If the cluster ends later than normal, March precipitation is low in

Somalia (-0.35). At the equator, instead of running parallel to the coast as normal, the wind comes from the sea, which is warmer than the land in March. This can explain the high temperatures. However, why it is so dry in the north cannot be explained based on the wind direction, they hardly differ from the prevailing wind patterns at this time. Overall, wd-2 is rather dry, with a moist influence on the Indies and the Ethiopian Highlands. wd-3 is very dry, especially in the Horn of Africa, which is well reflected in the April precipitation of Ethiopia (-0.45), Kenya (-0.42), and Somalia (-0.45). To this end, the frequency of extreme precipitation decreases throughout the region during wd-3 (-0.49). Instead of coming from the sea as usual during this month at the equator, the wind turns eastward while still on the sea, so that not as much moisture as usual reaches Kenya, Ethiopia, and Somalia. Overall, wd-3 is a dry cluster over land with a dipole south of the equator in the Indian Ocean.

wd-4 is the cluster with the most extreme precipitation events, which can occur in any country. It is the cluster with the lowest DTI, especially Kenya and South Sudan are warmer than usual (Fig. 14). If wd-4 occurs frequently, April precipitation increases in Eritrea (0.37) and the number of dry days of Djibouti (-0.37) and Eritrea (-0.34) decreases. The length of the rainy season of Eritrea is prolonged by a frequent occurrence of wd-4 (0.45). An early first occurrence leads to increased March precipitation in the northeast of the area (Djibouti -0.41; Eritrea -0.37, Ethiopia -0.43); an early end to less May precipitation in Ethiopia (0.46). Normally, the coastal inflow is from the south-southeast, but in the case of wd-4, the wind comes directly from the east. There, the sea is warmer than in the south at this time in April, so the air masses are warmer and more humid. This is reflected in the DWI and DTI as well. Overall, wd-4 is a cold, and especially over the Gulf of Aden moist cluster in the middle of the MAM season.

wd-5, as the first cluster of the MAM season, is the last cluster of the "Bega" dry season. It is particularly dry, both from absolute values and with an average DWI of 1.28 (Fig. 14). Thus, wd-5 correlates with both the number of dry days in Kenya at 0.41 and the March precipitation of Kenya (-0.45), Tanzania (-0.41), and Rwanda (-0.44). If the cluster extends well into March, the precipitation of Burundi (-0.49), Kenya (-0.42), and Rwanda (-0.42) is lower. The wind comes from the northeast during wd-5 in the Horn of Africa; normally it shifts to an easterly wind at the equator and goes inland, but here it maintains its coast-parallel course, blocking moist air that would otherwise flow inland. Overall, wd-5 is the driest cluster north of the equator.

wd-6, on the other hand, is the last cluster to appear. It is the dominant cluster of the last weeks of May, only rarely another cluster occurs at that time, therefore it has a DWI close to 1 (Fig. 14). In absolute terms wd-6 is rather dry. Consequently, when it occurs slightly earlier than normal, May is drier, which is reflected in the May precipitation of Kenya (-0.40), Burundi (-0.58), Rwanda (-0.64), and Tanzania (-0.58). Accordingly, its frequency correlates with the dry day count of Burundi (0.52) and

Rwanda (0.49) and with the end of the rainy season of Burundi (-0.59) and Rwanda (-0.48). The wd cluster 6 occurs 43.5 % more frequently in 2018 than in 1980 (see Fig. A2), consistent with the trend toward a shortening of the rainy season. Overall, wd-6 is the first cluster of the dry season and the driest cluster south of the equator. wd-7 usually starts around the end of April and has its main occurrence in May. If wd-7 occurs clustered, the April precipitation of Djibouti (-0.47) and Eritrea (-0.49) drops. The date of last occurrence slightly influences May precipitation in Rwanda (0.38) and Burundi (0.37). In terms of wind direction, wd-7 looks very similar to normal wind patterns, with no major changes evident to interpret DWI values (Fig. 14). Overall, wd-7 is a rather dry cluster, especially over the Gulf of Aden and in the mountainous regions.

wd-8 is the first cluster of the rainy season. Excess rainfall is particularly evident in Kenya with a DWI of 0.52; this means that almost twice as much rain as normal falls. Figure 14 shows well the double dipole structure of wd-8. On the one hand, there is a wet-wet dipole inland and, on the other hand, a wet-dry dipole between inland and ocean. The latter dipole could be caused by the conspicuous reversal of the wind direction by 180° over the southern Indian Ocean. Normally, at that time in the MAM season, the wind over the ocean comes directly from the south and the wind inland comes from the ocean. However, during wd-8, a wind reversal is present so that the air masses stay longer over the Indian Ocean. Inland, the wind direction is south-west during wd-8, and consequently the air masses come roughly from Zambia, where the peak of the rainy season was shortly before. The very moist, warm air masses hit the East African Highlands, rise and rain down as in the case near Tanzania. Cluster 8 occurs 10.2 days later than previously at the end of the study period, plus it is 46.5 % more frequent (see Fig. A2). This is clearly associated with the trend towards a later onset of the rainy seasons. Overall, wd-8 is the wettest of the wind direction clusters with a DWI of 0.77.

wd-9 is very inconspicuous based on the countries' DWI. However, if Fig. 14 is examined, it becomes apparent that there is a distinct dry region in the northeast of the study area. This is primarily over the ocean, so it is less noticeable. The Horn of Somalia is also affected, but the effect is more than offset by excess rainfall in the south of the country. This is also shown by the negative correlation of wd-9 frequency with Somalia's dry days (-0.43) and that when it occurs frequently, Somalia's rainy season arrives earlier (-0.38). Kenya is also affected by wd-9, Kenya's March precipitation then increases (0.50). The region in the northeast is so dry because the wind in Somalia does not hit the coast as it usually does, but runs parallel to the coast. It is precisely this stretch that is marked by the DWI as being significantly drier. Overall, wd-9 averaged over the region by temperature and precipitation is the cluster closest to normal.

4.4.4. Drying influence of the clusters

Figure 15 provides a summary overview of which clusters tend to be warm/cold or dry/wet. The DWI and the DTI were used for this purpose. The more a cluster corresponds to the normal state, the closer it is to (1|1). The SST clusters are expected to line up along the y-axis due to the clear correlation of SSTs to temperatures. Precipitation, in turn, should arrange along the x-axis. Deviations from this scheme mean that these clusters are exceptional and should be examined more closely.

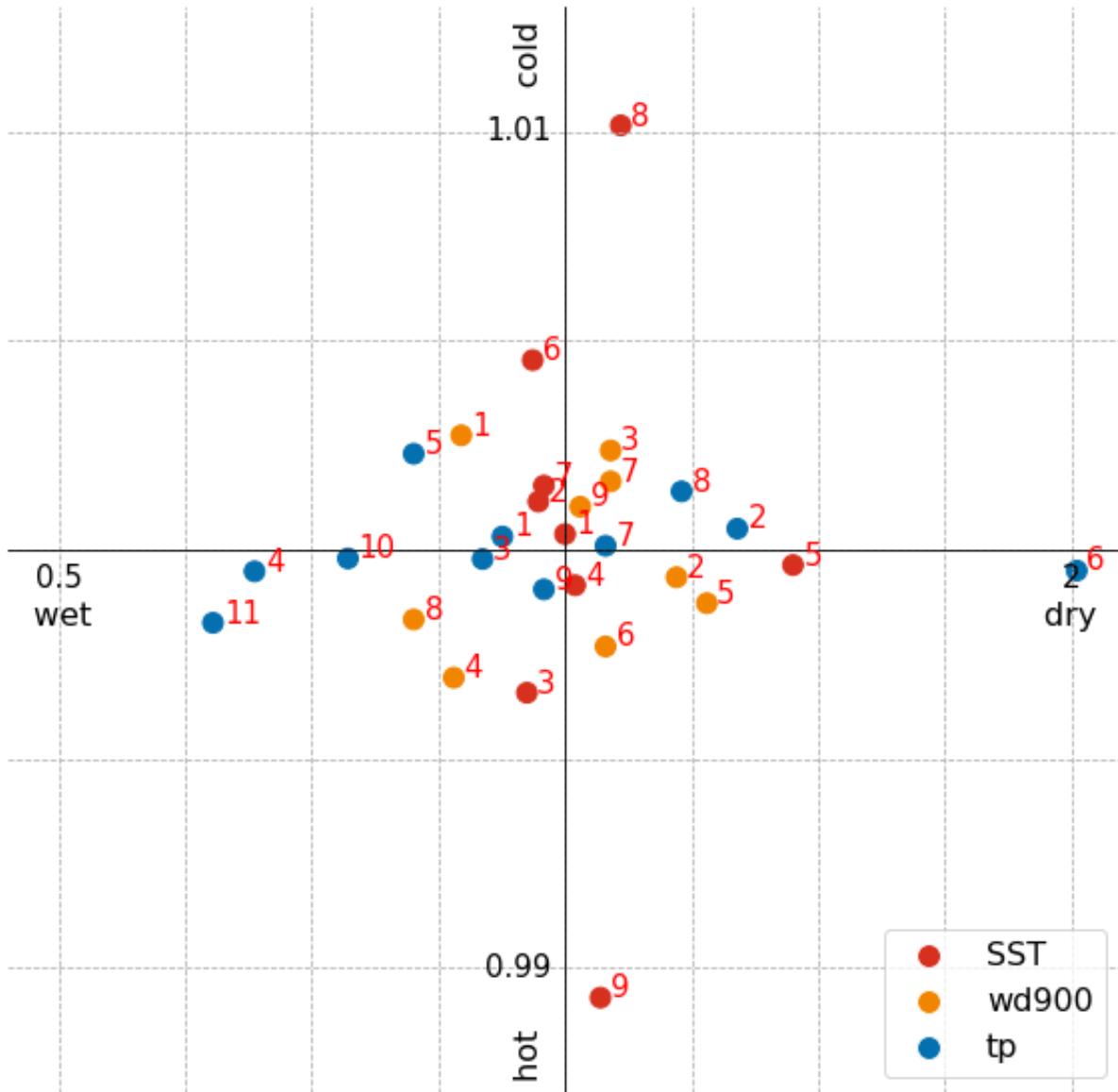


Figure 15: Wetness and heat of all clusters using the Daily Wetness Index (DWI) and Daily Temperature Index (DTI), averaged over the whole area. The axes intersect at (1|1). The red numbers are the cluster numbers. Values of 0.5 correspond to twice as wet/hot conditions as normal, values of 2 correspond to twice as dry/cold conditions, 1 describes the normal condition. Since Kelvin was used for the calculation of the DTI, the values deviate only slightly from 1, the relative distance must therefore be considered.

Accordingly, SST-5 is remarkable; all other SST clusters tend to be close to a DWI of 1, while SST-5 is clearly in the dry range. This is consistent with the characteristics described earlier and makes the cluster stand out even more. In addition, the particularly warm cluster SST-9 and the particularly cold SST-8 also stand out because they are so distinctly different from the other clusters. Regarding the precipitation clusters, tp-6 is to be emphasized, which is significantly drier than all other clusters.

The dryness of tp-6 is also striking in Tab. 3a. While all other tp clusters have a maximum of 20 % of their days falling below mean-std, for tp-6 it is every second day that is considered particularly dry. In contrast, tp clusters 3, 4, 5, 9, 10, and 11 cause almost no dry days. If the wet days are considered, especially tp-4 and tp-11 have a high percentage of days that can be classified as especially wet. 25.9 % of all tp-4 days are even more than twice the standard deviation above the mean. This means that a quarter of the days can be classified as extremely wet. If these clusters occur, the probability of wetter than normal conditions is consequently particularly high. Among the SSTs, SST-5 again stands out most in Tab. 3. While the other clusters are relatively close to the theoretically expected 16 %, one-third of the days of SST-5 are labeled very dry and almost none are labeled very wet. With SST-8, the effect of factoring out the seasonal pattern of precipitation can be well illustrated. Due to its location at the beginning of the study period, SST-8 occurs dominantly during a time when large parts of the study area are still in the dry season. As a result, many days at Mean are classified as very dry, because the mean value for calculating the percentage is rather high since the entire rainy season is factored in. Relative to the time axis, however, it is only slightly drier than average, which is evident by calculating the daily mean (Daily), as the percentage of days drops from 37 to 15 %. This happens due to the adjustment of the mean, which now only considers the time in which the cluster occurs dominantly, the most pronounced part of rainy season is excluded. The same effect is seen when comparing wd-1 to wd-8. In Mean, wd-1 is shown as the wettest cluster in both Tabs. 3(a) and (b). However, when the seasonal influence is removed, wd-8, which is considered average in Mean, is the wettest. For some clusters, removing the seasonality of rain also confirms the picture, e.g. wd-5 is marked as the driest cluster in both Mean and Daily.

4.5. Interactions of the clusters with each other

During the MAM season, the passage of the ITCZ ensures that each year the seasonal sequence of clusters is roughly the same, as SST, wind, and precipitation patterns repeat annually. In the case of SST, a variable that reacts rather slowly to external influences, the sequence is almost the same every

year (see Fig. 16 b). The crucial differences are not in the sequence, but in the duration of the individual clusters and their start and end times. Thus, as already explained in chapter 4.4.2, a significant influence on the precipitation develops.

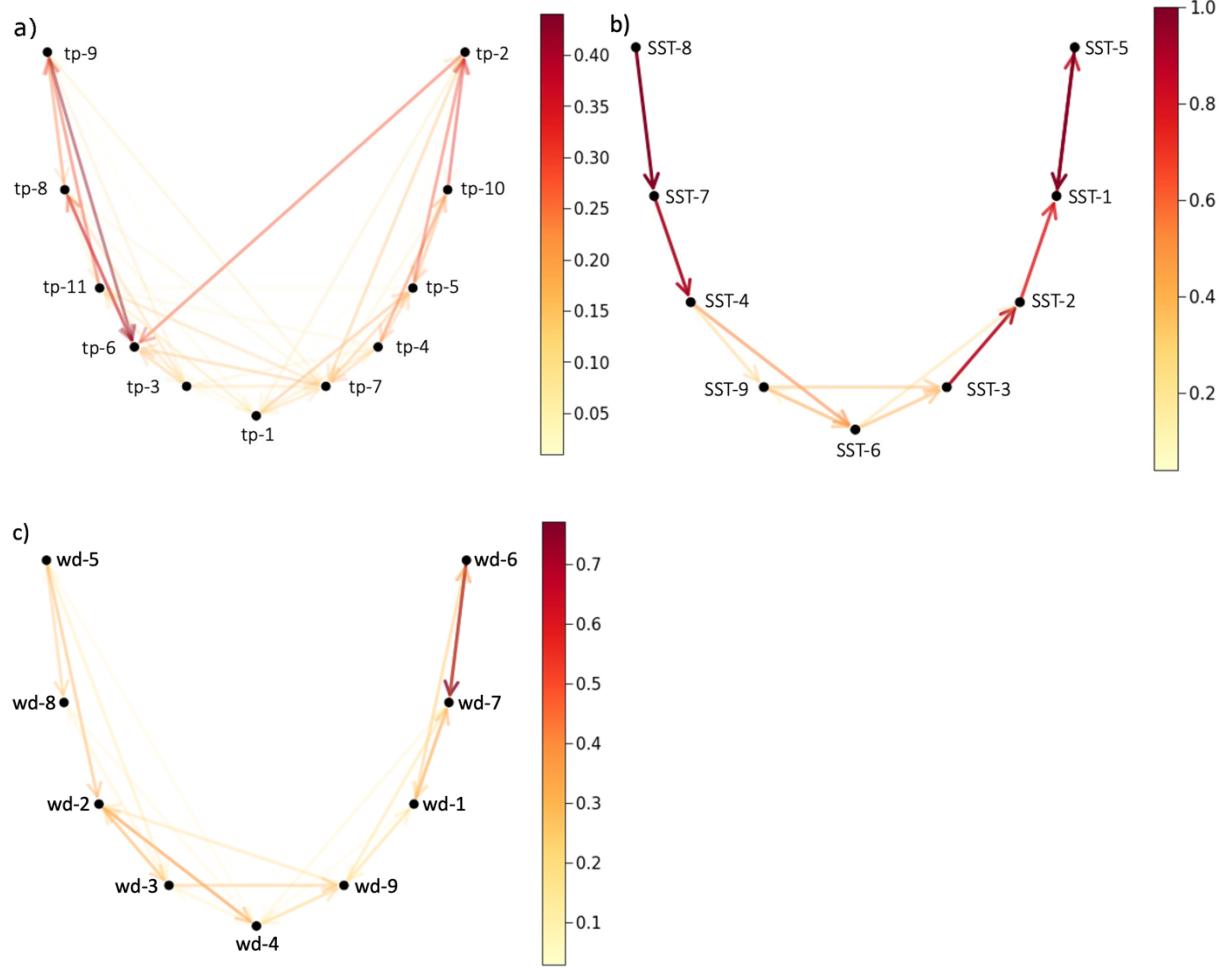


Figure 16: Multi-Directed Graph depicting the transition probabilities [...] of one cluster into another. The transition of a cluster into itself the next day is excluded. a) shows the precipitation clusters, b) the Sea Surface Temperature clusters and c) the clusters of the wind direction at 900 hPa level. The usual sequence during the season is from top left to top right. The shape of a horseshoe was preferred to a network because the tendency to a linear progression is better represented that way. There is no y-axis, so the height of the individual clusters is irrelevant.

The exception here are the clusters SST-3, -6 and -9. These clusters alternate from year to year, when SST-6 occurs frequently, SST-3 and SST-9 are less frequent and vice versa. SST-3 and SST-9 occur together (correlation of frequency: 0.74), to SST-6 the correlation is -0.64 and -0.77, respectively. This is interesting in that the rainfall patterns of the clusters differ; for example, SST-6 is wetter in the east than the other two clusters (see Fig. 13). Consequently, years with dominant SST-6 may mean higher precipitation for this region than in years of SST-3/SST-9. In temperature, this difference is even more pronounced; while SST-6 is the second coldest cluster, SST-3 and -9 are the two warmest clusters (see

Fig. 15). However, there is no clear periodicity in terms of alternating occurrence. SST-2 is also involved, it occurs frequently similar to SST-6 when SST-3 and -9 are rare (-0.61 and -0.50, respectively). However, there is no significant correlation between SST-2 and -6.

The wd clusters are much more dynamic. They also follow the dominant influence of the ITCZ, but a change from the sequence occurs more frequently, especially during the rainy season the different clusters interact a lot with each other. It is noticeable that although wd-2 transitions evenly into wd-3 and wd-4, these two virtually do not interact. This probably happens because they are quite opposite clusters and wd-2 acts as a transition phase. Depending on when wd-2 occurs in the season it changes. If wd-2 occurs early, the transition into wd-3 is dominant; if it occurs late in the season, wd-4 is preferred. There is no evidence for annual or otherwise periodic alternation between wd-3 and -4. The sequence of tp clusters (see Fig. 16 a) has even more dynamics because precipitation can differ significantly over a few kilometers. For example, while a temporal regression, e.g., from SST-7 to SST-8, rarely occurs in SST, tp-11 frequently merges into tp-9, although it is actually behind tp-9 in time on average. tp-6 should be considered separately again, because it is the only cluster that can always occur from the first to the last day of the MAM season. In addition, it occurs much more frequently than the other tp clusters, so that the transition probabilities to tp-6 increase significantly. Thus, tp-6 arises primarily from clusters 2, 8, and 9 and itself transitions back primarily to cluster 8.

SST-1, like SST-5, also promotes the occurrence of tp-2 (0.37), but not that of tp-6. Instead, cluster tp-8 occurs more frequently (0.45). It is also considered dry, but significantly less so than tp-6, which could be where the difference in wetness between SST-1 and SST-5 comes from. The fact that SST-4 is slightly drier than normal could be because it reduces the frequency of occurrence of the moist cluster tp-11 (-0.43). The SST clusters also have an influence on the wind direction clusters. When SST-5 occurs frequently, wd-5 becomes less frequent (-0.53) and wd-6 occurs more frequently (0.56). This could be because the overall long rains shift forward, so that with wd-5 the first wd cluster of the season is pushed into February and the last cluster wd-6 gets more weight. If the combination SST-3/SST-9 occurs, there is also a decrease in the frequency of wd-7 (-0.42). The wind direction then does not turn to the southwest like a monsoon, but hits the coast directly (cf. Figs. 13 and 14).

In addition to SSTs, wind also has significant effects on precipitation clusters. Tab. 4 shows this for all cluster combinations. Except for tp-1 and tp-4, at least a slight correlation to one or more of the wind clusters can be inferred for all clusters. A stronger effect has wd-1, which mainly promotes the formation of the wetter clusters tp-5 and tp-11 and hinders the formation of the drier clusters tp-2 and tp-8 (see Tab. 4). wd-9 also has an effect on several clusters, important here is the correlation with tp-6, which otherwise does not interact significantly with any other predictor cluster. In addition, wd-8 has a significant impact on tp-9. A comparison of the DWI maps of the two clusters shows that the

south forms almost an identical precipitation pattern (Figs. 11 and 14). There is one connection between the wd clusters themselves that stands out clearly, namely that wd-1 is infrequent when wd-7 is frequent (-0.58). These two clusters are relatively similar in terms of wind patterns (Fig. 14) and also occur nearly simultaneously in the MAM season, but where wd-1 is significantly wetter than normal, wd-7 is particularly dry. Therefore, for these regions, it would be valuable to find an explanation for the irregular periodic occurrence to better understand the preferential formation of each cluster.

Table 4: Correlation of the frequency of the individual wind clusters per year with the frequency of the precipitation clusters. Significant correlations ($\alpha = 0.05$) are marked in red.

	wd-1	wd-2	wd-3	wd-4	wd-5	wd-6	wd-7	wd-8	wd-9
tp-1	0,20	-0,11	-0,15	0,25	0,09	0,10	-0,21	-0,16	-0,10
tp-2	-0,36	-0,01	0,19	-0,16	-0,29	0,33	0,34	-0,16	0,08
tp-3	-0,11	0,33	0,22	0,11	0,05	-0,16	-0,01	0,05	-0,33
tp-4	0,22	-0,04	-0,30	0,16	0,04	-0,31	-0,10	0,16	0,18
tp-5	0,46	-0,01	-0,22	0,04	0,13	-0,04	-0,42	0,15	-0,13
tp-6	-0,18	0,20	0,28	-0,04	0,26	0,04	-0,02	-0,09	-0,34
tp-7	0,00	-0,04	-0,10	0,04	-0,27	-0,01	0,05	-0,18	0,45
tp-8	-0,39	0,02	0,04	-0,11	0,32	0,02	0,30	-0,22	0,10
tp-9	-0,16	-0,15	0,08	-0,09	-0,22	-0,02	0,20	0,57	-0,13
tp-10	0,25	-0,13	-0,13	-0,27	0,25	-0,42	0,04	0,17	0,16
tp-11	0,37	-0,20	-0,23	0,10	-0,28	0,11	-0,28	-0,01	0,24

4.6. Influence of teleconnections

The irregular periodic occurrence found in section 4.5 often comes from the fluctuations of different teleconnections. To test the influence of the teleconnections, first a simple correlation analysis between the indices described in chapter 3.3.2 and the atmospheric variables was performed. In addition to these calculations, it was tested whether the correlation increases when the teleconnections are detrended. This was generally not the case, on average the original values have slightly higher correlation values, so that the detrended index values were not further used. Subsequently, the properties of the cluster days, i.e. start, end and frequency, are compared with the teleconnections in order to test an influence on the position of the individual clusters.

The influence of the teleconnections on the temperature of the study area is clearly evident: The Western Indian Ocean Dipole (W-IOD) has a particularly strong influence with a correlation of 0.81 to the mean temperature of the MAM season. This was to be expected, since air temperature and SST

interact strongly. The correlation also extends to daily values. The temperature of the hottest day of the season correlates with the El Niño Southern Oscillation (ENSO) (0.31–0.53, except for the northernmost two countries of Sudan and Djibouti), significantly with W-IOD (0.43–0.58), and not with the Indian Ocean Dipole (IOD) (both with the March month of the teleconnection indices). Over the sea, of course, this correlation is stronger (ENSO: 0.56; W-IOD: 0.80). No significant correlation can be seen with the appearance day of the hottest day.

A strong interaction is also to be expected between the teleconnections and the average air pressure. This is also present, but ENSO has the strongest influence (0.70). Thus, the wind direction also experiences an influence of ENSO. With a negative W-IOD, the sea wind at 900 hPa comes from the east, with a positive one rather from the south. Potentially, this results in less inflow to the mainland from the moist ocean air masses with a positive W-IOD, so the coastal region could become drier. However, this could be compensated somewhat, since evaporation is higher due to the higher ocean temperatures.

Regarding the precipitation, the teleconnections also have an influence, this is further elaborated with the following examples. Low IOD brings higher rainfall in the northeast of the study area (Djibouti - 0.54; Eritrea -0.41; Ethiopia -0.34). Thus, a low March IOD results in fewer days of drought in Djibouti (0.49). No other effects of IOD are apparent. However, this is not true for the two individual indices, the Western and Eastern IOD. For example, the March W-IOD is positively correlated with rainfall in Tanzania (0.36), leading to an increase in dry days (0.37). In addition, March rainfall in Sudan and South Sudan decreases with a strong W-IOD. The Pacific Ocean also has an influence by means of ENSO and PDO. With a higher PDO in March, May rainfall is greater in the north of the region, especially in Ethiopia (0.49), but also in Djibouti (0.39) and Eritrea (0.36). With a positive ENSO, rainfall is increased in Kenya (0.43). To this end, March's ENSO is positively correlated with the most extreme heavy rainfall event of the season, significantly at the coast (e.g., Kenya 0.34) and over the Indian Ocean (0.43). However, no correlation is significant inland. In contrast, the influence of the Atlantic Ocean is more pronounced there. With a low E-ATL, April precipitation increases in Burundi (-0.34), Rwanda (-0.41), Sudan (-0.45), and South Sudan (-0.51). In March and May, however, precipitation is hardly influenced, only in Tanzania (-0.35). ATL does not have any significant influence. Last, the MJO also has an influence. With a more pronounced MJO, March precipitation increases in Kenya (0.33) and Somalia (0.38) and April precipitation of Ethiopia (0.34), Kenya (0.36) and Somalia (0.40). In contrast, the main precipitation of Djibouti (-0.35), Eritrea (-0.34), and Ethiopia (-0.38) decreases. Due to the high correlation between teleconnections and clusters, the correlations of SPEI are slightly higher than those of precipitation.

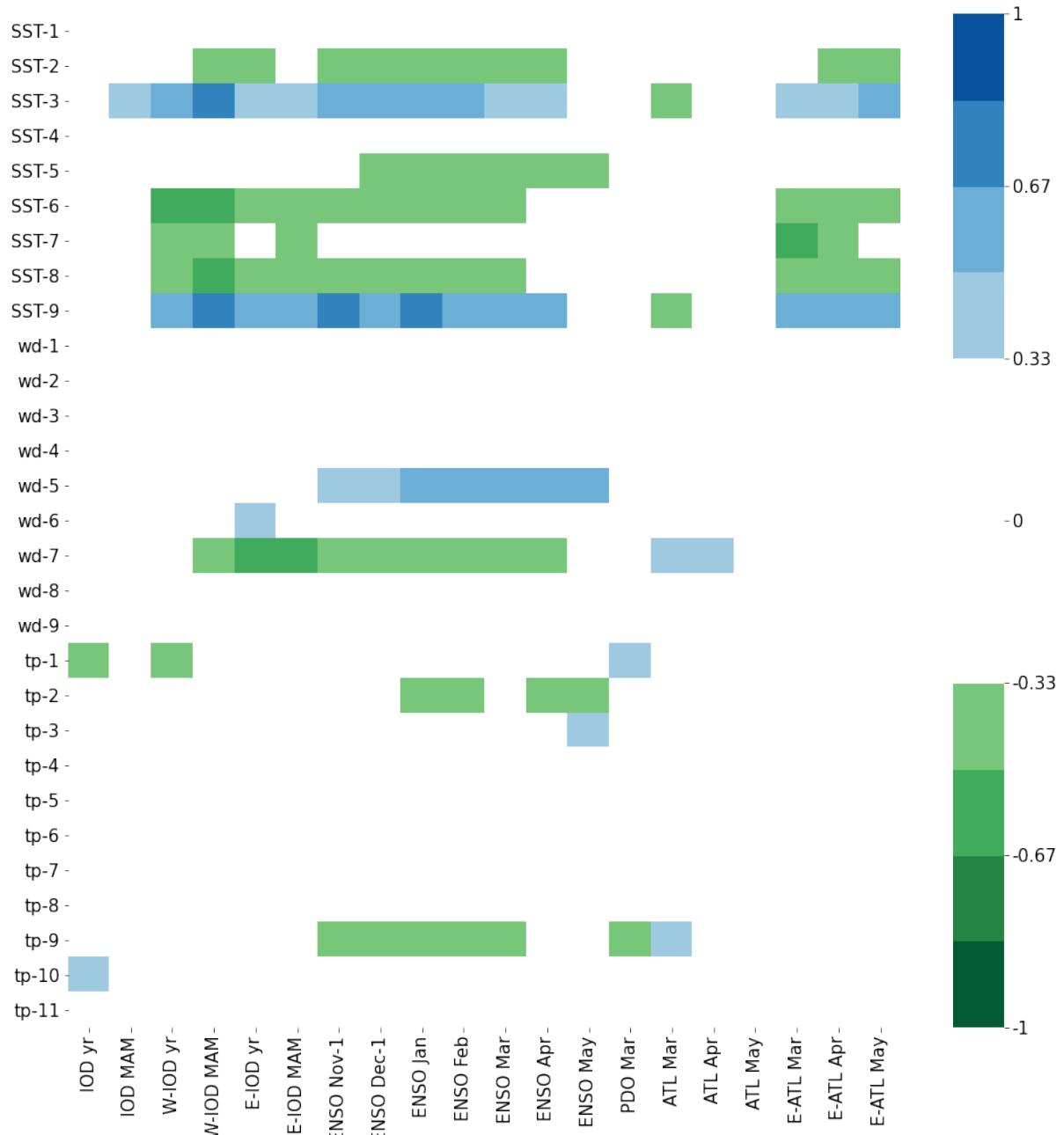


Figure 17: Heatmap of the correlation of the frequency of the individual clusters with the teleconnections. Teleconnections are plotted by specific months or MAM season. Nov-1 and Dec-1 means that November/December data from the previous year was used. A correlation is significant at 0.32 and -0.32, respectively.

According to the above analysis, there is only a slight correlation between IOD and precipitation. This and all other influences of the teleconnections shall now be examined in more detail by means of the analysis of the cluster properties. For this, Fig. 17 illustrates the relationship between the teleconnection indices and the frequency of the individual clusters. As expected, the correlation between the SSTs and the teleconnections is relatively high. It is interesting to note which clusters are positively correlated and which are significantly negatively correlated. While it was originally assumed that mainly IOD and ENSO influence the cluster formations, SST-7 shows that Atlantic Ocean temperature can also have

an influence. While the relationship with teleconnections is relatively clear for SSTs, the influence on wind cluster formation is only indirect. This makes wd-5 particularly interesting because it shows a clear correlation to ENSO and the W-IOD. The correlation to ENSO is already clearly present in December of the previous year (0.47). In addition, wd-5 is the driest cluster, both by DWI and in absolute terms, it has a direct impact on precipitation, especially in Somalia (DWI of 1.63) and Ethiopia (DWI of 1.42). This provides a new way of looking at the interaction between in this case ENSO and precipitation. Such correlations could possibly make local precipitation forecasts more accurate, because knowledge of the currently prevailing wind patterns allows conclusions to be drawn about the areas particularly affected by a lack of precipitation. Another example is wd-7, which is less common with positive E-IOD. wd-7 is partly responsible for April precipitation in the region. The surprising part is that E-IOD has a stronger correlation (-0.55) with wd-7 formation than W-IOD, which is spatially closer. Even ATL (0.37) and ENSO (-0.42) have a more significant influence than the W-IOD. wd-9, which is a factor in the amount of precipitation in Kenya in March, on the other hand, is again more dependent on the W-IOD. If the W-IOD is negative, the first occurrence date of wd-9 is later in the MAM season (-0.35). The frequency of its occurrence is more correlated with ENSO of February (0.36). The correlations with the first and last occurrence dates are not recorded in Fig. 17, so the significant ones are presented in Tab. 5.

Since the teleconnection indices replicate ocean temperatures from specific locations, it is expected that the relationship with the SST clusters is greater than with the tp or wd clusters, even if none of the teleconnections are located inside the study area. This is proven true, the clusters most likely to be influenced by the teleconnections are SST-3, SST-6 and SST-9, recognizable by the dark colors in Fig. 17. This illustrates their correlation mentioned in Section 4.5. If the W-IOD is positive, SST-3 and SST-9 occur more frequently. With negative W-IOD, i.e. colder temperatures of the Western Indian Ocean, SST-6 is formed more often. This can already be observed in February, the correlations are with 0.65; -0.54 and 0.74 only marginally weaker than in March. Due to the strong correlation with the teleconnections and a clear statement about the temperature of the clusters (SST-6 is the second coldest cluster and SST-9 by far the warmest, Fig. 12), it should already be predictable in February whether April will be warmer or colder than usual. With regard to precipitation, the statement is not so clear, because SST-3 and SST-6 are both rather wet, SST-9 rather dry (DWI: 0.93; 0.94; 1.07). In addition, about the same regions are particularly wet in SST-6 that are particularly dry in SST-9. However, with the help of the teleconnections, it can be determined that when the W-IOD is high in February, SST-3 (-0.43) and SST-9 (-0.69) appear earlier in the season. For SST-3, E-ATL is also a factor (-0.40) in addition to W-IOD. The date of occurrence of SST-6 has no correlation with W-IOD; it correlates only with PDO (-0.49). An early occurrence of SST-6 mostly means more rain in April especially for Somalia, Kenya, and Ethiopia, hence the rainy season starts earlier. The end of the phase also depends on the W-IOD,

-0.57 for SST-3 and -0.45 for SST-6. Accordingly, the duration clearly follows the W-IOD (0.72), although here the E-IOD (0.45), the ENSO (0.57), the May E-ATL (0.49), and the ATL of March (-0.42) also influence the frequency of SST-3. The values of SST-9 are even slightly stronger, while those of SST-6 are slightly weaker and with opposite sign due to the alternation with the other clusters.

The strongest phase of the rainy season is during SST-3. If it does not occur in a year, as already explained in chapter 4.5, SST-2 is the cluster of the main rainy season. Cluster 2 is highly dependent on ENSO, the monthly values from December of the previous year to April correlate with the date of occurrence of SST-2 with more than 0.64 each. In addition, the start of SST-2 correlates with the W-IOD of May with 0.61 and with the March ATL with -0.49. Thus, the formation of cluster 2, which in contrast to SST-3 occurs every year, shifts back several days in the case of positive ENSO and W-IOD values. In addition, the number of days assigned to cluster 2 decreases in such years (ENSO Dec: -0.46; W-IOD Jan: -0.49).

Another interesting connection exists for the first cluster of the MAM season. Under El Niño conditions with an ENSO index above 0.5, SST-7 tends to be preferred over SST-8. This should mean a rather wet start to the MAM season, especially for Kenya and Ethiopia, where SST-7 is much less dry compared to -8. In contrast, a dry start is more likely for negative ENSO conditions < -0.5 . This was tested using the Mann-Whitney U test at $\alpha = 0.05$ significance level. How strongly negative ENSO is, is not decisive, provided there are no neutral conditions between -0.5 and 0.5, which have no predictive power on March's start. This relationship is reflected in the influence of ENSO on the first day of SST-7 (-0.44 in February). Another influence is by W-IOD, which also correlates with the start, even more strongly than ENSO at -0.60. The influence of W-IOD can be found in the frequency of SST-7 (0.45). With SST-8 it behaves similarly, at -0.63 with the February W-IOD and -0.49 of the November ENSO. The end of this initial phase with dominant influence of SST-7 and -8 is also decided by the W-IOD (-0.80 and -0.63, respectively), but E-ATL also plays a role at -0.60. Consequently, with warmer sea temperatures than usual, the initial phase of dryness ends more quickly.

After this phase, SST-4 usually occurs. This cluster, which forms the beginning of the rainy season in several countries, is strongly influenced by the W-IOD in particular. This manifests itself in an earlier start (-0.74) and an earlier end of the cluster (-0.55) when the W-IOD is positive in February. E-ATL plays a minor role in the start (-0.55) and end (-0.35). Interestingly, the time between the start and the end does not depend on any teleconnection. This could be due to the fact that the duration is always approximately the same and the influence of the teleconnections is only reflected in a shift of the time period.

Also interesting is the end, the last cluster of the MAM season. Usually SST-5 is the conclusion of the MAM Season. This is especially true with negative ENSO, in 17 of 20 cases the season then ended with SST-5. With positive ENSO, SST-1 is more preferred, the relationship is confirmed by the Mann-Whitney U test. Since SST-1 tends to be neutral and SST-5 very dry, this could indicate that dry conditions are more common than usual when ENSO is negative, especially in Somalia and Kenya. This fits well with the precipitation data; the correlation between the number of days assigned to cluster 5 in May correlates significantly with April ENSO (-0.45) and with actual precipitation in Ethiopia (-0.32), Kenya (-0.55), Somalia (-0.43), and Tanzania (-0.54). The negative correlation with ENSO is already significant even in December of the previous year with -0.34. The other teleconnections like the IOD and PDO do not have a statistically significant impact with a correlation of about -0.23 each, but an influence in a subordinate role to ENSO is possible. SST-1 is also influenced by ENSO, with the cluster appearing earlier at negative ENSO values (0.48). The effect of SST-1 and SST-5 can possibly be explained by the favoring of wd-7 (0.41) and tp-8 (0.45) by SST-1, and the favoring of wd-6 (0.56) and the hinderance of tp-10 (-0.42) by SST-5. Over the years, the number of days in SST-1 and -5 tends to increase, which corresponds very well with the general decrease of rain in the same period. There is a correlation of -0.63 with the precipitation of the complete study area. March and April are not affected by this conclusion, their precipitation is independent of SST-1 and -5. June, however, is affected almost as much as May, if SST-5 occurs frequently at the end of May, the precipitation in June decreases significantly (-0.43), so that the rainy season ends earlier. The earlier SST-5 starts, the earlier the rainy season ends. This affects the entire region (0.49), Somalia especially (0.53). Here, the start date correlates strongly with May IOD (0.48), somewhat weaker with April IOD (0.46), and with April ENSO (0.37). Thus, there is a good possibility that ENSO and IOD can be used to predict several months in advance whether May in the Horn of Africa will be wet or dry and when the rainy season will tend to end.

For the precipitation clusters, the dependence of the first and last day on ENSO is not so pronounced. The W-IOD has a clear effect for it. A negative Western IOD of December of the previous year favors the rather humid tp cluster 9 at the beginning of March. A positive W-IOD, on the other hand, leads more often to the very dry tp-6. If the W-IOD is clearly positive, tp-9 does not occur. This can also be seen in the date of the first occurrence of tp-9; if the W-IOD is positive, the date moves back (0.42). The E-IOD has about the same influence (0.44), so the occurrence depends on the SSTs of the whole Indian Ocean. Subordinately, the frequency is also influenced by ENSO (December: -0.33) and the March ATL (0.35). tp-6, on the other hand, occurs independently of the teleconnections. Consequently, it is mainly tp-9 that determines the start of the season; if it remains absent, the gap is filled by tp-6. While this is significant, the effect of SSTs is more dominant.

If the W-IOD and/or E-IOD is clearly negative, then the MAM season ends with tp-2. Here it depends again on the absolute temperature of the Indian Ocean, the IOD is not important. In contrast, with warm Indian Ocean temperatures in December of the previous year, tp cluster 6 is more likely to be favored, and consequently it becomes drier. ENSO has no influence on it, but correlates with the frequency of tp-2 (-0.35). As already explained, tp-6 shows virtually no dependence on a teleconnection, the dependence of the tp-2 start date on the E-IOD (0.33) and the W-IOD (0.39) is the decisive factor which of the clusters occurs.

Further correlations of precipitation with teleconnections can be found in Tab. 5. It is worth highlighting the very wet tp-11, whose last occurrence is clearly dependent on the March IOD (-0.50) and the ENSO as early as December (0.46). However, some clusters, e.g., tp-3, do not interact with the clusters of their own variables, other variable clusters, or any teleconnections. It is therefore difficult to determine the occurrence of these clusters. In general, these are rather average clusters with little influence on the precipitation variability, but to understand the whole system it is necessary to be able to trace also these clusters.

In terms of wind direction clusters, March tends to start with wd clusters 3, 5, or 8. wd-5 has a very dry DWI, wd-8 a very wet one. While the frequency of wd-5 is clearly dependent on ENSO (0.50 as early as January), there is still no clear signal of a teleconnection that could give a clue about the start of the MAM season. There is a tendency for wd-5 to start when W-IOD is negative, but this is only significant at the $\alpha = 0.1$ level. The ENSO signal is ambiguous. Since wd-3 also occurs as a start date, a subordinate influence of IOD cannot be ruled out, as it correlates with the start date of wd-3 (-0.38). Nothing can be concluded about the start date of wd-8, since the start is not significantly correlated with any teleconnection. The end, on the other hand, does, with warm water temperatures of the W-IOD (-0.46) and the E-ATL (-0.35) ending cluster 8 earlier.

May, on the other hand, is more distinct, always ending with wd900 cluster 6 or 7, with wd-6 accounting for the bulk of 90 % of cases. The difference between the two clusters manifests itself primarily in more precipitation to the east over Somalia and the Indian Ocean at wd-6. Otherwise, the clusters are similarly slightly dry. Cluster 6 promotes the occurrence of extremely hot days north of 12° N; wd-7 tends to occur around 10° N. In addition, wd-6 slightly inhibits the occurrence of the wet cluster tp-10 (-0.42), wd-7 that of tp-5 (-0.42), which is also wet. In the formation of the clusters, ENSO and the E-IOD are the most important influences. If the E-IOD is negative, wd-7 is formed more frequently (-0.55), and the same is true if ENSO is negative (-0.44). The correlation is already established in March and February, respectively. The formation of wd-6 is independent of the teleconnections. On the other hand, it is dominantly influenced by SST-5. If SST-5 occurs frequently, the formation of wd-6 is favored (0.56). Since the frequency of SST-5 also depends on ENSO, there is a logical connection. If ENSO is

clearly negative, the MAM season ends with wd-6. If ENSO is strongly positive, the probability that wd-7 can occur in addition to wd-6 is significantly higher. In addition, with positive ENSO, the previously occurring cluster wd-1 is also pushed back (0.43). Thus, ENSO has a significant influence on the last clusters of both predictor variables. If ENSO is negative, particularly dry clusters emerge. wd-6 and wd-7 occur earlier and earlier in the season, especially cluster 6 occurs about half a day earlier per year on average. This could be a trigger for the tendency to advance the end of the rainy season.

Overall, Fig. 17 also provides an indication of whether the gradient between two SSTs is the dominant factor, as in the IOD or ATL, or the absolute water temperature, as in the W-IOD or E-ATL. The colors in the columns of the W-IOD and the E-IOD are darker overall than those of the IOD, meaning they correlate more strongly with abundance. In the Atlantic, the situation is similar, the temperature gradient ATL has only a low correlation, while the absolute temperature E-ATL has a much higher correlation overall. Similarly, if the teleconnections are correlated with the start or end date, it can be assumed that the influence of the absolute water temperatures is the more important one.

Table 5: Correlation between the properties of the individual clusters and the teleconnections. If a teleconnection correlates significantly ($p\text{-value} < 0.05$) with a property, it is listed; the order corresponds to the strength of the correlation. The Occurrence Frequency is shown in detail in Fig. 17. The frequently occurring teleconnections ENSO and W-IOD are highlighted in green and blue.

	<i>Occurrence Frequency</i>	<i>Date first Cluster</i>	<i>Date last Cl.</i>
SST-1	-	ENSO, ATL	ENSO
SST-2	ENSO, W-IOD, E-ATL	ENSO, E-IOD, ATL, W-IOD, E-ATL, PDO	ENSO, ATL
SST-3	W-IOD, ENSO, E-ATL, E-IOD, ATL, IOD	W-IOD, ATL	ENSO, ATL
SST-4	-	W-IOD, E-ATL, E-IOD, ENSO	W-IOD, E-IOD, ENSO, E-ATL
SST-5	ENSO	ENSO	-
SST-6	W-IOD, E-ATL, ENSO, E-IOD	-	W-IOD
SST-7	E-ATL, W-IOD, E-IOD	W-IOD, ENSO, E-ATL, E-IOD	W-IOD, E-ATL, E-IOD, ENSO
SST-8	W-IOD, E-IOD, ENSO, E-ATL	-	W-IOD
SST-9	W-IOD, ENSO, E-IOD, E-ATL, ATL	W-IOD, E-ATL, E-IOD	ENSO
wd-1	MJO	ENSO	ENSO
wd-2	-	-	-
wd-3	-	-	ENSO

<i>wd-4</i>	-	-	-
<i>wd-5</i>	ENSO	-	ENSO
<i>wd-6</i>	-	-	-
<i>wd-7</i>	E-IOD, ENSO, W-IOD, ATL	-	ENSO
<i>wd-8</i>	-	-	W-IOD, E-ATL
<i>wd-9</i>	-	W-IOD	ENSO
<i>tp-1</i>	PDO	ATL, ENSO	-
<i>tp-2</i>	ENSO	E-ATL, E-IOD	PDO
<i>tp-3</i>	-	-	-
<i>tp-4</i>	-	ATL, PDO	-
<i>tp-5</i>	-	E-ATL	-
<i>tp-6</i>	-	-	-
<i>tp-7</i>	MJO	-	-
<i>tp-8</i>	-	E-ATL	ENSO
<i>tp-9</i>	ATL, ENSO, PDO	E-IOD, W-IOD	ENSO
<i>tp-10</i>	-	E-IOD	-
<i>tp-11</i>	-	-	ENSO, E-IOD, PDO

5. Discussion

In this work, a cluster analysis is used to test a new approach to study the MAM season in terms of circulation patterns. The predictor variables identified as most suitable for clustering are, first, the wind direction at 900 hPa level (wd900). LAUX et al. (2020) used a similar approach, focusing on the v-wind at 700-hPa level combined with the mean sea level pressure. The wd900 can indicate the onset and end of the rainy season by showing the location of the ITCZ, and, on the other hand, can indicate wind effects such as a for a month atypical wind direction parallel to the coast, which prevents the moisture of the sea from advancing far inland. The second predictor variable is Sea Surface Temperature (SST), which primarily influences precipitation with an early or delayed onset of certain SST patterns. The Cluster analysis confirms the trend toward less precipitation that has persisted since 1998. For the

countries Ethiopia, South Sudan, Somalia, Kenya, Uganda, Tanzania, Burundi and Rwanda, the cluster analysis gives a deeper understanding of the influence of the different drivers that affect the rainy season. The rainy season of Djibouti, Eritrea, and Sudan is only partially within the study period, thus the conclusions about their MAM season are not applicable to the complete rainy season; for that, June would have to be additionally studied. The clusters can be analyzed with regard to their moisture using the Daily Wetness Index (DWI), so that statements can be made about whether precipitation is favored or hindered. Based on the calculation of the DWI, the influence of the clusters can be decoupled from the course of the rainy season, so that a clear gain in knowledge is achieved especially in the transition periods between dry and rainy season. This complements the approach of Laux et al. (2020), who also located clusters for the northeast region of Brazil that are linked to dry and wet conditions, but did not compare them with each other for their impact.

For some characteristics of some clusters (e.g., their start date, end date, or frequency), a trend can be seen several months in advance through the teleconnections. Through the interactions of the clusters with each other and the analysis by means of the DWI, precipitation deficits and surpluses can be inferred partly one month in advance and partly already in December of the previous year. For example, the main precipitation of the entire region, especially at the level of the equator, correlates strongly with the date of the first occurrence of SST clusters 1, 2, and 5, which in turn correlate strongly with ENSO. La Niña phases thus tend to lead to an early end of the rainy season. Particularly affected are Kenya and Somalia, Tanzania to a lesser extent. The correlation is most pronounced with ENSO of May, but already in December of the previous year the correlation is significant. Thus, six months in advance, the first indications could point to an earlier end of the rainy season than usual. This would be important information because the end of the rainy season is important for total precipitation (0.38), especially in the southern regions.

Getting a good estimate for the start of the rainy season would be even more important, since the total amount of rainfall for the Long Rains is even more significantly dependent on the start than the end, according to the calculations (-0.45), confirmed by CAMBERLIN et al. (2009). At the start, a preference of a certain SST cluster by positive ENSO values can also be found, but the alternative cluster is similar to it in terms of the impact on precipitation, so that no reliable statement can be made to what extent the clusters influence precipitation. Besides ENSO, the crucial teleconnection is the Western IOD, as it has a large influence on the SST clusters. This manifests itself particularly with respect to temperature. The variation of the average temperature of the region over the years almost parallels the variation of the W-IOD. For the precipitation distribution, the W-IOD is also important due to its correlation with the SST clusters, the E-IOD to a lesser extent. This confirms the results of ENDRIS et al. (2015) who, using CORDEX model data, also concluded that W-IOD contributes to the precipitation of

East Africa. UMMENHOFER et al. (2009) were also able to make a similar statement about the SSTs during the short rains. This thesis shows that other teleconnections like the MJO, the PDO or the temperatures of the Atlantic Ocean play only minor roles, ENDRISS et al. (2015) agrees. Whether the start and end statements can be of use in forecasting the rainy season remains to be determined. But it is certain that single factors cannot be assumed as the only source of variability; several variables must always be included in the calculations, which also differ in time and place, sometimes significantly, CAMBERLIN et al. (2009) agrees.

Another difficulty here could be that the correlation between SST-5, the cluster responsible for the dry season start, and ENSO has tended to break down since 2003. While from 1980–2002 there was still a very strong underlying interaction (correlation of 0.82), this has declined significantly over the last 15 years of observation. The timing of this break corresponds exactly to the point in time from which the influence of climate change on the GHA became increasingly strong. Whether this is merely a coincidence or whether climate change is introducing disorder into the system in this way remains to be investigated in a future study. Nonetheless, the findings regarding the reasons for the early end of the rainy season are very important because it allows for a much better understanding of the mechanics of the MAM season. In addition, the findings regarding the interrelationships of the various clusters, particularly the SST clusters with the tp clusters, shed new light on how and to what extent Sea Surface temperatures influence precipitation in specific regions. This shows that cluster analyses, as previously shown by LAUX et al. (2020), can improve our understanding of the climate system.

The analysis of the location of the greatest influence of the clusters in relation to the teleconnections is on the one hand a strength of the cluster analysis. It makes it possible to learn about correlations that are not apparent when looking only at precipitation. However, it also carries a weakness. Because the process of clustering primarily groups similar patterns into a cluster, information about the progression of the pattern's strength is lost because the cluster centroids merely reflect an intersection. For example, tp-11, the wettest of the tp clusters, does not significantly change its abundance over the years. Without further investigation, this would indicate that the length of the wet season would remain roughly the same. However, if we examine how precipitation behaves during the days assigned to the cluster, a significant trend toward lower precipitation is observed. Thus, for the most part, it is not the wet clusters that become less frequent and the dry clusters that become more frequent, but instead many of the clusters produce significantly less precipitation. This complicates the issue because in addition to frequency, the internal variability of the clusters must also be considered. Another weakness of the cluster analysis is that the ECV of precipitation is relatively low compared to the wind clusters and the SST clusters due to the high spatial differentiation, so that a misinterpretation of the cluster assignments cannot be excluded. Therefore, the statements still have to be checked for their

robustness. Some correlations of the clusters are not completely transferable to the rain, because the correlation then slips below the significance level. For example, for the start of March, the conclusions are more or less in line with the trend, but the precipitation does not fit perfectly, the significance level is more in the range of $\alpha = 0.1$. Therefore, it is difficult to evaluate the usefulness of the clustering well, further research in this area is needed to make solid statements.

Besides the SST clusters, the wind direction clusters also have an influence on precipitation. The characteristics of some wd clusters correlate well with the precipitation of the countries on the one hand, on the other hand they also lead to the formation of special tp clusters, which in turn influence the precipitation. One country particularly influenced by wd clusters is Kenya. Six of the nine wd clusters significantly influence precipitation with the frequency of their occurrence, four of them even over the entire MAM season. Of particular note is wd-1, which in addition to Kenya also influences Burundi, Rwanda, Somalia, and parts of the Indian Ocean. This marks an advance over YANG et al. (2015), where wind direction has already been studied at multiple levels, but only resolved at a seasonal scale and with less reference to the intraseasonal influence of wind direction.

One issue that arose from the topographic location of the study region was whether and to what extent the complex topography affects the cluster analysis. While differences due to mountain ranges were still clearly noticeable when looking at the absolute deviation of the individual tp clusters from the Mean, the DWI is a suitable tool to put the influence of topography on precipitation into context. While the effects of the mountains are still evident in some clusters in the DWI, their impact is substantially smaller. Thus, the effects of mountains on the clustering of precipitation are less dominant than initially expected. This contrasts with the findings of JORBA et al. (2004) who found an impact of the Pyrenees on the cluster of Barcelona wind trajectories and thus identified an influence of a mountain range on the wind direction clusters. For example, the highest elevation in the study area, Mount Kilimanjaro, only affects a maximum of six pixels. With respect to the wind clusters, the influence of the mountain ranges is weak in that only small day-to-day differences occur there during the MAM season, so the wd clusters are not influenced that much. As a comparison, the diurnal differences at sea are about four times stronger for the u and v winds (ERA5 data) and have a significantly stronger influence on the assignment to a certain cluster. Thus, the wind clusters are more likely to represent the wind systems over the Indian Ocean, which are more important to the overall weather pattern of the GHA as they transport moisture.

Furthermore, this work shows that there is a clear drought trend during the MAM season. This is especially true for the drought index SPEI, which was particularly low between 1998 and 2012. While this has already been confirmed by a large number of authors (z.B. LYON and DEWITT, 2012; GLEIXNER et al., 2017), the attribution of the reduced precipitation to a statistically significant shortening of the rainy

season of most countries has not been clearly described so far. The earlier ending of the rainy season also manifests itself in the clusters in that the entire chain of SST clusters shifts slightly forward. This is particularly pronounced in La Niña years. This also affects the June precipitation of Ethiopia and the other northern countries. This fits with the statement of KORECHA and BARNSTON (2007), that summer precipitation is determined by ENSO. In addition, in the north of the region the rainy seasons are influenced by the W-IOD, warm water temperatures in the Indian Ocean lead to a shortening of the rainy season.

5.1. Outlook

The findings on clusters can be applied to a seasonal forecast model. The monthly issued forecasts of the SEAS5 model, produced by the *European Centre for Medium-Range Weather Forecasts* (ECMWF), could be well used for this purpose. On the one hand, the forecasts can be checked specifically with respect to the teleconnections. Due to the sometimes strong correlation with the cluster distributions, conclusions could be drawn about e.g. the intensity or the end of the rainy season months in advance by an accurate forecast of the W-IOD and ENSO. In addition, Seasonal to Subseasonal forecasts (S2S) could predict, for example, the occurrence of specific wind clusters and thus precipitation distributions in shorter time windows. The SEAS5 forecasts can be compared to the clusters and assigned to the one that most closely matches. Since SEAS5 has 51 ensemble members, this would have to be done for each member individually. This gives a probability distribution of which cluster is most likely to be dominant in the near future and how the transitions between the individual clusters behave. Whether SEAS5 and S2S forecasts can form patterns similar to the clusters remains to be evaluated. This could be analyzed using the SEAS5 hindcasts and comparing the results to ERA5 data. A bias correction of the predictions could increase the skill because it leads to a better agreement of the predicted values with the reanalysis data (HERMANSON et al., 2017; LORENZ et al., 2020). The challenge here is, among other things, that the ensemble spread of the predictions may not allow unambiguous conclusions, since each ensemble member must be clustered individually. This also implies a large computational effort. Bias correction may help decisively against the ensemble spread. Against the high expenditure of computer capacity, one could concentrate on using only those ensemble members which are already quite close to the pressure situation in the lead month 0 instead of all the 51 ensemble members. If the analysis of the forecasts overall shows significant skill, the MAM season precipitation could be forecast better than previously possible, providing valuable information to the region's stakeholders.

5.2. Acknowledgements

This research was supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 869730 (CONFER).

V. Bibliography

- ADAMS, H.D., M. GUARDIOLA-CLARAMONTE, G.A. BARRON-GAFFORD, J. CAMILO VILLEGAS, D.D. BRESHEARS, C.B. ZOU, P.A. TROCH and T.E. HUXMAN (2009): Temperature sensitivity of drought-induced tree mortality portends increased regional die-off under global-change-type drought. In: Proceedings of the National Academy of Sciences of the United States of America 106 (17): 7063–7066. DOI: 10.1073/pnas.0901438106.
- BAHAGA T.K., G. MENGISTU TSIDU, F. KUCHARSKIB and G.T. DIRO (2015): Potential predictability of the sea-surface temperature forced equatorial East African short rains interannual variability in the 20th century. In: Quarterly Journal of the Royal Meteorological Society 141: 16–26. DOI: 10.1002/qj.2338.
- BAHAGA T.K., A. FINK and P. KNIPPERTZ (2019): Revisiting interannual to decadal teleconnections influencing seasonal rainfall in the Greater Horn of Africa during the 20th century. In: International Journal of the Royal Meteorological Society 39 (5): 2765–2785. DOI: 10.1002/joc.5986.
- BALSAMO, G., C. ALBERGEL, A. BELJAARS, S. BOUSSETTA, E. BRUN, H. CLOKE, D. DEE, E. DUTRA, J. MUÑOZ-SABATER, F. PAPPENBERGER, P. DE ROSNAY, T. STOCKDALE and F. VITART (2015): ERA-Interim/Land: a global land surface reanalysis data set. In: Hydrology and Earth System Sciences 19: 3 89–407. DOI: 10.5194/hess-19-389-2015.
- BAMSTON, A.G., M. CHELLIAH and S.B. GOLDENBERG (1997): Documentation of a highly ENSO-related SST region in the equatorial pacific: Research note. In: Atmosphere–Ocean 35 (3): 367–383. DOI: 10.1080/07055900.1997.9649597.
- BÁRDOSSY, A. (2010): Atmospheric circulation pattern classification for South-West Germany using hydrological variables. In: Physics and Chemistry of the Earth, Parts A/B/C 35 (9-12): 498–506. DOI: 10.1016/j.pce.2010.02.007.
- BEHERA, S., J.J. LUO, S. MASSON, S.A. RAO, H. SAKUMA and T. YAMAGATA (2006): A CGCM Study on the Interaction between IOD and ENSO. In: Journal of Climate 19 (9): 1688–1705. DOI: 10.1175/JCLI3797.1.
- Bureau of Meteorology (BOM) (2021): Madden-Julian Oscillation (MJO). Available from: <http://www.bom.gov.au/climate/mjo/graphics/rmm.74toRealtime.txt>.

- CAMBERLIN, P. and N. PHILIPPON (2002): The East African March–May Rainy Season: Associated Atmospheric Dynamics and Predictability over the 1968–97 Period. In: *Journal of Climate* 15 (9): 1002–1019. DOI: 10.1175/1520-0442(2002)015<1002:TEAMMR>2.0.CO;2.
- CAMBERLIN, P., V. MORON, R. OKOOLA, N. PHILIPPON and W. GITAU (2009): Components of rainy seasons variability in Equatorial East Africa: onset, cessation, rainfall frequency and intensity. In: *Theoretical and Applied Climatology* 98 (3). DOI: 10.1007/s00704-009- 0113-1.
- COOK, K.H. and E.K. VIZY (2013): Projected Changes in East African Rainy Seasons. In: *Journal of Climate* 26 (16): 5931–5948. DOI: 10.1175/JCLI-D-12-00455.1.
- CHAN, R.Y., M. VUILLE, D.R. HARDY and R.S. BRADLEY (2007): Intraseasonal precipitation variability on Kilimanjaro and the East African region and its relationship to the large-scale circulation. In: *Theoretical and Applied Climatology* 93: 149–165. DOI: 10.1007/s00704-007-0338-9.
- DAVIS, R.E. (1976): Predictability of Sea Surface Temperature and Sea Level Pressure Anomalies over the North Pacific Ocean. In: *Journal of Physical Oceanography* 6 (3): 249–266. DOI: 10.1175/1520-0485(1976)006<0249:POSSTA>2.0.CO;2.
- DIRO, G.T., E. BLACK and D.I.F. GRIMES (2008): Seasonal forecasting of Ethiopian spring rains. In: *Meteorological Applications* 15: 73–83. DOI: 10.1002/met.63.
- DONOHOE, A., J. MARSHALL, D. FERREIRA and D. MCGEE (2013): The Relationship between ITCZ Location and Cross-Equatorial Atmospheric Heat Transport: From the Seasonal Cycle to the Last Glacial Maximum. In: *Journal of Climate* 26 (11): 3597–3618. DOI: 10.1175/JCLI-D-12-00467.1.
- ENDRIS, H.S., C. LENNARD, B. HEWITSON, A. DOSIO, G. NIKULIN and G.A. ARTAN (2015): Teleconnection responses in multi-GCM driven CORDEX RCMs over Eastern Africa. In: *Climate Dynamics* 46: 2821–2846. DOI: 10.1007/s00382-015-2734-7.
- ENDRIS, H.S., C. LENNARD, B. HEWITSON, A. DOSIO, G. NIKULIN and G.A. ARTAN (2019): Future changes in rainfall associated with ENSO, IOD and changes in the mean state over Eastern Africa. In: *Climate Dynamics* 52: 2029–2053. DOI: 10.1007/s00382-018-4239-7.
- ENYEW B. and G. STEENEVELD (2014): Analysing the Impact of Topography on Precipitation and Flooding on the Ethiopian Highlands. In: *Journal of Geology and Geosciences* 2014, 3 (173). DOI: 10.4172/2329-6755.1000173.
- FISHER, R.A. (1922): On the Mathematical Foundations of Theoretical Statistics. In: *The Royal Society* 222: 594–604. DOI: 10.1098/rsta.1922.0009.

- FUNK, C. and Co-Authors (2018): Examining the potential contributions of extreme “Western V” sea surface temperatures to the 2017 March–June East African drought. In: HERRING S.C., N. CHRISTIDIS, A. HOELL, M.P. HOERLING and P.A. STOTT (Ed.)(2018): Explaining extreme events of 2017 from a climate perspective. Special Supplement to the Bulletin of the American Meteorological Society Vol. 100, No. 1, January 2019. DOI: 10.1175/BAMS-D-18-0108.1.
- GEBREMICHAEL, A., S. QURAISHI and G. MAMO (2014): Analysis of Seasonal Rainfall Variability for Agricultural Water Resource Management in Southern Region, Ethiopia. In: Journal of Natural Sciences Research 4 (11): 56–79.
- GISSILA, T., E. BLACK, D.I.F. GRIMES and J.M. SLINGO (2004): Seasonal forecasting of the Ethiopian summer rains. In: International Journal of Climatology 24 (11): 1345–1358. DOI: 10.1002/joc.1078.
- GLEIXNER, S., N.S. KEENLYSIDE, T.D. DEMISSIE, F. COUNILLON, Y. WANG and E. VISTE (2017): Seasonal predictability of Kiremt rainfall in coupled general circulation models. In: Environmental Research Letters 12: 114016. DOI: 10.1088/1748-9326/aa8cfa.
- GLEIXNER, S., T. DEMISSIE and G. DIRO (2020): Did ERA5 Improve Temperature and Precipitation Reanalysis over East Africa? Atmosphere 2020, 11, 996; DOI: 10.3390/atmos11090996.
- GUDOSHAVA, M. and Co-Authors (2020): Projected effects of 1.5 °C and 2 °C global warming levels on the intra-seasonal rainfall characteristics over the Greater Horn of Africa. In: Environmental Research Letters 15 (3). DOI: 10.1088/1748-9326/ab6b33.
- HAMED, K. and A. RAO (1998): A modified Mann-Kendall trend test for autocorrelated data. In: Journal of Hydrology 204: 182–196. DOI: 10.1016/S0022-1694(97)00125-X.
- HASTENRATH, S., D. POLZIN and P. CAMBERLIN (2004): Exploring the predictability of the ‘Short Rains’ at the coast of East Africa. In: International Journal of Climatology 24 (11): 1333–1343. DOI: 10.1002/joc.1070.
- HERMANSON, L., H.-L. REN, M. VELLINGA, N.D. DUNSTONE, P. HYDER, S. INESON, A.A. SCAIFE, D.M. SMITH, V. THOMPSON, B. TIAN and K.D. WILLIAMS (2017): Different types of drifts in two seasonal forecast systems and their dependence on ENSO. Climate Dynamics 51: 1411–1426. DOI: 10.1007/s00382-017-3962-9.
- HERSBACH, H. and Co-Authors (2020): The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society 146 (730): 1999–2049. DOI: 10.1002/qj.3803.

HOFFMANN, P. and H. SCHLÜNZEN (2013): Weather pattern classification to represent the urban heat Island in present and future climate. In: *Journal of Applied Meteorology and Climatology* 52 (12), 2699–2714. DOI: 10.1175/JAMC-D-12-065.1.

HUBERT, L. and P. ARABIE (1985): Comparing Partitions. In: *Journal of Classification* 2: 193–218. DOI: 10.1007/bf01908075.

JORBA, O., C. PÉRÉZ, F. ROCADENBOSCH and J.M. BALDASANO (2004): Cluster Analysis of 4-Day Back Trajectories Arriving in the Barcelona Area, Spain, from 1997 to 2002. In: *Journal of Applied Meteorology and Climatology* 43 (6): 887–901. DOI: 10.1175/1520-0450(2004)043<0887:CAODBT>2.0.CO;2

KASPAR F., U. SCHULZWEIDA and R. MÜLLER (2010): “Climate Data Operators” as a user-friendly processing tool for CMSAF's satellite-derived climate monitoring products. Available from https://code.mpimet.mpg.de/attachments/download/1095/pdf_conf_p57_s5_02_kaspar_v.pdf.

KILAVI, M., D. MACLEOD, M. AMBANI, J. ROBBINS, R. DANKERS, RICHARD GRAHAM, H. TITLEY, A.A.M. SALIH and M.C. TODD (2018): Extreme Rainfall and Flooding over Central Kenya Including Nairobi City during the Long-Rains Season 2018: Causes, Predictability, and Potential for Early Warning and Actions. In: *Atmosphere* 9 (12). DOI: 10.3390/atmos9120472.

KORECHA, D. and A.G. BARNSTON (2007): Predictability of June–September Rainfall in Ethiopia. In: *Monthly Weather Review* 135 (2): 628–650. DOI: 10.1175/MWR3304.1.

LAUX, P., B. BÖKER, E. SÁVIO MARTINS, F. DAS CHAGAS VASCONCELOS JUNIOR, V. MORON, T.C. PORTELE, C. LORENZ, A. PHILIPP and H. KUNSTMANN (2020): A semi-objective circulation pattern classification scheme for the semi-arid Northeast Brazil. In: *International Journal of Climatology* 41: 51–72. DOI: 10.1002/joc.6608.

LORENZ, C., T.C. PORTELE, P. LAUX and H. KUNSTMANN (2020): Bias-corrected and spatially disaggregated seasonal forecasts: A long-term reference forecast product for the water sector in semi-arid regions. In: *Earth System Science Data* 13 (6), 2701–2722. DOI: 10.5194/essd-13-2701-2021.

LYON, B. and D.G. DEWITT (2012): A recent and abrupt decline in the East African long rains. *Geophysical Research Letters* 39: L02702. DOI: 10.1029/2011GL050337.

MACLEOD, D. (2019): Seasonal forecasts of the East African long rains: insight from atmospheric relaxation experiments. In: *Climate Dynamics* 53: 4505–4520. DOI: 10.1007/s00382-019-04800-6.

- MANATSA, D., Y. MORIOKA, S.K. BEHERA, C.H. MATARIRA and T. YAMAGATA (2013): Impact of Mascarene High variability on the East African ‘short rains’. In: Climate Dynamics 42: 1259–1274. DOI: 10.1007/s00382-013-1848-z.
- MANN, H.B. and D.R. WHITNEY (1947): On a test of whether one of two random variables is stochastically larger than the other. In: Annals of Mathematical Statistics 18: 50–60.
- MCKEE, T.B., N.J. DOESKEN and J. KLEIST (1993): The relationship of drought frequency and duration to time scales. Eighth Conference on Applied Climatology, 17-22 January 1993, Anaheim, California.
- MEYNARD, C.N., M. LECOQ, M.-P. CHAPUIS and CYRIL PIOU (2020): On the relative role of climate change and management in the current desert locust outbreak in East Africa. In: Global Change Biology 26: 3753–3755. DOI: 10.1111/gcb.15137.
- MILTON, R. (1964): An Extended Table of Critical Values for the Mann-Whitney (Wilcoxon) Two-Sample Statistic. In: Journal of the American Statistical Association 59 (307): 925–934. DOI: 10.1080/01621459.1964.10480740.
- MUTAI, C.C., M.N. WARD and A.W. COLMAN (1998): Towards the prediction of the East Africa short rains based on sea-surface temperature–atmosphere coupling. In: International Journal of Climatology 18 (9): 975–997. DOI: 10.1002/(SICI)1097-0088(199807)18:9<975::AID-JOC259>3.0.CO;2-U.
- NACHAR, N. (2008): The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. In: Tutorials in Quantitative Methods for Psychology 4 (1): 13–20. DOI: 10.20982/tqmp.04.1.p013.
- NOAA (2021a): Download Climate Timeseries. Dipole Mode Index (DMI). Available from https://psl.noaa.gov/gcos_wgsp/Timeseries/DMI/.
- NOAA (2021b): ESRL/NOAA Nino 3.4, 1870-present based on HadISST. Available from https://psl.noaa.gov/gcos_wgsp/Timeseries/Data/nino34.long.data.
- OKOOLA, R.E. (1999): A diagnostic study of the eastern Africa monsoon circulation during the northern hemisphere spring season. In: International Journal of Climatology 19 (2): 143–168. DOI: 10.1002/(SICI)1097-0088(199902)19:2<143::AID-JOC342>3.0.CO;2-U.
- OMONDI P.A.O., J.L. AWANGE, E. FOROOTAN, L.A. OGALLO, R. BARAKIZA, G.B. GIRMAW, I. FESSEHA, V. KULULETERA, C. KILEMBE and M.M. MBATI (2014): Changes in temperature and precipitation extremes over the Greater Horn of Africa region from 1961 to 2010. In: International Journal of Climatology 34: 1262–1277. DOI: 10.1002/joc.3763.

PAI, D.S. and M. RAJEEVAN (2005): Empirical prediction of Indian summer monsoon rainfall with different lead periods based on global SST anomalies. In: Meteorology and Atmospheric Physics 000: 1–11. DOI: 10.1007/s00703-005-0136-9.

PHILIPP, A., J. BARTHOLY, C. BECK, M. ERPICUM, P. ESTEBAN, X. FETTWEIS, R. HUTH, P. JAMES, S. JOURDAN, F. KREIENKAMP, T. KRENNERT, S. LYKOUDIS, S.C. MICHALIDES, K. PIANKO-KLUCZYNSKA, P. POST, D. RASILLA ÁLVAREZ, R. SCHIEMANN, A. SPEKAT and F.S. TYMVIOSK (2010): Cost733cat – A database of weather and circulation type classifications. In: Physics and Chemistry of the Earth, Parts A/B/C 35 (9–12): 360–373. DOI: 10.1016/j.pce.2009.12.010.

PHILIPPON, N., P. CAMBERLIN and N. FAUCHEREAU (2002): Empirical predictability study of October–December East African rainfall. In: Quarterly Journal of the Royal Meteorological Society 128 (585): 2239–2256. DOI: 10.1256/qj.01.190.

PHILIPP, A., C. BECK, P. ESTEBAN, F. KREIENKAMP, T. KRENNERT, K. LOCHBIHLER, S. P. LYKOUDIS, K. PIANKO-KLUCZYNSKA, P. POST, D. R. ÁLVAREZ, A. SPEKAT and F. STREICHER (2014): Cost733class-1.2 User guide. Available from https://www.researchgate.net/publication/269335894_COST733_CLASS_v12_User_guide.

POHL, B. and P. CAMBERLIN (2006): Influence of the Madden–Julian Oscillation on East African rainfall: II. March–May season extremes and interannual variability. In: Quarterly Journal of the Royal Meteorological Society 132: 2541–2558. DOI: 10.1256/qj.05.223.

RAND, W.M. (1971): Objective Criteria for the Evaluation of Clustering Methods. In: Journal of the American Statistical Association 66 (336): 846–850. DOI: 10.1080/01621459.1971.10482356.

ROWELL, D.R., B.B.B. BOOTH, S.E. NICHOLSON and P. GOOD (2015): Reconciling Past and Future Rainfall Trends over East Africa. In: Journal of Climate 28 (24): 9768–9788. DOI: 10.1175/JCLI-D-15-0140.1.

SANTOS, J.M. and M. EMBRECHTS (2009): On the Use of the Adjusted Rand Index as a Metric for Evaluating Supervised Classification. Lecture Notes in Computer Science 5769: 175–184. DOI: 10.1007/978-3-642-04277-5_18.

SEELY, M.K. and J.R. HENSCHEL (1998): The Climatology of Namib Fog. Available from: https://www.researchgate.net/profile/Joh-Henschel/publication/289525691_The_climatology_of_Namib_fog/links/568e7afb08aeaa1481b01e85/The-climatology-of-Namib-fog.pdf.

TAREK, M., F.P. BRISSETTE and R. ARSENAULT (2020): Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences* 24: 2527–2544. DOI: 10.5194/hess-24-2527-2020.

UMMENHOFER, C.C., A. SEN GUPTA1, M.H. ENGLAND and C.J.C. REASON (2009): Contributions of Indian Ocean Sea Surface Temperatures to Enhanced East African Rainfall. In: *Journal of Climate* 22 (4): 993–1013. DOI: 10.1175/2008JCLI2493.1.

UPPADA, S.K. (2014): Centroid Based Clustering Algorithms- A Clarion Study. In: *International Journal of Computer Science and Information Technologies* 5 (6): 7309–7313. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.658.3904&rep=rep1&type=pdf>.

VICENTE-SERRANO, S.M., S. BEGUERÍA, and J.I. LÓPEZ-MORENO (2010): A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. In: *Journal of Climate* 23 (7): 1696–1718. DOI: 10.1175/2009JCLI2909.1.

VIGAUD, N., B. LYON and A. GIANNINI (2017): Sub-seasonal teleconnections between convection over the Indian Ocean, the East African long rains and tropical Pacific surface temperatures. In: *International Journal of Climatology* 37 (3): 1167–1180. DOI: 10.1002/joc.4765.

WALKER, D.P., C.E. BIRCH, J.H. MARSHAM, A.A. SCAIFE, R.J. GRAHAM and Z.T. SEGELE (2019): Skill of dynamical and GHACOF consensus seasonal forecasts of East African rainfall. In: *Climate Dynamics* 53 (7–8). DOI: 10.1007/s00382-019-04835-9.

WALLACE, J.M. and D.S. GUTZLER (1981): Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter. In: *Monthly Weather Review* 109 (4): 784–812. DOI: 10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2.

WANG, F., W. SHAO, H. YU, G. KAN, X. HE, D. ZHANG, M. REN and G. WANG (2020): Re-evaluation of the Power of the Mann-Kendall Test for Detecting Monotonic Trends in Hydrometeorological Time Series. In: *Frontiers in Earth Science* 8 (14). DOI: 0.3389/feart.2020.00014.

WOLTER, K. and M.S. TIMLIN (2011): El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index. In: *International Journal of Climatology* 31 (7): 1074–1087. DOI: 10.1002/joc.2336.

World Meteorological Organization (WMO) (2020): State of the Climate in Africa. Available from: https://library.wmo.int/doc_num.php?explnum_id=10421.

YANG, W., R. SEAGER, M.A. CANE and B. LYON (2015): The annual cycle of East African precipitation. In:
Journal of Climate 28 (6): 2385–2404. DOI: 10.1175/JCLI-D-14-00484.1.

VI. Appendix

Table A1: Occurrence of clusters by month. Mean Day 0 is March 1st, Mean Day 1 is March 2nd, etc.

CLUSTER	MARCH	APRIL	MAY	MEAN DAY
SST-1	0	0	439	79,8
SST-2	0	90	382	67,2
SST-3	0	279	148	57,8
SST-4	375	147	0	25,3
SST-5	0	0	231	85,7
SST-6	14	385	9	44,2
SST-7	435	4	0	13,2
SST-8	280	0	0	6,5
SST-9	105	265	0	35,5
WD-1	7	181	192	59,3
WD-2	276	85	8	22,0
WD-3	258	139	0	24,3
WD-4	168	228	24	35,1
WD-5	253	16	0	12,2
WD-6	1	22	523	80,0
WD-7	1	193	439	67,2
WD-8	172	10	0	13,3
WD-9	73	296	23	41,7
TP-1	41	134	65	49,8
TP-2	6	42	506	76,8
TP-3	95	110	18	34,8
TP-4	1	86	79	60,8
TP-5	0	73	149	65,8
TP-6	427	179	124	32,0
TP-7	40	242	112	50,8
TP-8	201	73	2	21,6
TP-9	278	84	1	19,4
TP-10	2	26	151	71,0
TP-11	118	121	2	30,0

Table A2–4: Mean daily precipitation per Cluster for all countries.

<i>tp</i>	<i>BDI</i>	<i>DJI</i>	<i>ERI</i>	<i>ETH</i>	<i>KEN</i>	<i>RWA</i>	<i>SOM</i>	<i>TZA</i>	<i>SSD</i>	<i>SDN</i>	<i>UGA</i>
1	5,1	2,9	2,3	6,6	2,9	4,7	3,1	3,3	2,7	0,2	7,1
2	1,3	0,3	0,3	3,5	1,5	1,8	1,0	1,1	3,8	1,6	4,7
3	6,0	0,5	0,5	2,4	2,0	4,4	0,9	4,8	1,7	0,3	4,5
4	6,1	1,4	1,2	5,5	5,8	6,1	4,1	5,5	4,2	0,9	8,5
5	4,9	1,2	0,9	6,4	6,2	6,1	3,3	2,9	6,4	1,6	10,0
6	3,0	0,2	0,2	1,3	0,7	1,9	0,3	2,3	0,6	0,2	1,9
7	7,1	0,3	0,3	3,4	3,9	6,0	2,1	4,2	3,7	0,6	7,9
8	5,0	0,3	0,4	1,6	1,5	3,4	0,4	5,4	1,3	0,2	3,7
9	7,2	0,5	0,4	1,6	2,3	5,3	0,4	8,4	1,1	0,0	5,5
10	3,7	0,5	0,4	4,2	3,5	4,0	2,9	2,2	3,8	1,4	6,6
11	9,5	1,5	1,7	4,8	8,5	8,8	2,1	11,0	3,3	0,2	10,8

<i>SST</i>	<i>BDI</i>	<i>DJI</i>	<i>ERI</i>	<i>ETH</i>	<i>KEN</i>	<i>RWA</i>	<i>SOM</i>	<i>TZA</i>	<i>SSD</i>	<i>SDN</i>	<i>UGA</i>
1	1,6	0,6	0,4	3,9	2,1	2,3	1,5	1,4	3,7	1,6	5,4
2	4,4	0,7	0,6	4,4	3,5	4,7	2,6	2,7	4,3	1,1	7,1
3	5,5	0,8	0,7	4,8	4,4	5,4	3,3	3,5	3,5	0,8	7,2
4	6,0	1,0	0,9	2,6	2,8	4,4	0,7	6,0	1,6	0,1	5,0
5	0,6	0,3	0,2	2,9	1,2	0,8	0,5	0,6	3,3	1,6	3,8
6	7,0	1,0	0,9	4,2	4,2	5,8	2,0	5,4	3,0	0,4	7,2
7	5,7	0,7	0,6	1,8	1,9	4,3	0,4	5,7	1,3	0,1	4,5
8	4,7	0,3	0,2	0,9	1,0	3,3	0,1	5,1	0,7	0,1	3,1
9	6,0	0,6	0,8	2,8	3,4	4,7	1,3	5,5	1,9	0,1	5,7

<i>wd900</i>	<i>BDI</i>	<i>DJI</i>	<i>ERI</i>	<i>ETH</i>	<i>KEN</i>	<i>RWA</i>	<i>SOM</i>	<i>TZA</i>	<i>SSD</i>	<i>SDN</i>	<i>UGA</i>
1	6,0	0,8	0,9	5,0	5,2	6,6	2,9	4,1	4,8	1,0	8,5
2	5,1	0,5	0,4	1,7	1,6	3,7	0,5	4,2	1,0	0,1	3,6
3	5,8	0,4	0,4	1,8	2,5	4,3	0,6	5,5	1,3	0,1	4,5
4	6,9	1,7	1,6	4,4	3,8	5,8	1,7	6,3	2,9	0,3	7,5
5	5,0	0,4	0,3	1,2	1,2	3,1	0,2	5,3	0,6	0,1	2,9
6	1,5	0,7	0,5	4,0	2,1	2,0	1,2	1,2	3,6	1,6	5,2
7	3,2	0,4	0,4	3,7	2,4	3,3	2,3	2,2	3,2	1,0	5,5
8	6,5	0,7	0,7	2,2	3,8	5,2	0,5	8,2	1,6	0,2	5,9
9	6,5	0,5	0,4	3,3	3,9	5,2	2,0	4,9	2,5	0,3	6,5

Table A5–7: Mean daily temperature at 2 m height above ground level per Cluster for all countries.

<i>tp</i>	<i>BDI</i>	<i>DJI</i>	<i>ERI</i>	<i>ETH</i>	<i>KEN</i>	<i>RWA</i>	<i>SOM</i>	<i>TZA</i>	<i>SSD</i>	<i>SDN</i>	<i>UGA</i>
1	19,9	27,9	28,8	23,8	25,7	20,2	27,3	22,5	29,5	31,3	23,0
2	20,0	31,2	31,2	24,2	25,2	20,3	27,6	22,1	28,3	31,7	22,8
3	20,0	27,0	28,6	24,7	26,2	20,3	27,4	22,8	30,4	31,6	23,8
4	19,7	28,8	30,1	23,9	25,2	20,0	27,0	22,1	29,0	32,2	22,6
5	19,7	29,4	30,3	23,7	25,0	20,0	26,9	22,2	28,1	31,7	22,4
6	20,4	27,4	28,8	25,0	26,5	20,8	27,3	23,0	31,1	31,6	24,4
7	19,7	28,6	29,8	24,4	25,5	20,0	27,3	22,3	29,4	32,1	22,9
8	20,2	26,5	28,0	25,0	26,8	20,6	27,3	23,1	30,9	30,8	24,3
9	19,9	26,2	27,4	24,7	26,4	20,3	27,1	22,6	30,4	29,9	23,7
10	19,7	29,9	30,8	24,0	25,0	19,9	27,0	22,0	28,5	32,0	22,5
11	19,5	26,8	27,7	23,9	25,7	19,9	27,3	22,3	29,2	30,3	22,7

<i>SST</i>	<i>BDI</i>	<i>DJI</i>	<i>ERI</i>	<i>ETH</i>	<i>KEN</i>	<i>RWA</i>	<i>SOM</i>	<i>TZA</i>	<i>SSD</i>	<i>SDN</i>	<i>UGA</i>
1	20,0	31,0	31,2	24,0	25,2	20,2	27,4	22,0	28,2	31,6	22,6
2	19,7	29,7	30,5	23,9	25,1	19,9	27,1	22,1	28,5	32,2	22,6
3	19,9	29,0	30,4	24,2	25,4	20,1	27,2	22,5	29,4	32,6	22,9
4	20,1	26,6	28,0	24,8	26,6	20,4	27,5	22,9	30,7	30,8	23,9
5	20,2	32,6	31,5	24,2	24,9	20,6	27,8	21,8	28,1	31,3	22,7
6	19,5	27,7	29,1	24,1	25,4	19,8	27,4	22,2	29,6	31,9	22,9
7	20,1	25,8	27,1	24,8	26,7	20,5	27,0	23,0	30,8	30,1	24,3
8	20,2	24,7	26,7	24,7	26,9	20,7	26,2	23,1	31,2	29,7	24,6
9	20,3	27,7	29,0	25,0	26,3	20,6	27,9	22,9	30,9	31,9	23,8

<i>wd900</i>	<i>BDI</i>	<i>DJI</i>	<i>ERI</i>	<i>ETH</i>	<i>KEN</i>	<i>RWA</i>	<i>SOM</i>	<i>TZA</i>	<i>SSD</i>	<i>SDN</i>	<i>UGA</i>
1	19,6	29,1	30,2	24,0	25,1	19,9	27,1	22,1	28,7	32,1	22,5
2	20,4	26,3	27,7	24,9	26,8	20,8	27,2	23,2	31,1	30,7	24,6
3	19,9	26,3	28,0	24,7	26,4	20,3	27,1	22,7	30,6	30,8	23,9
4	20,0	27,2	28,6	24,6	26,3	20,3	27,4	22,7	30,4	31,4	23,6
5	20,4	25,5	27,1	24,8	26,8	20,9	26,9	23,2	31,2	30,1	24,7
6	20,0	31,5	31,0	24,0	25,1	20,3	27,6	22,0	28,0	31,2	22,6
7	19,9	29,9	30,8	24,3	25,3	20,1	27,3	22,2	29,1	32,5	22,9
8	19,9	25,7	27,0	24,4	26,5	20,3	26,9	22,6	30,3	29,6	23,6
9	19,7	27,8	29,3	24,5	25,7	20,0	27,4	22,4	29,9	32,0	23,1

Figure A1: The temperature trend during the season over the ocean and the land. As temperature data, grid-points in a strip from 4° N to 1° S and 42° E to 51° E were assigned to either the ocean or the land.

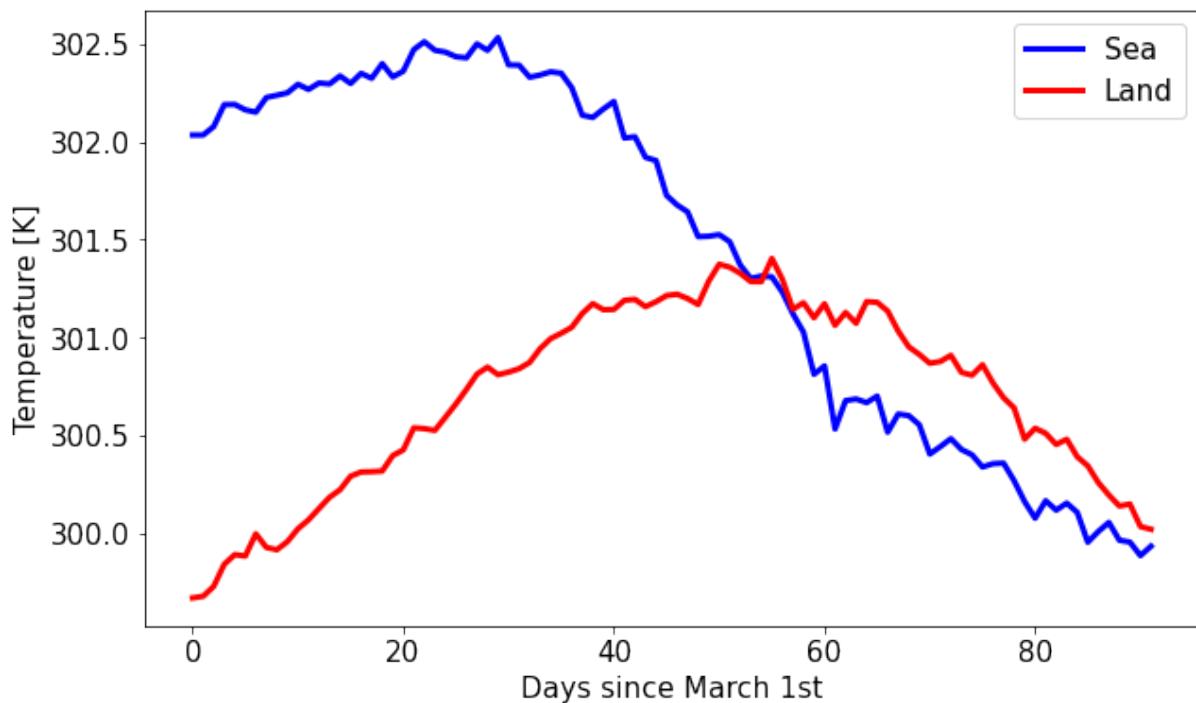


Figure A2: Frequency increase in each cluster [%]. Calculation by the end point through the start point of a trend line laid through the years (using the method of least squares) minus 1.

