Seasonal Forecasting of Rainfall in Equatorial East Africa using an artificial neural network

Colin J. McMechan

Jurian Beunk

Utrecht University

Seasonal Forecasting of Rainfall in Equatorial East Africa using an artificial neural network

MSc Thesis
02-2021

Author: Jurian Beunk
Student number: 5745861
E-mail: j.beunk@students.uu.nl
1st supervisor: Geert Sterk

2nd supervisor: Niko Wanders
ECTS Credits: 30

MSc Programme: Earth Surface and Water
Faculty of Geosciences
Department of Physical Geography
Utrecht University

Abstract

The eastern African region continues to be extremely vulnerable to droughts. In Kenya alone, about 50 million people have been affected by droughts since 1983. Rainfall in the region is bi-modal, with the long rains in the boreal spring and the short rains in the boreal autumn. Seasonal forecasts of rainfall are essential to improve drought preparedness and mitigation for water managers and farmers. The aim of this study was to forecast monthly rainfall totals at a set of locations in the equatorial east African region, using an Artificial Neural Network (ANN) called a Long Short-Term Memory (LSTM). The input dataset consisted of seasonal forecasts (SEAS5) of precipitation and temperature from the European Centre for Medium-Range Weather Forecasts (ECMWF) in combination with several climate indices related to the El Niño Southern Oscillation (ENSO), the Indian Ocean Dipole (IOD) and the Madden-Julian Oscillation (MJO). The performance was assessed by means of the WNDI (Weighted Non-Dimensional Index) and the anomaly correlation. Benchmarks for the evaluation included: climatology, precipitation forecasts by SEAS5 and a multi-variate linear model based on the same input data as the LSTM. In addition, the ability of the model to forecasts anomalous seasonal rainfall was assessed based on the 2x2 contingency table and associated scores (hit rate, false alarm rate and Clayton Skill Score). Sensitivity analysis was carried out to evaluate the relative importance of the features in the input dataset.

For all lead times, the LSTM outperformed both the linear model and SEAS5 in terms of both anomaly correlation and WNDI, although it is not clear whether or not this is statistically significant. Especially at longer lead times the LSTM shows improved performance relative to SEAS5, due to a good coupling with the climate indices. At a lead time of 4 months and in terms of WNDI and the anomaly correlation, differences in performance between the long rains and the short rains were small. The LSTM model showed slightly better performance with a WNDI of 0.75-0.85 (equal to an RMSE of 6-7% of the mean annual rainfall) and an anomaly correlation of 0.76-0.77, compared to the linear model. In the context of anomalous seasonal rainfall, the performance in the short rains was substantially better, on average, with hit rates between 40-50%, false alarm rates of 4% and Clayton Skill scores of roughly 0.7. The linear model performed better at forecasting below-normal rainfall in the short rains. The LSTM showed better performance in the long rains, especially when forecasting below-normal rainfall, with a hit rate of 42%, false alarm rate of 8% and a Clayton Skill Score of 0.55, on average. With regard to operational use, especially forecasts of anomalous rainfall are of interest, as they may be associated with the occurrence of floods and droughts. The model developed in this study is underconfident in forecasting these anomalies and are therefore not sophisticated enough for operational use. However, due to the observed low false alarm rates, it may still provide valuable information to farmers and water managers near the stations. Several suggestions are made for the improvement of the model developed in this study.

Acknowledgements

The idea for this thesis was conceived after a good talk with Geert Sterk. It was my ambition to explore some deep learning techniques, since I wanted to learn something new and explore a more practical side of the earth sciences: forecasting. Since Geert is an East African expert familiar with the potential detrimental impacts of droughts, we figured it would be a nice idea to look at the possibility of forecasting rainfall in East Africa. Together with Niko Wanders, an expert in the field of drought, climate models and machine learning, a team was formed. I would like to thank both Geert and Niko for their support and feedback during the process of this research. Also, I would like to thank Sharon Nicholson for helping me out tremendously by sharing her personally compiled rainfall dataset, forming the basis of this research.

For their support and love, I would like to thank my family, friends and girlfriend.

Table of contents

Introduction 1
Background 1
Problem definition 1
Objectives and Research Questions 3
Study Area 4
Topography & rainfall 5
Coastal climate 5
Methods 6
Data 6
Station data 6
Seasonal forecast data 6
Large scale climate indices 6
The modelling process 7
RNNs, memory cells & LSTM 8
Pre-processing the data 10
Training the model 10
Forecast verification 12
Sensitivity analysis 15
Repeated experiments 15
Robustness experiment 15
Feature importance 16
Software 16
Results 17
Literature study 17
Defining drought 17
Rainfall variability in equatorial east Africa 18
Predictability and forecasting 20
Seasonal forecasting by dynamical models 22
Machine learning 23
Forecast overview 24
Forecast verification 27
Anomalous seasonal rainfall (4-month lead time) 27

Seasonal differences (4-month lead time) 28
Lead time and performance 29
Sensitivity analysis 29
Discussion 31
Methodology and data 31
Data quantity and quality 31
Seasonality, long-term trends and nonstationary relationships 31
Direct Multi-Step Forecasting 32
Sensitivity analysis 33
Performance evaluation 33
Lead time and performance 33
Differences between the seasons 33
Value of the forecasts for operational use 34
Recommendations 35
Conclusion 36
References 37
Appendix A: Basic statistics at all rainfall stations 40
Appendix B: Overview climate indices 41
Appendix C: Rainfall records at additional stations 42
Appendix D: Mean rainfall distributions at additional stations 44
Appendix E: Forecasts at additional stations 45
Appendix F: Seasonal rainfall totals at additional stations 48
Appendix G: 2x2 contingency tables and scores for individual stations 54
Appendix H: performance at individual stations 57
Appendix I: model calibration and robustness 58

List of figures

Figure 1 Digital Elevation Model (GTOPO30) of the study area (USGS, 1993) 4
Figure 2 Mean Annual Rainfall in EEA. Adopted from: Woodman (2020). 4
Figure 3 Rainfall stations in the research area. 5
Figure 4 Mean vector wind and divergence at 850 mb during Nov-Mar (a) and May-Sep (b). Adopted from: Nicholson (2016). 5
Figure 5 Schematic overview of the workflow in this study. 6
Figure 6 a) linear threshold unit (LTU) b) simple recurrent neural network c) schematic (memory) cell (left), through time (right)(d) layer of neurons (left) through time (right) (modified after: Géron, 2017) 8
Figure 7 Architecture of an LSTM cell. Adopted from: Géron (2017). 9
Figure 8 Schematic overview of the single- step, lead time=1 model. 11
Figure 9 The 2x2 contingency table comparing observed and forecast frequencies of occurrence and non-occurrence of a dichotomous event (adapted from: Wilks, 2006). 13
Figure 10 The defined categories for below normal conditions (to the left of the lower boundary), normal conditions (between the lower and upper boundary) and above-normal conditions (to the right of the upper boundary), based on the normal distribution. Boundaries are located at +/- 0.8 standard deviations from the mean. 14
Figure 11 Schematic overview of the robustness experiment. 15
Figure 12 Drought definition framework. Adopted from: Van Loon et al. (2016). 17
Figure 13 Schematic representation of Local Geographic Factors, Remote Forcing, Regional Circulation and Coastal Influences affecting the rainy seasons in eastern Africa. Adopted from: Nicholson (2017). 18
Figure 14 Meteorological mechanisms affecting East African rainfall, associated with the MJO. MSE refers to the Moist Static Energy and SLP to Sea Level Pressure (modified from Berhane & Zaitchik, 2014). 19
Figure 15 Sliding 20-year correlation between October and November rainfall and the indices for ENSO and IOD. Vertical lines indicate regime shifts. Values above the striped line are significant (alpha = 0.05). Adopted from: Nicholson (2014). 22
Figure 16 The locations of the 7 best performing stations (x-axis = longitude, y-axis = latitude). 24
Figure 17 Observed monthly rainfall at station 9 in the period between 1994 and 2011. 24
Figure 18 Mean monthly rainfall at station 9 between 1994 and 2011. 25
Figure 19 Forecasts of monthly rainfall by SEAS5 (bottom), the linear model (middle) and the LSTM (top) for station 9 at a lead time of 4 months. 25
Figure 20 Long rains total rainfall at station 9, sorted by actual rainfall. 26
Figure 21 Short rains total rainfall at station 9, sorted by actual rainfall. 26
Figure 22 Median performance comparison between LSTM, Linear and SEAS5 in the long rains and short rains based on the Anomaly Correlation and WNDI at a lead time of 4 months 28
Figure 23 Performance as a function of lead time based on the top 7 stations. 29
Figure 24 The number of stations at each lead time where a feature is increasing performance. 30
Figure 25 Count of lead times in which ENSO/ IOD or MJO was improving model performance (note: the count lead time does not correspond to the actual lead time). 30
Figure 26 Geographical distribution of the top 7 stations (at a lead time of 4 months). 32

Figure 27 Drying and wetting trends in Equatorial East Africa (SPI06, 75% confidence level). Adopted from: Woodman (2020). 32

List of tables

Table 1 Historical drought years between 1980-2020 and number of affected people sorted by country (source: EM-DAT, 2020) 1
Table 2 Cross-correlation in a 12-month window between rainfall, DMI and MEI expressed by the r (modified after: Woodman, 2020) 7
Table 3 Conversion table of statistical and neural network jargon (Sarle, 1994). 10
Table 4 Location and elevation of the top 7 stations. 24
Table 5 Mean scores (Bias, Hit Rate, False Alarm Rate (F) and CSS skill score) over the best 7 stations for below-normal and above-normal conditions. 28

Introduction

Background

The natural disaster of drought can be characterized by its severity, duration and spatial extent. Due to these characteristics, drought is ranked the number one natural hazard in terms of impacts on 1) loss of life 2) economy 3) social impacts and 4) long-term impacts (Sivakumar *et al.,* 2014). Due to Africa's large dependence on rainfed agriculture and growing population, vulnerability to this natural hazard is high (Van Loon *et al.,* 2016; Mondiale, 2008). In Africa, about 95% of agriculture is rainfed (Van Aalst *et al.,* 2007) and about 70% of the populations livelihood is characterized by dependence on uncertain rainfall and exposure to climate risk (Hansen *et al.,* 2011). Economically, the agricultural sector in east Africa makes up for about 40% of the GDP (FAO, 2014). The population of east Africa has doubled between 1983-2008, whereas per-capita agricultural land has decreased by 33% (Funk *et al.,* 2008). Soil moisture drought (or agricultural drought) may result in reduced income for farmers, hunger, increased food prices, unemployment and migration (Sivakumar *et al.,* 2014). On the seasonal timescale, rainfall totals below 500 mm are a rough indicator for inadequate water availability for agriculture (Sheffield *et al.,* 2014). Table 1 shows a historical record of drought occurrence in the study area between 1980-2020 and the number of affected people. Kenya suffered from the most droughts, followed by Tanzania. The EM-DAT database was consulted but did not indicate economic damage since information on this topic is scarce. Due to this large vulnerability to droughts, seasonal forecasts can help to improve drought preparedness and mitigation for water managers and smallholder farmers (Van Loon, 2015).

Table 1 Historical drought years between 1980-2020 and number of affected people sorted by country (source: EM-DAT, 2020)

Country	Drought years	People affected (*10^6)
Burundi	1999, 2003, 2005, 2008, 2009, 2010	3.1
Kenya	1983, 1991, 1994, 1996, 1999, 2004, 2005, 2008, 2010, 2012, 2014, 2016, 2019	50.2
Tanzania	1984, 1988, 1990, 1996, 2003, 2004, 2006, 2011	12.7
Uganda	1987, 1998, 1999, 2002, 2005, 2008, 2010	5.0
Rwanda	1984, 1989, 1996, 1999, 2003	4.2

Problem definition

Rainfall in equatorial eastern Africa is bimodal and concentrated in two rainy seasons: the "long rains" in the months March, April and May and the "short rains" in the months October, November and December. The main source of variability throughout the year is associated with the short rainy season even though rainfall is generally lower during this season (Camberlin & Phillipon, 2002). In the long rains, sources of rainfall variability have proven difficult to pin down and interannual variability of rainfall in individual months of the long rains are uncorrelated. Nicholson and Entekhabi (1986) found that the opposite is true for the short rains. This is largely due to the strength and persistence of the El Niño Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD) during the short rains and is associated with above-average rainfall (Ogallo et al.1988, Kiladis & Diaz, 1989; Woodman, 2020; Keijzer, 2020). In the long rains, Zorita & Tilya (2002) found that March and April rainfall is linked to surface zonal winds and the zonal contrast between the Indian Ocean and the East African landmass. For May they found a strong link between the Indian monsoon, characterized by the meridional temperature contrast between the Asian continent and the Indian Ocean and meridional surface winds. The maximum possible lead time for prediction of the long rains is about 2 months due to the spring predictability barrier. The spring predictability barrier is a result of the fast changing state of the

tropical pacific ocean during spring, compromising the predictability of ENSO based on pre-spring variables (Nicholson, 2017). In addition to interannual variability, Pohl & Camberlin (2006) found a strong influence of the Madden-Julian Oscillation (MJO) on long rain intraseasonal variability (44% of long rains seasonal rainfall variance between 1979-1995). The MJO is a 30-60 day oscillation around the equator and proven to be a source of weather variability in the tropics on an intraseasonal timescale (Madden & Julian, 1994).

Seasonal prediction of rainfall in equatorial east Africa is of large interest to the climate community, with a large amount of research carried out over the last 30 years (Nicholson, 2017). Although reliable seasonal forecast have strong potential benefits, both analysis and prediction of rainfall is challenging because of the complexity of the rainfall regime (Diro *et al.,* 2011; Segele *et al.,* 2009). Due to the large correlation between the months and strength and persistence of teleconnections, the short-rains are more predictable than the long rains in both statistical and dynamical models (Dutra *et al.,* 2013, Mwangi *et al.,* 2014). Generally, statistical models outperform dynamical models in terms of forecast skill and have higher predictability at longer lead times (Chen & Georgakakos, 2015). The use of atmospheric variables (such as zonal winds and low-level circulation) in statistical models is underexploited, but is shown to be able to improve forecast skill (Nicholson, 2017). Physically based and dynamical General Circulation Models (GCMs) are able to issue seasonal forecasts of rainfall and are available from different climate forecast systems. Wanders & Wood (2016) applied a linear weighting method to four seasonal forecasts to create a multi-model. The multi-model was shown to have higher skill compared to a standard equally weighted multi-model mean.

Cohen *et al.* (2019) argue that the climate community and seasonal forecasting centres rely almost entirely on dynamical models and the potential of modern statistical techniques referred to as 'machine learning' is underestimated and underexploited. With regard to seasonal prediction of rainfall, most studies employ simple linear regression techniques. However, nonlinear relationships are shown to exist between large scale factors such as ENSO, and the surface and upper level zonal winds over the central equatorial Indian Ocean and rainfall, on an interannual basis (Black *et al.,* 2003). For example, Nicholson (2015) showed that in wet years, 3 to 4 of the aforementioned factors were associated with anomalous conditions, whereas in dry years, merely 1 or 2 and sometimes even none of the factors were associated with anomalous conditions. In addition to the amount of factors, wet years showed strong linear relations with the factors, but in dry years the sign of individual factors seemed to play a role, indicative of a non-linear response to changes in these factors.

In contrast to linear models, Artificial Neural Networks (ANN) are a highly flexible model class based on data driven learning. As a result, ANNs make no implicit assumptions about a systems behaviour and are therefore applicable to both linear and nonlinear tasks. In the context of the problem in this study, this is highly desirable. In practice, ANNs have proven successful in the field of rainfall prediction due to the ability of ANNs to 1) learn from the past (data driven learning) and 2) to account for the aforementioned linear relationships between rainfall and the large scale climate indices (Parmar *et al.,* 2017; Nayak, *et al.,* 2013). Mwale & Gan (2005) acknowledged the nonlinear and nonstationary characteristics of the data (rainfall and SST) and showed that using an artificial neural network model calibrated by a genetic algorithm (ANN-GA) outperforms a traditional linear canonical correlation model (CCA) often used in prediction studies. The Pearson corelation was used as a performance measure and was observed to be 0.25-0.55 for the CCA versus 0.70-0.90 for the ANN-GA, in East Africa.

Objectives and Research Questions

The aim of this study is to forecast monthly rainfall totals at a set of locations in the equatorial east African region, using an ANN model with seasonal forecasts from ECMWF SEAS5 in combination with climate indices.

The following specific objectives are defined:

1. Compile a pre-processed dataset holding all inputs, including rainfall data at multiple gauged locations.
2. Set up a model routine that allows forecasting at a 1-6 month lead time.
3. Evaluate the skill of the ANN model relative to: climatology, a multivariate linear model and the seasonal forecast of ECMWF-SEAS5.
4. Evaluate the skill of the ANN model with regard to anomalous seasonal rainfall.
5. Evaluate the feature importance of individual features for all lead times and stations.

Study Area

The study area is located between roughly 30-41°E and 4°N-7°S. It encompasses the countries of Tanzania, Kenya, Uganda, Rwanda and Burundi, and is similar to the study area of Woodman (2020). A digital elevation model is shown in Figure 1. A map of mean annual rainfall is shown in Figure 2. Largest rainfall amounts are found in Uganda and around Lake Victoria. Smallest rainfall amounts are found in large parts of northern and eastern Kenya.

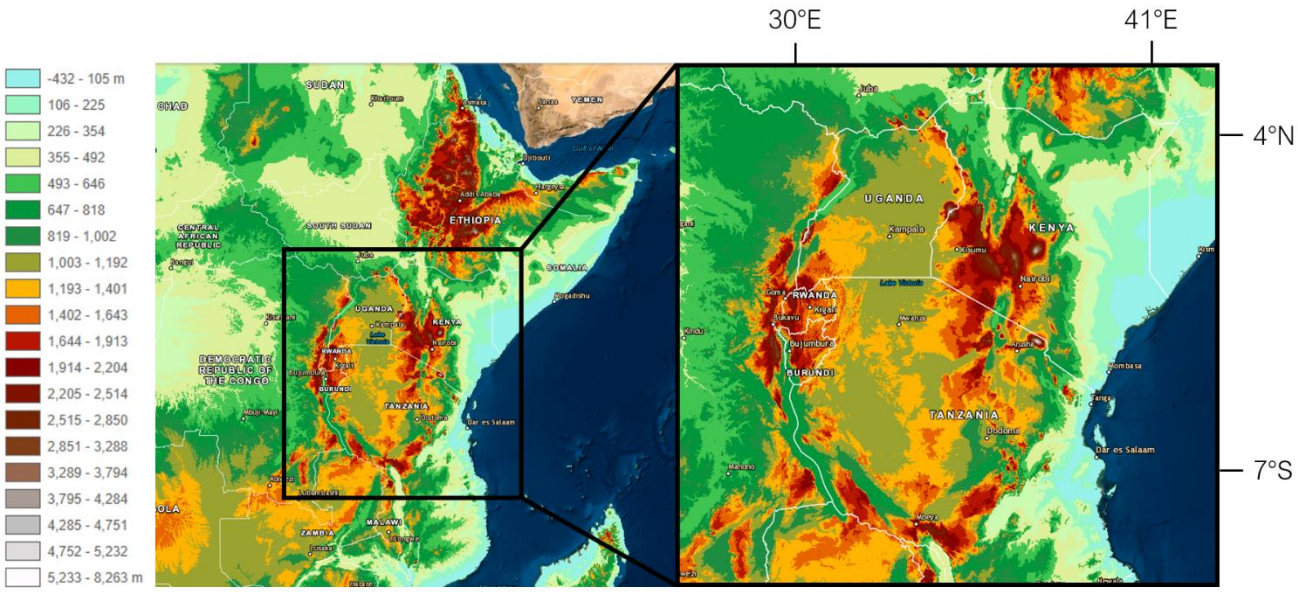


Figure 1 Digital Elevation Model (GTOPO30) of the study area (USGS, 1993) .

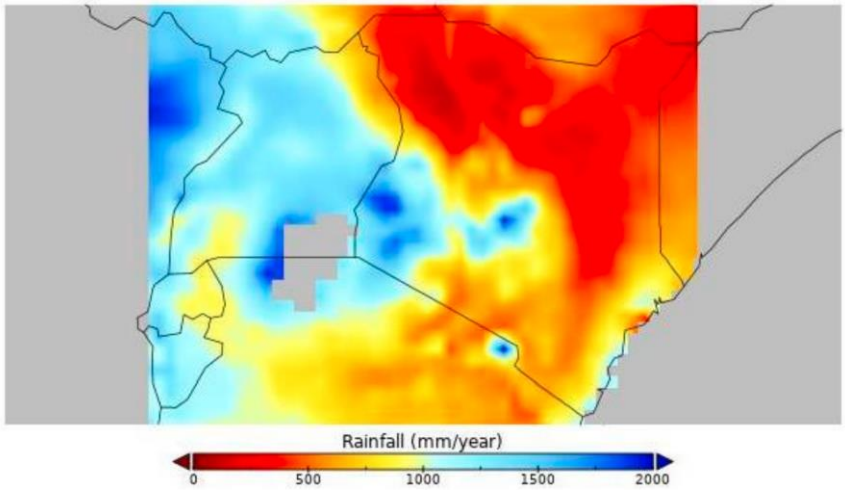


Figure 2 Mean Annual Rainfall in EEA. Adopted from: Woodman (2020).

Figure 3 gives an overview of the locations of the 17 rainfall stations considered in this study, and their elevation. Most of the stations are located in Kenya and few in Northern Tanzania and Uganda. Elevations range from 138 m to 1869 m.

Topography & rainfall

The topography is highly variable due to the presence of the East African rift that has a north-south orientation and two branches running through the area. The area is characterized by lowlands and highlands, with Mt. Kilimanjaro peaking at 5895 meters above sea level. Although mean elevation has little influence on rainfall amounts, it has strong control on rainfall frequency (Camberlin *et al.,* 2014). Low topography in northern parts of Kenya combined with high topography in Ethiopia result in a channel-like shape of the landmass with a south-east to north-west orientation. As a result, a low-level jet, referred to as the 'Turkana Jet' is formed. It is shown to affect low-level divergence in the area and is linked to the November-March northeast monsoon and the May-September southwest Monsoon (Figure 4; Nicholson, 2016). Spatially, increased values of divergence are associated with decreased values of rainfall (Figure 2).

Coastal climate

The coastal regions up to 50 kilometres inland have a different climate compared to inland regions due to sea breeze circulation in this area and its interaction with the Indian ocean monsoon. Long rains are enhanced, rainfall maximum occurs in May (instead of April for inland areas) and a moderate decrease of rainfall in the summer season (Camberlin & Planchon, 1997).

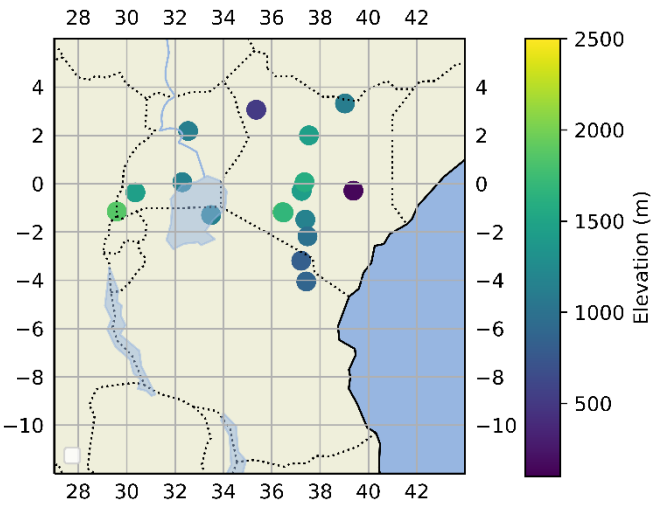


Figure 3 Rainfall stations in the research area.

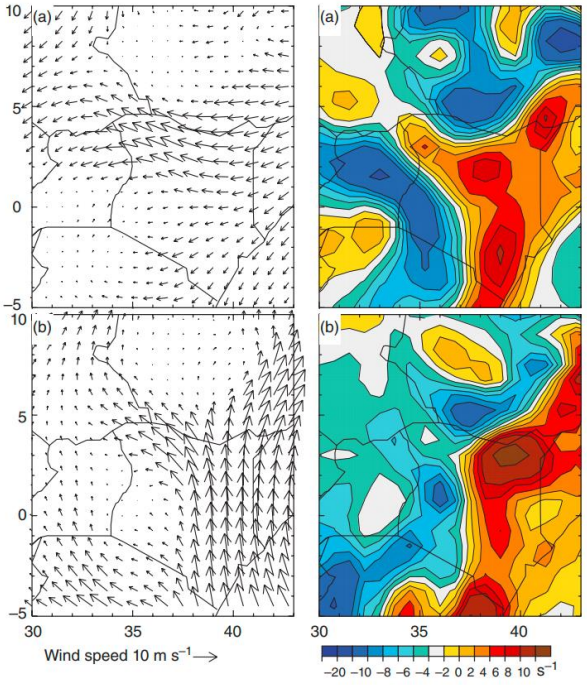


Figure 4 Mean vector wind and divergence at 850 mb during Nov-Mar (a) and May-Sep (b). Adopted from: Nicholson (2016).

Methods

Figure 5 shows a schematic overview of the workflow of this study. In this chapter, the data and the modelling process will be discussed consecutively.

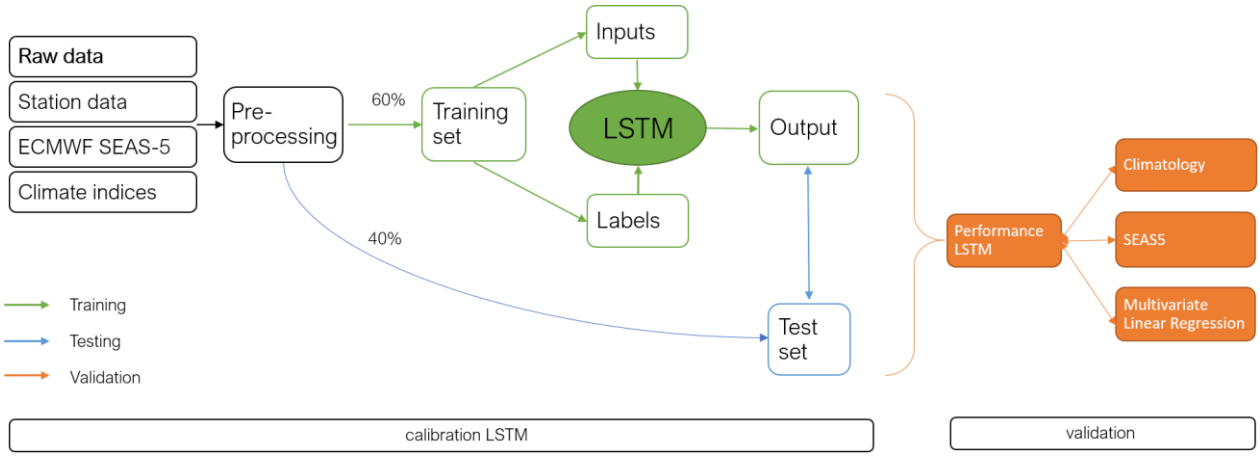


Figure 5 Schematic overview of the workflow in this study.

Data

Station data

The availability of monthly rainfall data on a station basis has been in the decline since the 1980's in equatorial east Africa (Camberlin & Phillipon, 2002). Therefore, post-1990 temporal coverage of station data was relatively limited. However, Professor Sharon S. Nicholson (Florida State University) has provided a helping hand by sharing her personal gauge dataset for this study. The dataset compromised rainfall data derived from meteorological and agricultural bulletins and from unpublished archives from meteorological services (Nicholson, 1986). Stations were selected to meet the following criteria: 1) near full temporal coverage in between 1993-2011 and 2) <2 missing values. Missing values in the station data were replaced by the mean monthly rainfall, corresponding to the month that is missing. Furthermore, a basic analysis of the mean rainfall for each station was carried out to gain insight in the distribution of the rainfall. To be able to investigate differences in the long rains and short rains, stations with a clear bimodal distributions were of specific interest in this study. Stations without a clear bimodal distributions were removed. These included mostly stations where peak rainfall was observed outside of the defined rainy seasons.

Seasonal forecast data

Seasonal forecasts at a lead time varying from 1-6 months were obtained from the long-range forecasting system SEAS5, from the European Centre for Medium-Range Weather Forecasts (ECMWF). Both forecasts for precipitation as well as temperature were used in this study. More specifically, so called hindcast products (forecasts that are produced retrospectively) were used. The hindcast period available was 1993-2016, setting a lower limit to the temporal range of the study.

Large scale climate indices

Table 2 shows the results of a cross-correlation test with a 12-month window between rainfall in 4 sub-regions and two lagged climate indices carried out by Woodman (2020). The climate indices include the Dipole Mode Index (DMI) and the Multivariate Enso Index (MEI). It is shown by the

correlation coefficient (r) that significant correlation exists between rainfall and climate indices up to a lag of -4 months for MEI in subregion 4. In this study DMI and MEI were also be used. Although the correlation between rainfall and an individual predictor may be low, the ANN model is able to model the nonlinear relationships between multiple predictors and rainfall, and after training the model output is expected to have a larger correlation with rainfall (Abbot & Marohasy, 2012). Berhane and Zaitchik (2014) showed a large influence of indices of the Madden-Julian oscillation between 80° and 120° of on rainfall in the study area. The indices described were retrieved from the Royal Dutch Meteorological Institute (KNMI) Climate Explorer.

Table 2 Cross-correlation in a 12-month window between rainfall, DMI and MEI expressed by the r (modified after: Woodman, 2020)

Region	Description	DMI		MEI	
		r	Lag (months)	r	Lag (months)
EEA	Full region: North-Tanzania, Kenya, Uganda, Rwanda, Burundi	0.34	-1	0.20	-3
SR1	Eastern half of Kenya	0.37	0	0.26	-2
SR2	Northwest Kenya	0.30	-1	0.15	-3
SR3	Western Uganda, Rwanda	0.22	-1	0.14	-4
SR4	North Tanzania, southwest Kenya	0.29	-2	0.19	+3

(EEA: Equatorial East Africa, SR1-SR4: sub-region 1 – sub-region 4)

The modelling process

For each individual lead time, a separate model was created. This method is also referred to as a direct multi-step forecasting strategy (Brownlee, 2018). For each individual lead time model, the corresponding ECMWF SEAS5 data was selected (i.e. for the model at a lead time of two months, the SEAS5 input data was also produced at a lead time of two months). A function was developed to automatically generate the correct input dataset based on the lead and lag times of the model and features, respectively.

The machine learning model that was used in this study is an artificial neural network (ANN). This model class is highly flexible model for relating a collection of inputs (e.g. lagged measurements or model forecasts at a set of locations) to a collection of outputs (in this case rainfall at different stations/locations) (Cohen *et al.,* 2019). The inputs undergo a series of non-linear transformations in the ANN's hidden layers based on the training data. These transformations are associated with weights that are optimized to minimize the prediction error of the model.

In this study, the main interest was to forecast monthly rainfall based on a set of inputs. There are many ANN types and it is impossible to know in advance which type of network will perform the best (Abbot & Marohasy, 2014). Abbot & Marohasy (2012) used a simple recurrent network (SRN) called a Jordan-Elman network (Elman, 1990; Ding *et al.,* 2013) to produce medium-range monthly rainfall forecasts at 1-3 month lead based on several climate indices and atmospheric variables. In this study, a Recurrent Neural Network (RNN) called a Long Short-term Memory (LSTM) was used. The LSTM cell is more successful at capturing long-term patterns and faster at training than other recurrent neural networks (Géron, 2019). In the following section, RNNs, memory cells and the LSTM will be described and explained schematically.

RNNs, memory cells & LSTM

Inspired by biological neurons, the most basic neural network is based on an artificial neuron called a linear threshold unit (LTU). The LTU computes a weighted sum of the inputs (Figure 6a). A weighted sum (lower half of the circle in the figure) of the input values is applied to a step function (upper half of the circle in the figure) resulting in the output.

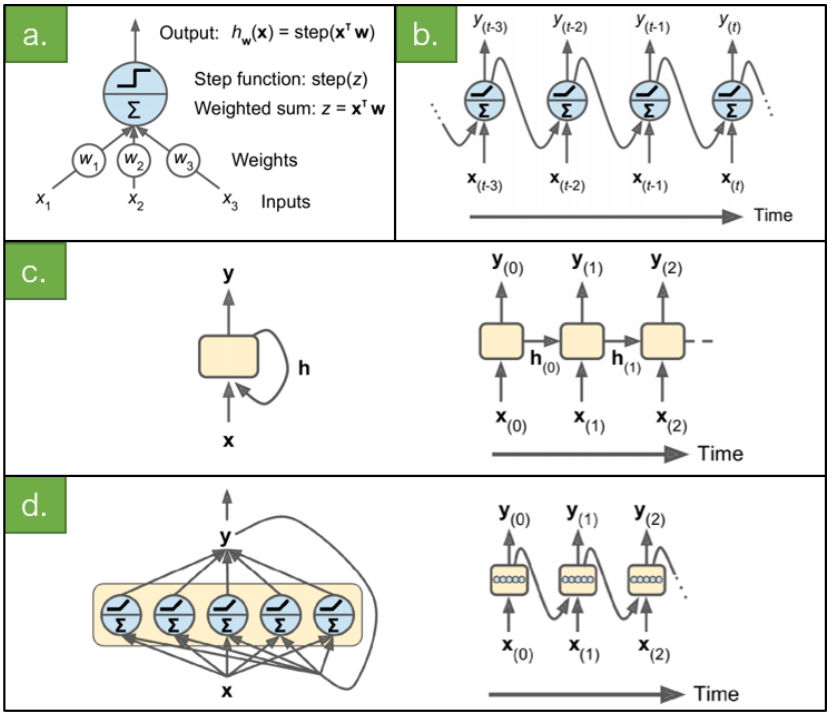


Figure 6 a) linear threshold unit (LTU) b) simple recurrent neural network c) schematic (memory) cell (left), through time (right)(d) layer of neurons (left) through time (right) (modified after: Géron, 2017)

Although there exists many types of RNNs, the most simple one consists of one neuron that feeds its output back into itself as shown schematically in Figure 6b. Since the output of a recurrent neuron is dependent on all the inputs of previous timesteps, it has a form of memory. A (memory) cell is the part of the neural network that preserves a certain 'state' across timesteps (Figure 6c). Generally, the state of the cell at timestep t ($\boldsymbol{h}_t$) is a function of inputs at the timestep and the state in the previous timestep ($\boldsymbol{h}_t = f(\boldsymbol{h}_{t-1}, \boldsymbol{x}_t)$. The output $\boldsymbol{y}_t$ at the timestep is also a function of the state at the previous timestep and the inputs at the current timestep. Stacking multiple recurrent neurons such as the one in Figure 6a, results in a layer of recurrent neurons (Figure 6d). Note that because of the stack the output is no longer a scalar, but a vector.

The LSTM cell is an advanced version of a basic cell as described above and introduced by Hochreiter & Schmidhuber (1997). The LSTM cell is more successful at capturing long-term patterns and faster at training than other neural networks. The structure of an LSTM cell is shown in Figure 7. Instead of only one vector representing the (short-term) state of the cell, as in a basic cell as described above, the state of an LSTM cell is split into two vectors $\boldsymbol{h_t}$ and $\boldsymbol{c_t}$ that represent the short-term and long-term state, respectively. The main idea behind an LSTM cell is that it can learn what it should store in, discard from and read from its long-term state. This is controlled by the input, forget and output gates, respectively. The cell contains four layers, of which the layer outputting $\boldsymbol{g_t}$ is the main layer. In a basic cell, this is only a layer that is present and will return its output to $\boldsymbol{y_t}$ and $\boldsymbol{h_t}$ directly. The behaviour of this layer is the same as explained above, analysing the current input and previous (short-term) state. In contrast to a basic cell, in an LSTM cell the most important output of this layer is stored in the long-term state $\boldsymbol{c_t}$. The other three layers are gate controllers and ouput values are between 0 and 1, based on the logistic activation fuction. The element wise multiplication allows for the control of the gates, closing when they are 0 and opening when they are 1.

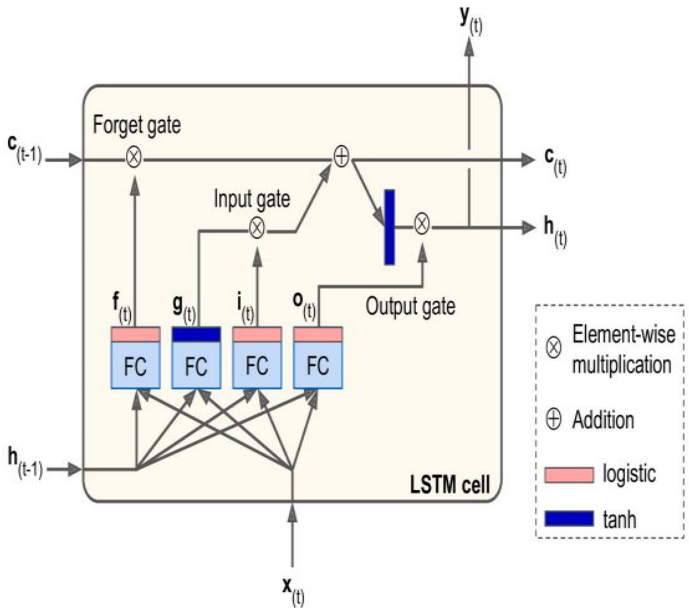


Figure 7 Architecture of an LSTM cell. Adopted from: Géron (2017).

The LSTM model has two important parameters that can be adjusted manually: the number of neurons and the batch size. These parameters are referred to as hyperparameters. Increasing the number of neurons in the LSTM network, increases the learning capacity of the network and training time (Brownlee, 2018). Using too many neurons might cause overfitting. During training, the LSTM network updates its weights every time a batch of training data is fed into it. The batch size therefore determines how often network weights are updated. In the case of this study, a batch size of 6 corresponds to 6 months of training data. To properly align the batches with the training and test data, batch sizes tested were chosen to be a factor of both the train and test set size.

With regard to the terminology, names and definitions vary between the statistical literature and the neural network literature. Table 3 indicates some of these differences to prevent confusion. In this study, neural network jargon will be used. In neural network literature, there are no equivalents for the terms ‘sample’ and ‘population’ but data is often divided in a ‘training set’ and ‘test set’ that are used for cross-validation (Sarle, 1994).

Table 3 Conversion table of statistical and neural network jargon (Sarle, 1994).

Statistical jargon	Neural Network jargon
Variables	Features
Independent variables	Inputs
Predicted values	Outputs
Dependent variables	Targets or training values
Residuals	Errors
Estimation	Training
Estimation criterion	Error function
Parameter estimates	(Synaptic) weights

Pre-processing the data

Before training, all features were normalized so that the feature distribution had a range between 0 and 1. This was done by applying the following statistical transformation:

$$X_{std} = (x - \frac{\min(X)}{\max(X)} - \min(X)) \quad 1)$$

$$X_{scaled} = X_{std} * ((\max(X) - \min(X)) - \min(X)) \quad 2)$$

where x represents a datapoint taken from the distribution X.

Removing trends and seasonality from input data is common practice in time series forecasting with models such as the autoregressive integrated moving average (ARIMA) model. However, RNNs are capable of learning seasonal trends from the data. Géron (2019) noted that detrending is therefore not necessary, as performance is expected to only increase slightly at best.

Training the model

The quality of a machine learning model can be expressed in terms of its variance and bias. The variance in this case refers to the spread in performance when the model is exposed to data that it was not trained on. A high-variance model might be able to reach a very good fit to the training data but fail to generalize well to unseen data. This is also referred to as overfitting. Bias, on the other hand, refers to how well the model is able to capture essential patterns in the data. A model with high bias will fail to capture these essential patterns and result in underfitting the data. An optimal model can be described as one that has low variance and low bias so that is both captures the essential patterns in the training data, but also generalizes well to unseen data. As variance and bias are negatively correlated, at the basis of every machine learning problem is the so called bias-variance trade-off.

Before training, the data was split into two sets: the training set and the test set. The training and test sets compromised 60% and 40% of the original data, respectively. Since the data is a time-series and a certain level of time-dependency is assumed, it was not desirable to shuffle the data and split randomly. Therefore, the training data compromised a period from the start of the dataset to 60% of the total time period of the dataset. The test set was then defined as the last 40% of the time within the dataset. This created a realistic approach to the problem, since the test results are based on data that is collected after the training period. This is comparable to an operational forecast scenario, where future conditions are not known.

During the training phase, the neural network optimizes the weights so that the model output matches the training target as well as possible. In this study, the training targets were the monthly rainfall totals as measured at different stations. The RNN network processes the training set that is a time-series, in a step-by-step manner. Since it was desirable for the model to be trained on multiple timesteps, the LSTM layer was defined to make a prediction based on the inputs at every timestep. In this study a separate model is created for every lead time. Figure 8 shows the process as described above for a lead 1 model. In this set-up, the model produces a continuous time-series and was able to account for seasonal biases in the seasonal forecasts of the GCMs.

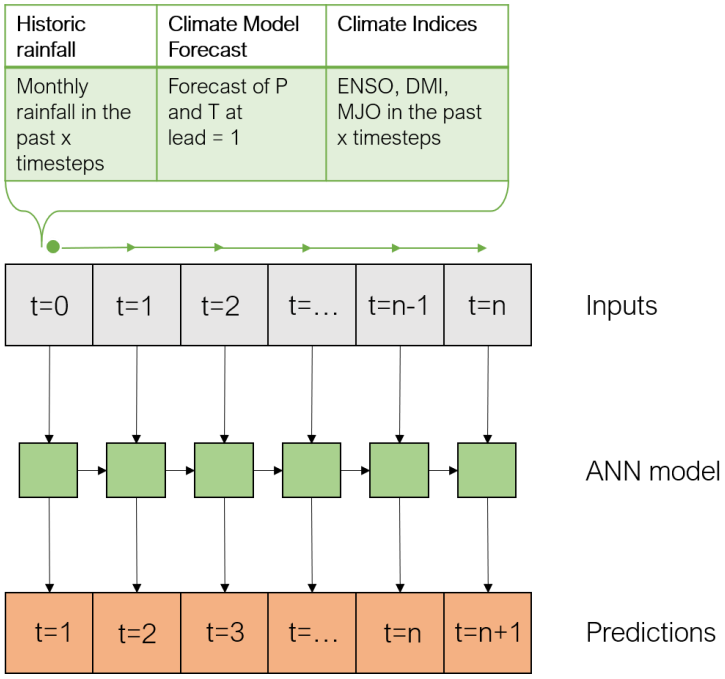


Figure 8 Schematic overview of the single- step, lead time=1 model.

To prevent both the linear and the LSTM models from overfitting, an automated procedure was implemented that monitored both the training and test error during training. The procedure stopped training after reaching a minimum train and test error on the data. There were several optimization algorithms available that could be implemented for this procedure. In this study, the Adam optimisation algorithm was implemented for its computational efficiency and the little memory requirements (Kingma & Ba, 2014). The stochastic nature of the optimization algorithm may produce slight differences in performance (Brownlee, 2018).

Forecast verification

Verification methods for a continuous predictand

The forecast model designed in this study produced forecasts of a continuous predictand (rainfall) in a non-probabilistic fashion. In this case, two scalar measures are in common use: the Root Mean Square Error and the Mean Absolute Error (Wilks, 2006).

The Root Mean Square Error (RMSE) is a very basic measure of the difference between predicted and observed values. It is the square root of the average of squared errors as can be seen in equation 3, where $\mathbf{y}_t$ represents a value from the prediction and o_t represents the observed value. Values of RMSE are always positive and values close to 0 indicate a near-perfect fit. In addition, it takes the square root of the Mean Squared Error (MSE), which is also applicable, but the RMSE has the advantage that errors can be expressed in the same physical dimension as the variable that is evaluated: rainfall.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}(\mathbf{y}_t - o_t)^2}{N}} = \sqrt{MSE} \qquad 3)$$

The second scalar measure, the Mean Absolut Error (MAE), is less sensitive to outliers compared to the MSE where forecast errors are squared. Since both RMSE and MAE have the same physical dimensions as rainfall, they may be interpreted as a typical magnitude for forecast errors (Wilks, 2006).

$$MAE = \frac{\sum_{t=1}^{N}|\mathbf{y}_t - o_t|}{N} \qquad 4)$$

Since the RMSE squares the errors, it is more sensitive to outliers than the MAE. In this study, the largest errors were expected to be found in the months with the largest amounts of rainfall: the long rains and the short rains. Combining these two observations, the RMSE was considered the most relevant metric in this study. To prevent underfitting of rainfall in the rainy seasons, the RMSE was the selected metric to be observed by the Adam optimization algorithm during training.

The model used in this study predicted monthly rainfall at different stations. However, the mean annual rainfall may have differed between rainfall stations, affecting the magnitude of the RMSE. To allow for a comparison between stations, a weighted version of the RMSE was used: the Weighted Non-dimensional Index (WNDI). The WNDI was defined as the RMSE relative to the mean annual rainfall as described in equation 5 (Johns *et al.,* 2006). A WNDI of 0 indicates a perfect score. A WNDI of 1 equals an RMSE value equal to 1/12th of the mean annual rainfall and indicates a moderate score. A WNDI of 0.5 can be interpreted as a good score.

$$WNDI = \frac{RMSE * 12}{mean\ annual\ rainfall} \qquad 5)$$

As a measure of linear correlation between the forecasts and the actual values, the Pearson product-moment coefficient of linear correlation was used. Correlating the forecasts directly with the observations may be misleading, as seasonal variation might cause very high values of correlation (Persson, 2001). To account for seasonal variation, the anomaly correlation coefficient (ACC) was calculated according to equation 6.

$$ACC = \frac{\overline{(f-c)(a-c)}}{\sqrt{\overline{(f-c)^2}\overline{(a-c)^2}}} \quad 6)$$

Here, f and a represent the forecasted and actual values of precipitation and c represents the mean monthly rainfall. For this reason, the ACC is a skill score relative to the climate. An ACC of 0 represents no correlation, ACC>0 represents an improvement over the climatology and ACC=1 represents a perfect forecast.

In order to objectively assess the quality of the forecasts produced by the LSTM, a systematic comparison was made with respect to climatology, ECMWF SEAS5 and a multivariate linear model. The assessment was based on the WNDI and the ACC. Since the ECMWF SEAS5 was used as input to the models, both the linear and LSTM model were expected to have better performance. It was, however, useful to use SEAS5 as a benchmark to quantify the actual improvement over SEAS5. Finally, the linear model was used as a third reference to test the potential advantage of a non-linear model in seasonal rainfall forecasting.

Verification methods for a discrete predictand

So far, methods of forecast verification of a continuous variable (rainfall) have been presented. To assess the ability of the models to forecast extreme seasonal rainfall totals, the 2x2 contingency table was used (Figure 9, Wilks (2006)). The contingency table assumes a dichotomous outcome (either the occurrence or non-occurrence of a certain event) and compares the observed frequencies of occurrence and non-occurrence with the forecasted frequencies. Four possible scenarios are present: a) the event is observed and forecasted, b) the event is not observed but forecasted, c) the event is observed but not forecasted and d) the event is not observed and not forecasted.

To analyse the total rainfall over a season three specific events were defined: below-normal seasonal rainfall, normal seasonal rainfall and above-normal seasonal rainfall. Assuming a near-normal distribution of the seasonal rainfall totals, normal seasonal rainfall was defined as the $mean \pm 0.8 \times std$ (Figure 10). The factorization of $0.8 \times std$ was chosen based on trial-and-error to obtain a representable discrimination between categories. For each of the events a 2x2 contingency table was constructed, indicating either the occurrence or non-occurrence of the event.

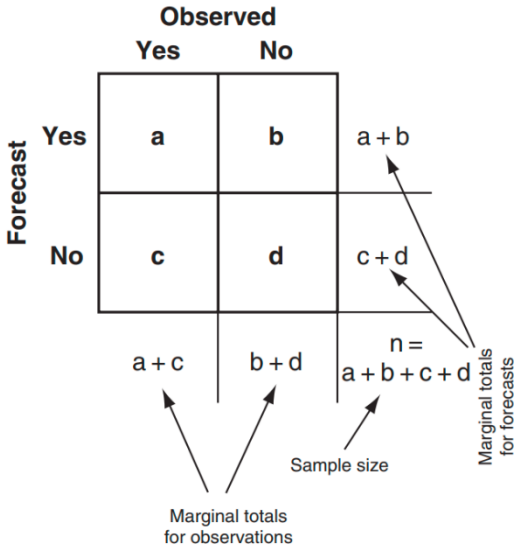


Figure 9 The 2x2 contingency table comparing observed and forecast frequencies of occurrence and non-occurrence of a dichotomous event (adapted from: Wilks, 2006).

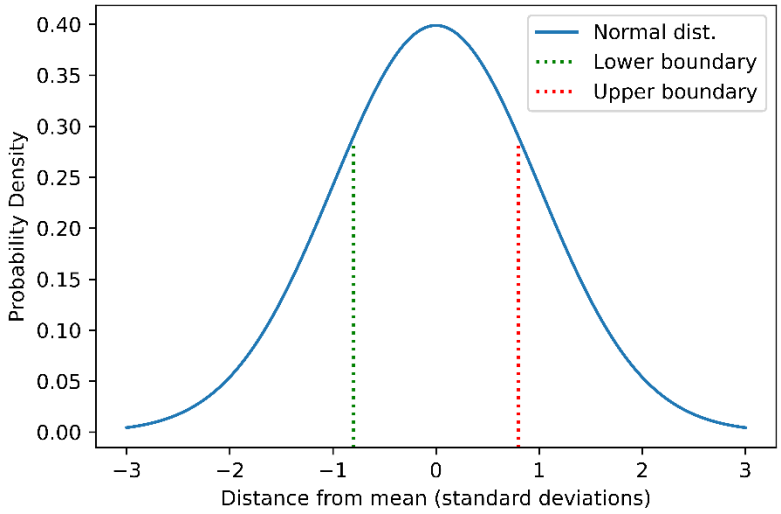


Figure 10 The defined categories for below normal conditions (to the left of the lower boundary), normal conditions (between the lower and upper boundary) and above-normal conditions (to the right of the upper boundary), based on the normal distribution. Boundaries are located at +/- 0.8 standard deviations from the mean.

Based on the results obtained from the constructed contingency tables, several scalar attributes were computed to allow for an evaluation of the results. The letters (a, b, c, d) used in the equations are consistent with the definitions in Figure 9.

The bias (B, equation 7), is the ratio between the number of yes forecasts and the number of yes observations. Unbiassed forecasts will have a bias of 1 (indicating that the number of yes forecasts is equal to the number of yes observations).

$$B = \frac{a+b}{a+c} \quad 7)$$

The hit rate (H, equation 8) is the ratio between the number of yes forecasts that are also observed and the total number of yes observations and is also called the probability of detections (POD). A perfect hit rate is equal to 1 (indicating that all observed occurrences are also forecasts).

$$H = \frac{a}{a+c} \quad 8)$$

The false alarm rate (F, equation 9) is the ratio between the number of false alarms and the total number of cases for which the event did not occur. The best possible false alarm rate is 0 (indicating that an event was never forecast when it also did not occur.

$$F = \frac{b}{b+d} \quad 9)$$

In addition, the Clayton Skill Score (CSS) was computed (Clayton, 1927). The CSS is defined following equation 10. When an event occurs more frequently when forecast than when not forecast, the CSS indicates a positive skill. A CSS of 0 indicates a random forecast.

$$CSS = \frac{a}{a+b} - \frac{c}{c+d} = \frac{ad-bc}{(a+b)(c+d)} \quad 10)$$

Sensitivity analysis

Repeated experiments

Multiple repeated experiments were carried out to find the optimal lag times for rainfall and the climate indices, as well as the optimal setting for the hyperparameters of the LSTM (number of neurons and batch size). A valid range of values was selected for the parameter under investigation and the model was run iteratively over each possible value. Each iteration was repeated 30 times to gain insight in the spread of performance for each setting, accounting for potential differences in performance due to early stopping and the stochastic nature of the learning algorithm. After the procedure was completed for a certain parameter under investigation, a setting with both lowest median RMSE and lowest spread was considered to be optimal.

Robustness experiment

The data split of 60% training data and 40% test data was initially done chronologically, as described before. However, Woodman (2020) showed significant trends throughout the research area in terms of SPI06, indicating both wetting and drying. SPI06 refers to the calculation of the SPI based on an accumulation period of 6 months. In case of a chronological split and the presence of a trend in the data, the training set does not represent the test set correctly. To test the variation in performance of the LSTM model when it is trained on shuffled data, a repeated experiment was designed. In the experiment, the years in the training and test sets were shuffled, but the months within each year were kept in chronological order. The procedure of the experiment is shown schematically in Figure 11.

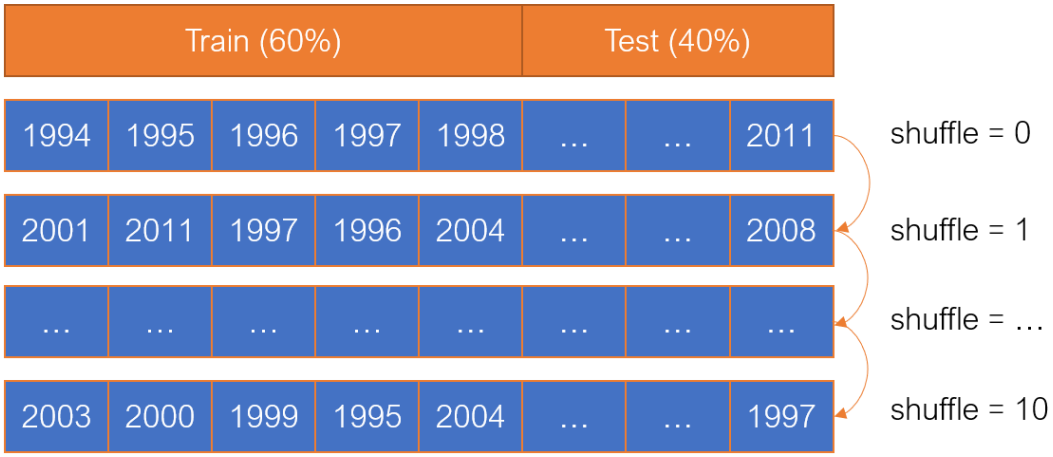


Figure 11 Schematic overview of the robustness experiment.

Feature importance

The input dataset consisted of lagged observed rainfall, lagged indices and seasonal forecasts from SEAS5. Recognizing the potential differences in the relative importance of the features that may have been be present between stations, an experiment was designed to indicate the relative importance of each feature.

In a linear model, it is valid to use the weights for feature importance testing after the model is trained, because the weights represent a direct linear relationship between the inputs and outputs. For a non-linear model, such as an LSTM, this is not valid and an alternative method should be used. In the case of this study, the effect of individual features on the performance was tested in the following manner. First, a benchmark performance was created by running the LSTM model on none of the features, except the lagged rainfall amounts. Secondly, the model was run with the lagged rainfall amounts and one other feature under study. Finally, the difference in performance is assessed by means of the RMSE. This step was repeated for every feature in the input dataset and for all lag times.

The criterium for a feature to be considered 'relevant', an increase in performance of >1% was used as a threshold.

Software

All programming was carried out in the Python programming language (Python Software Foundation, 2020). Additional packages used included NumPy (Oliphant, 2006), TensorFlow (Abadi *et al,* 2016), Scikit-Learn (Pedregosa *et al.,* 2011). The Keras API (Géron, 2019) was used for its built-in layers such as LSTM, so that advanced configuration choices will not have to be made.

Results

Literature study

Defining drought

The qualitative description of drought is 'a deficit of water relative to normal conditions' (Lloyd-Hughes, 2013). Traditionally, the drivers for drought were exclusively viewed in relation to climate variability, but Van Loon *et al.,* (2016) propose the inclusion of human activities into the definition, since both natural and anthropogenic processes interact in a complex way. Here, drought is defined as a time-dependent propagation from meteorological drought to soil moisture drought to hydrological drought (Figure 12). Meteorological drought is defined as a prolonged and abnormal moisture deficiency in the atmosphere (Huschke, 1959). The occurrence of a meteorological drought, might propagate into a soil moisture drought depending on boundary conditions and initial soil moisture conditions after a certain amount of time. The definition of soil moisture drought is 'a period of below-normal soil moisture levels. In the context of crop growth, this type of drought is often referred to as 'agricultural drought': 'a shortage of water harmful to man's agricultural activities' (Heathcote, 1974). Again, soil moisture drought might propagate into hydrological drought, defined as below-normal streamflow, groundwater and lake or reservoir levels'. In this research meteorological drought will be addressed.

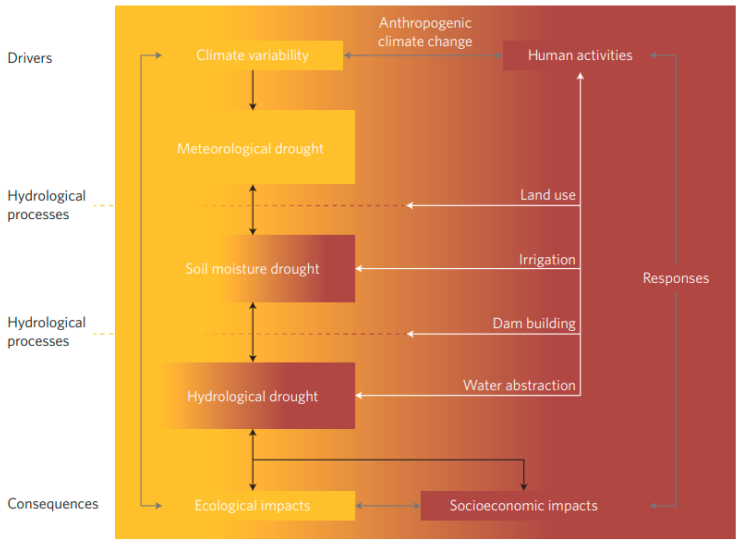


Figure 12 Drought definition framework. Adopted from: Van Loon et al. (2016).

To be able to study drought, a qualitative definition is not sufficient. McKee *et al.* (1993) noted that droughts are implicitly related to time and proposed a universal, versatile tool to be able to quantify drought: the Standardized Precipitation Index (SPI). The SPI is based on only one input parameter (precipitation) and is an index that is computed on the basis of a precipitation anomaly over a set time period relative to average climate conditions. Due to its universality, the World Meteorological Organization (WMO) has recommended the use of the SPI for all National Meteorological services around the world (Svoboda, 2012).

Rainfall variability in equatorial east Africa

Equatorial East Africa (EEA) has a semiarid-arid climate. Rainfall is controlled by a large set of controls that are depicted schematically in Figure 13 (Nicholson, 2017). These controls have effects on different timescales. For this reason, variability of rainfall will be discussed for different timescales in the context of underlying controls.

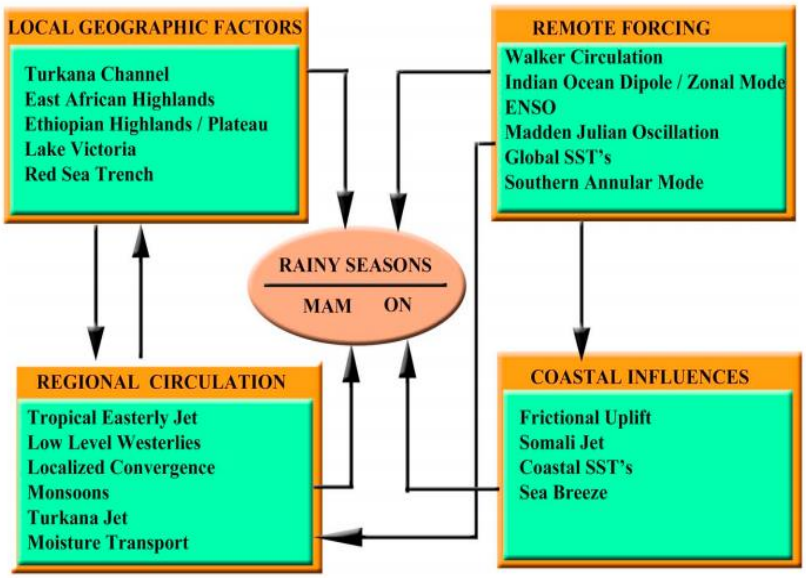


Figure 13 Schematic representation of Local Geographic Factors, Remote Forcing, Regional Circulation and Coastal Influences affecting the rainy seasons in eastern Africa. Adopted from: Nicholson (2017).

Seasonality

Rainfall in equatorial eastern Africa (EEA) is bimodal and concentrated in two rainy seasons: the "long rains" in the months March, April and May (MAM) and the "short rains" in the months October and November and December (OND). Differences in the definition of the rainy seasons exists between authors (for example, Nicholson (2017) defines the short rains to include only October and November (Figure 13)). Camberlin & Phillipon (2002) carried out principal component analysis to assess spatial variation in rainfall between months, and found that March and April and October and November had similar spatial patterns, but May and December did not. Consequently, a universal definition of the rainy seasons is not present. In this study, the long rains are defined from March-May (MAM) and the short rains from October-December (OND).

The definition of the migration of the Intertropical Convergence Zone (ITCZ) is one of the main controls with regard to seasonal rainfall variation. The ITCZ is defined as "a region near the equator where trade winds converge" (Miller, 1996). However, over equatorial East Africa and other tropical landmasses, trade winds do not exist and convergence of the northeasterlies and southwesterly monsoon flow are used here to refer to the ITCZ (Nicholson, 2018). The Indian monsoon creates a low-level easterly/notheasterly flow from November to March and a southerly flow from May to October and the peak of the monsoon months corresponds to the dry season in East Africa (Nicholson, 2017). The ITCZ is often associated with a rainfall maximum, but this assumption is not valid over EEA. For this reason, Nicholson (2009) proposed the use of the term tropical rain belt and suggested that the use of ITCZ should be avoided over EEA, except for coastal areas being influenced by trade winds. Yang *et al.* (2015) suggested that the bimodal annual cycle is a result of the annual cycle of monsoonal winds combining with the annual cycle of the Indian Ocean SST. Sea surface temperatures in the western Indian Ocean, near the coast of EEA, are higher during the rainy seasons and highest during the long rains. Therefore, the long rains are longer and more intense than the short rains (Yang *et al.,* 2015).

Intraseasonal variation and predictability

Intraseasonal variation of rainfall is highly crucial for agriculture, as it refers to the occurrence of dry spells and wet spells. Within the rainy seasons, dry spells occur about 3-4 times with a length of 5-10

days (Nicholson, 2017). The short rains are more variable and generally less intense compared to the long rains and suffer from more dry spells that exceed 7 and 14 days (Den Daas, 2016). The occurrence of a dry-spell within the rainy season can be detrimental to rainfed crop production and risk reduction for dry spell induced crop failures are among the key strategies for improvement of agricultural management (Rockström *et al,* 2010).

Pohl & Camberlin (2006) found a strong influence of the Madden-Julian Oscillation (MJO) on long rain variability (44% of long rains seasonal rainfall variance between 1979-1995). Common variance of MJO and rainfall on an interannual basis is variable (5%-53%). Dry conditions are associated with a weak MJO and a stronger Walker-like circulation over the Indian Ocean. The MJO is a 30-60 day oscillation around the equator and proven to be a source of weather variability in the tropics on an intraseasonal timescale (Madden & Julian, 1994). It is apparent as a large-scale eastward propagating system of convection, zonal winds and upper-level velocity potential (Hendon & Salby, 1994). Indices of MJO are constructed based on the geographical location of the convective centre that leads to increased precipitation and is expressed by the longitude. Berhane & Zaitchik (2014) studied specifically the impact of the MJO on east African rainfall on an intraseasonal basis between 1998 and 2012. They found that indices of MJO at 70°-80°E and 120°W show highest correlations with East African rainfall. They also found that, related to the MJO, two important mechanisms (thermodynamic and convergence) affect East African rainfall in both the long rains and the short rains (Figure 14). In the figure, the MSE refers to the Moist Static Energy and SLP to Sea Level Pressure. Both mechanisms result in increased precipitation in East Africa.

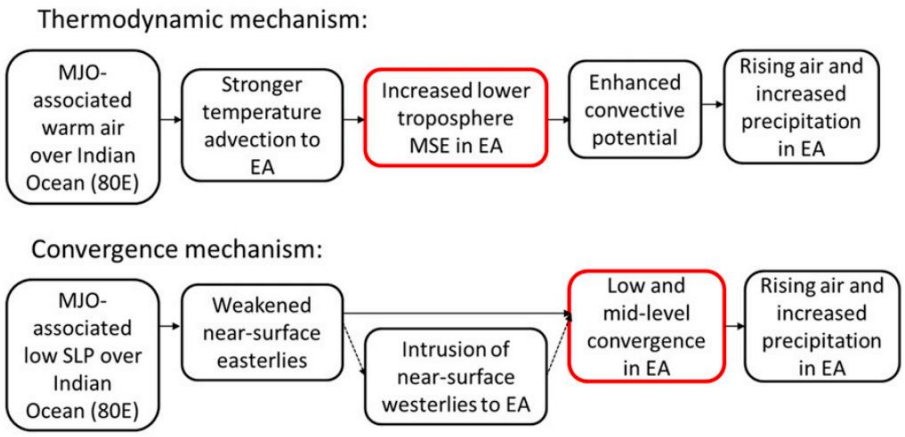


Figure 14 Meteorological mechanisms affecting East African rainfall, associated with the MJO. MSE refers to the Moist Static Energy and SLP to Sea Level Pressure (modified from Berhane & Zaitchik, 2014).

Interannual variation and teleconnections

When considering interannual variability of rainfall over the region, it is useful to note that the main source of interannual variability throughout the year is associated with the short rainy season even though rainfall is generally lower during this season (Camberlin & Phillipon, 2002). The long rains are less variable (or more reliable in other words) and have less of a relative effect on interannual variability (Den Daas, 2016).

Even though the annual rainfall field over the research area is extremely heterogenous, strong coherence in the patterns of interannual variability exists (Woodman, 2020; Nicholson, 2017). Traditionally, interannual variability of rainfall over the region and its periodicity was suggested to be related to the Southern Oscillation Index (SOI) (Nicholson & Entekhabi, 1986). This index is a

standardized index based on the observed sea level pressure differences between Tahiti and Darwin, Australia. SOI is indicative of the El Niño Southern Oscillation (ENSO), which is a periodic fluctuation (every 2-7 years) in sea surface temperature (El Niño) and the air pressure of the atmosphere (Southern Oscillation). Links of ENSO to interannual variability of east African rainfall are well established (Ropelewski & Halpert, 1986; Ogallo & Janowiak, 1988; Woodman, 2020; Keijzer, 2020).

The 1997/1998 El Niño event triggered a surge of research activity due to the strong and present link between ENSO and east African rainfall (Hansen *et al.,* 2011). In 1999, a sea surface temperature (SST) mode was discovered to exist in the Indian Ocean that is referred to as the Indian Ocean Zonal Mode (IOZM) or Indian Ocean Dipole (IOD) (Webster *et al.,* 1999; Saji *et al.,* 1999). The positive phase of the IOD is associated with anomalously warm SST's in the western Indian Ocean that are associated with anomalously large amounts of rainfall during the short rains (Black *et al.,* 2003; Woodman 2020; Keijzer 2020).

There exists uncertainty to whether or not ENSO and IOD are independent phenomena, since it is shown that an atmospheric response to variations in the Indian Ocean is required for ENSO to impact east African rainfall (Saji *et al.,* 1999; Goddard and Graham 1999). In addition, Goddard & Graham (1999) proved that zonal circulation over the Indian Ocean is critical for transmitting the ENSO signal to East Africa. Keijzer (2020) applied partial correlation between both ENSO, IOD and other teleconnections and rainfall in the Lake Manyara (Tanzania) catchment and found that IOD is least intercorrelated, whereas ENSO is correlated to a higher degree with other teleconnections. With respect to the actual correlation, Keijzer (2020) showed that IOD has the largest influence on OND rainfall compared to ENSO, MJO.

Phenomena such as ENSO, related to sea surface temperatures and air pressure have an effect on large scale circulation systems that affect rainfall patterns worldwide. Pohl and Camberlin (2011) looked at wind shear between the lower and upper troposphere that is embedded in these large scale atmospheric configurations and showed and linked this with regional rainfall anomalies. ENSO was shown to modulate this circulation most frequently on an interannual basis. La Niña (an exaggeration of normal zonal gradients) normally peaks between October and December encompassing the short rains in east Africa, causing below-normal rainfall (Camberlin & Philippon, 2002).

Predictability and forecasting

Seasonal prediction of rainfall in EEA is of large interest to the climate community, with a large amount of research carried out over the last 30 years (Nicholson, 2017). Although reliable seasonal forecast have strong potential benefits, both analysis and prediction of rainfall is challenging because of the complexity of the rainfall regime (Diro *et al.,* 2011; Segele *et al.,* 2009). Research has indicated that the short rains are relatively predictable whereas the long rains are not. With regard to types of forecasting models, statistical models generally outperform dynamical models in terms of forecast skill and have higher predictability at longer lead times (Chen & Georgakakos, 2015). The use of atmospheric variables (such as zonal winds and low-level circulation) in statistical models is underexploited, but is shown to be able to improve forecast skill (Nicholson, 2017).

Long Rains

For the long rains, sources of rainfall variability have proven difficult to pin down (Camberlin & Phillipon, 2002). Nicholson and Entekhabi (1986), showed that interannual variability of individual months in the long rains are uncorrelated. Camberlin & Phillipon (2002) obtained differences in variability between months by applying a principal component analysis and suggested that May should be considered separately from March and April due to different sources of variability. This was also recommended by Keijzer (2020). According to Zorita & Tilya (2002) the March and April rainfall is linked to surface

zonal winds and the zonal contrast between the Indian Ocean and the East African landmass. For May they found a strong link between the Indian monsoon, characterized by the meridional temperature contrast between the Asian continent and the Indian Ocean and meridional surface winds.

Camberlin & Phillipon (2002) were the first to accentuate atmospheric variables in their regression model. Among the predictors were SSTs in Nino 1.2, zonal wind over the Congo basin at 1000 mbar, geopotential height of the 500 mbar surface over the Near East and the east-west moist static energy gradient between the East African highlands and the Sahel. Correlations between predicted and observed March to May rainfall for Kenya/Uganda was 0.66 in cross-validation mode. They noted that although weak teleconnections are present during the long rains, there still is some predictability during this season.

Focusing on southern Ethiopia, eastern Kenya and southern Somalia, Funk *et al.* (2014) applied a principal component analysis (PCA) to CHIRPS March-May rainfall data over the East African Region between 1981 and 2013 to identify the two main modes of rainfall variability over the research area (PC1 & PC2). By recognizing more frequent droughts in the area due to a stronger Walker Circulation (Willams & Funk, 2011), a warming of the Indo-Pacific warm pool and increased western Pacific SST gradient, they showed that the two main modes of variability are tied to western-central Pacific and central Indian Ocean SST's. They used the WPG (West Pacific Gradient index; Hoell & Funk, 2013) and CIO (Central Indian Ocean index; Zhou *et al.,* 2017) in January to predict the two rainfall modes, respectively. The underlying idea here is that more recent analysis (1993-2012 (Funk *et al.,* 2013; Lyon & DeWitt, 2012) of east African rainfall indicates a link between a precipitation dipole over the western and central Pacific ocean. In addition, a combination of La Niña with a strong WPG produces large SST gradients across the entire equatorial Pacific and appear to be linked to increased subsidence over the Horn of Africa. These studies did not distinguish individual months within the long rains.

Nicholson (2014) used multiple linear regression and used several atmospheric variables including: omega, zonal and meridional winds at 925, 850, 700 and 200 mbar. Omega is a meteorological variable indicating the vertical motion of wind in the troposphere and at a synoptic-scale. An interesting conclusion of the study was that these variables provide a larger forecast skill than sea surface temperature and sea level pressure.

The maximum possible lead time for prediction of the long rains is about 2 months due to the spring predictability barrier, which is associated with limited predictability of ENSO in this time. This again, is related to the rapidly changing conditions in the tropical Pacific Ocean that make the prediction of the state of the ocean very difficult.

Short Rains

Nicholson and Entekhabi (1986), showed that interannual variability of individual months in the long rains are uncorrelated, while the opposite is true for the short rains. This is largely due to the strength and persistence of ENSO and IOD during the short rains and is associated with above-average rainfall (Ogallo et al.1988, Kiladis & Diaz, 1989; Woodman, 2020; Keijzer, 2020).

In more physical terms, ENSO and the IOD are both modulating the Walker Circulation. The Walker Circulation extends over the whole equator, but is characterized by an easterly flow at 200mbar and westerly winds as low-levels over the Indian Ocean and associated with rising in the east and subsidence in the west. This low-level flow modulates the short rains more strongly than ENSO (Bergonzini *et al.,* 2004). Increased westerlies, leading to a stronger Walker cell over the Indian Ocean and increased subsidence were related to 2005 and 2010 short rain droughts (Hastenrath, 2007; Hastenrath *et al.,* 2010). Lyon (2014) showed that SST anomalies in the Indian and Pacific

Oceans have a strong link to drought in the short rainy season, but that this link is much weaker during the long rains.

Camberlin *et al.* (2002) investigated the relationships between atmospheric dynamics over the global tropics with rainfall in the short rains. As a result, they constructed two indices by principal component analysis, indicative of circulation over the western Indian Ocean, meridional winds over the south-eastern tip of Africa and a monsoon index with a NE component at 200mb and a SW component at 850 mbar

Nicholson (2014) also predicted rainfall in the short rains (defined as October and November), again by implementing atmospheric variables (zonal, meridional winds and omega) and other variables (SST, SLP and IOD). Based on the values of the variable in May, a correlation of 0.80 was obtained, indicating the potential of forecasting a lead time of 6 months.

Woodman (2020) performed a time lag cross-correlation analysis in four subregions of EEA to investigate the link between drought variability and ENSO and IOD. He found a considerable influence of these factors on drought variability in the area, although this influence was not homogeneous.

The effect of ENSO and IOD on East African rainfall has been proven to be non-stationary (Figure 15, Nicholson, 2014). The figure indicates a decreasing correlation between ENSO and IOD and the short rains between 1982 and 1997, after which the correlation became stronger again after a regime shift (black line) in 1997.

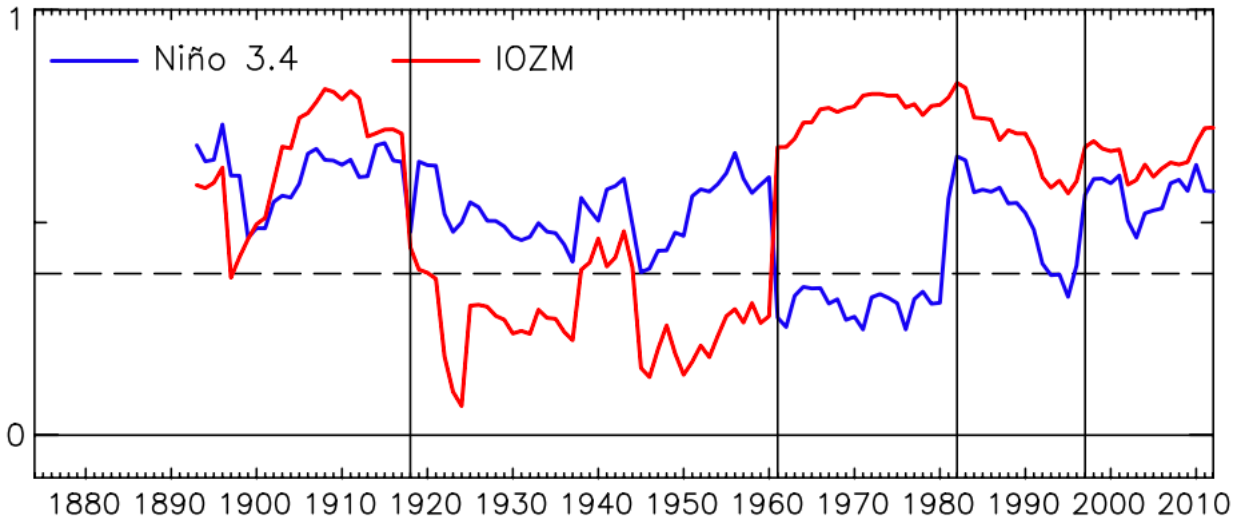


Figure 15 Sliding 20-year correlation between October and November rainfall and the indices for ENSO and IOD. Vertical lines indicate regime shifts. Values above the striped line are significant (alpha = 0.05). Adopted from: Nicholson (2014).

Seasonal forecasting by dynamical models

Physically based General Circulation Models (GCM's) are used to issue seasonal forecasts of rainfall and are available from different climate forecast systems. Forecasts are generally issued at a monthly or daily temporal resolution and 1° spatial resolution. These forecasts have improved for temperature but less so for precipitation (Kirtman & Pirani, 2009). This can be addressed to the inherent chaotic character of the atmosphere, where slightly different initial conditions may lead to a different resulting state of the atmosphere over time (Lorenz, 1963). To improve forecasting skill, multiple models can be combined to obtain a so called multi-model. Error cancellation can be achieved by running a multi-

model, whereas uncertainties in initial conditions are addressed by the ensembles within the individual models (Hagedorn *et al.,*2005; Sikder *et al.,* 2016). Error cancellation is a conceptual idea that refers to a situation where the single-model ensembles of two hypothetical models lie below and above the verification. When the two models are combined, the respective errors of the single-model ensembles with respect to the verification cancel each other out. Note that this is not always the case, but the concept is the most important reason for the superiority of multi-models (Hagedorn *et al.,* 2005) over individual model ensembles. The commonly used method to achieve this is to compute an ensemble mean (Hao *et al.,* 2018). Wanders & Wood (2016) applied a linear weighting method to four seasonal forecasts to create a multi-model. The multi-model was shown to have higher skill compared to a standard equally weighted multi-model mean, showing the potential of using a weighted multi-model approach.

Machine learning

Cohen *et al* (2019) argue that the climate community and seasonal forecasting centres rely almost entirely on dynamical models, and that the potential of modern statistical techniques, referred to as 'machine learning', is underestimated and underexploited. With regard to seasonal prediction of rainfall, most studies employ simple linear regression techniques. However, nonlinear relationships are shown to exist between large scale factors such as ENSO, the Indian Ocean Dipole and the surface and upper level zonal winds over the central equatorial Indian Ocean and rainfall on an interannual basis (Black *et al.,* 2003). For example, Nicholson (2015) showed that in wet years, 3 to 4 of the aforementioned factors were associated with anomalous conditions, whereas in dry years, merely 1 or 2 and sometimes even none of the factors were associated with anomalous conditions. In addition to the amount of factors, wet years showed strong linear relations with the factors, but in dry years the sign of individual factors seemed to play a role, indicative of a non-linear response to changes in these factors.

Mwale & Gan (2005) acknowledged the nonlinear and nonstationary characteristics of the data (rainfall and SST) and showed that using an artificial neural network model calibrated by a genetic algorithm (ANN-GA) outperforms a traditional linear canonical correlation model (CCA) often used in prediction studies (Pearson r 0.70-0.90 for ANN-GA versus 0.25-0.55 for CCA). Parmar *et al.* (2017) reviewed several machine learning models for rainfall prediction and concluded that the use of ANN models is preferable due to the ability of these models to 1) learn from the past and 2) account for aforementioned nonlinear relationships in rainfall data.

Forecast overview

In this study, forecasts were produced at 17 stations. Before starting to analyse the forecasts, this chapter provides a visual overview of the rainfall record, mean rainfall distribution and forecasts at the 7 best performing stations. The names, locations and elevations of these stations are presented in Figure 16 and Table 4. Most figures in this chapter are shown for station 9 specifically, but identical figures for the other stations are presented in the Appendices C-F.

Table 4 Location and elevation of the top 7 stations.

Station	Country	City	Elevation (m)
9	Kenya	Dagorett	1836
11	Kenya	Embu	1432
22	Kenya	Makindu	995
25	Kenya	Meru	1584
26	Kenya	Moyale	1112
27	Kenya	Nairobi-Wilson	560
37	Uganda	Entebbe	1146

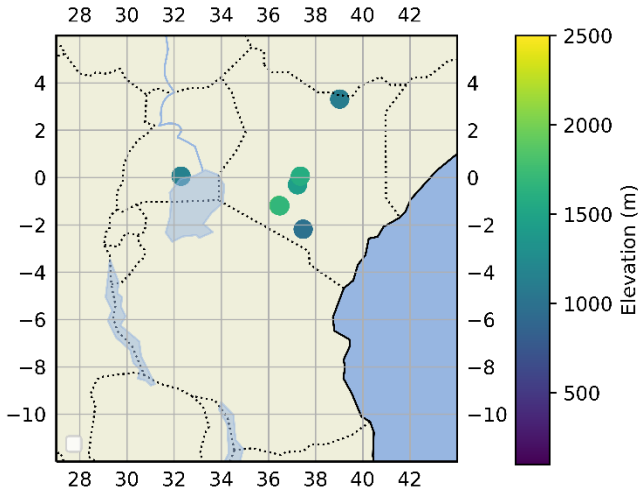


Figure 16 The locations of the 7 best performing stations (x-axis = longitude, y-axis = latitude).

Figure 17 provides an overview of the rainfall record at station 9. Some years are characterized by extreme rainfall peaks, whereas in other years rainfall is substantially less. The corresponding mean monthly rainfall observed between 1994 and 2011 is plotted in Figure 18. Consistent with the general pattern observed in East Africa, the long rainy season lasts longer and is more intense compared to the short rainy season. Similar results for the additional stations are provided in Appendix C and Appendix D. The mean rainfall and standard deviation in both seasons at the stations are presented in Appendix A. In agreement with the literature, the short rains are more variable than the long rains (standard deviation of 71 mm versus 66 mm, respectively).

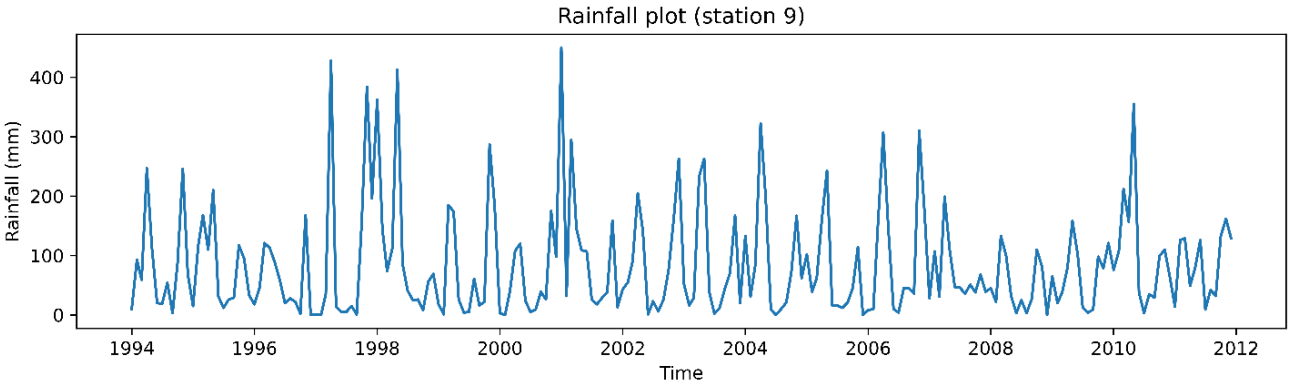


Figure 17 Observed monthly rainfall at station 9 in the period between 1994 and 2011.

The analysis of the forecasts that will be presented in the next section of the report mainly focusses on forecasts issued at a lead time of 4 months, as long-lead forecasts are especially relevant for water users near the stations. Forecasts made by SEAS5, the linear model and the LSTM model at a lead time of 4 months are presented in Figure 19. A wet bias of the SEAS5 forecasts can be recognized and both the linear model and the LSTM tend to underestimate some of the rainfall peaks throughout the record. Differences between the linear model and the LSTM are present, but no clear systematic differences can be recognized from this figure. A similar overview for the additional stations is presented in appendix E.

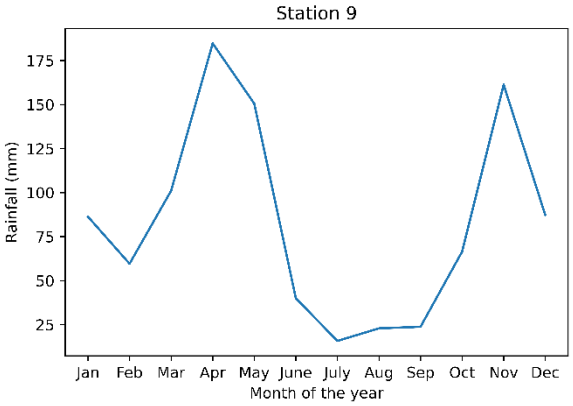


Figure 18 Mean monthly rainfall at station 9 between 1994 and 2011.

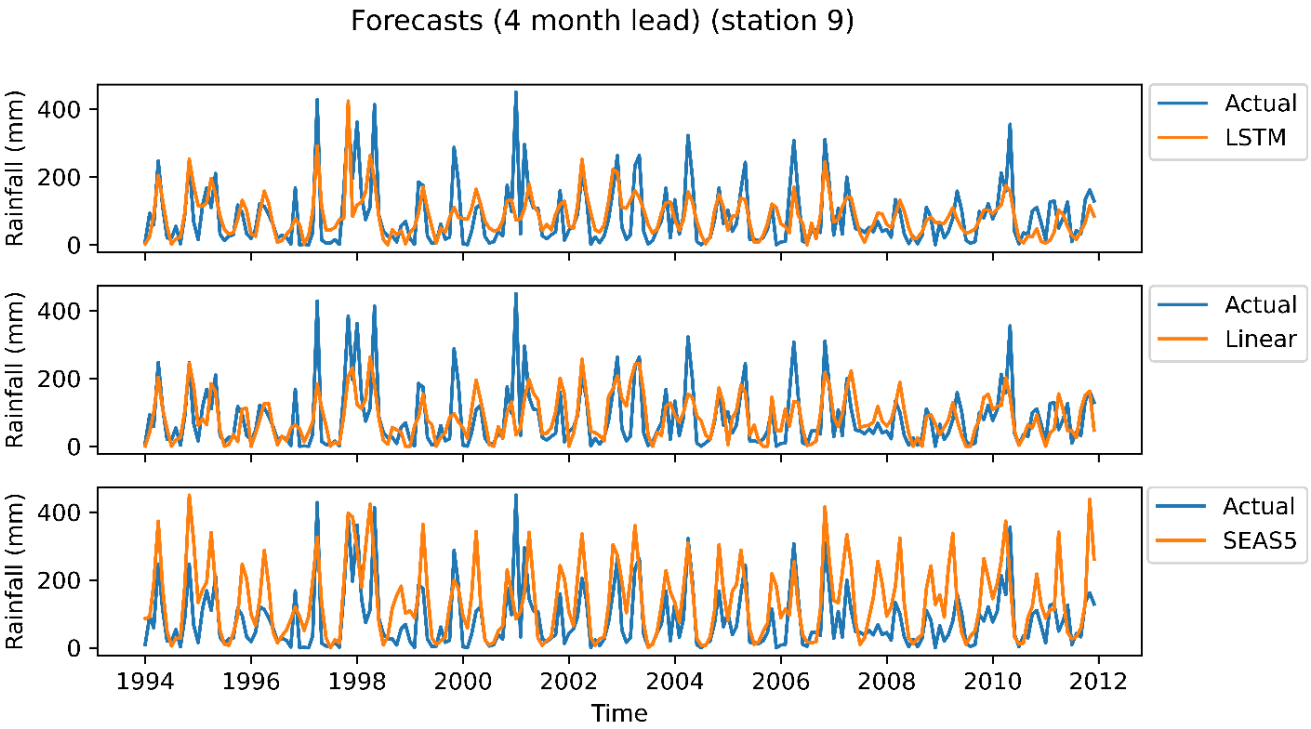


Figure 19 Forecasts of monthly rainfall by SEAS5 (bottom), the linear model (middle) and the LSTM (top) for station 9 at a lead time of 4 months.

Figure 20 & Figure 21 show the actual and forecasted total seasonal rainfall in the long rains and the short rains at station 9. The figures are sorted by the actual rainfall to provide an insight in the performance in the below-normal, normal and above-normal seasons. The most important observation is that below-normal rainfall is generally over-forecasted and above-normal rainfall is under-forecasted. This effect seems to be stronger in the long rains. Identical results for the other stations are presented in Appendix F, where similar general observations can be found.

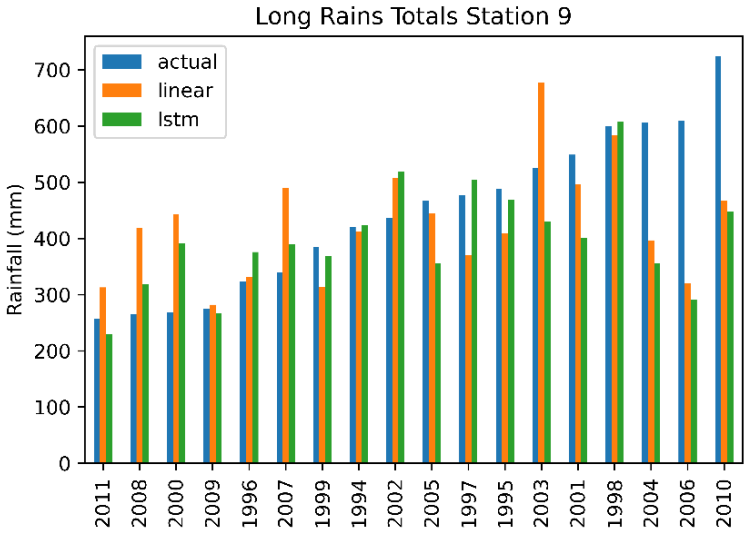


Figure 20 Long rains total rainfall at station 9, sorted by actual rainfall.

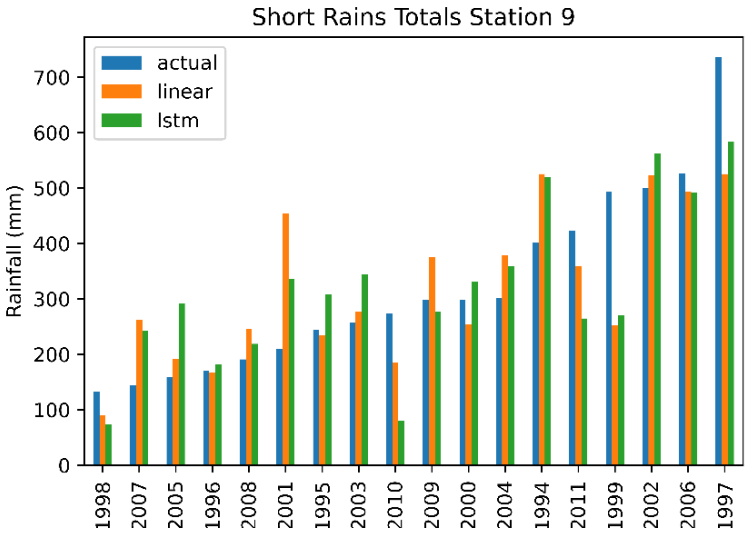


Figure 21 Short rains total rainfall at station 9, sorted by actual rainfall.

An overview of the behaviour of ENSO, DMI, MJO80 and MJO120 is provided in Appendix B. The extreme El Niño of 1997/1998 can be clearly recognized, as well as La Niña Events from 1998-2001 and 2010-2012.

Forecast verification

This chapter aims to give an extensive overview of the performance of the models to address multiple relevant aspects: differences between stations, models, lead time and the long and short rains. As mentioned before, the focus of this chapter will be on forecasts issued at a lead time of 4 months, as especially long-lead forecasts are relevant to water users near the stations. Especially relevant to water users near the stations is the ability of the model to forecast anomalous seasonal rainfall, as they may be associated with the occurrence of floods and droughts.

The metrics used are: WNDI (a weighted version of the RMSE), the anomaly correlation and several scores associated with the 2x2 contingency table.

Anomalous seasonal rainfall (4-month lead time)

To quantify the forecasts quality in the context of anomalous total seasonal rainfall, three categories of were defined: 1) below normal, 2) normal, 3) above normal (Figure 10). Based on these categories, forecasts issued at a lead time of 4 months were evaluated by means of the 2x2 contingency table.

Table 5 provides a comprehensive summary of these results in terms of the bias, hit rate, false alarms rate (F) and the Clayton Skill Score (CSS) for below-normal and above-normal seasons. The values represent the mean statistic of the scores obtained from the 7 best performing stations. Several observations can be drawn from Table 5.

First of all, the model is underconfident in forecasting anomalies. This is reflected by the hit rate. The hit rate ranges between 0.07 and 0.48, indicating that 7-48% of the times an anomaly occurs it is also forecasted. This is consistent with what can be observed visually from Figure 20 and Figure 21.

Second of all, false alarm rates are generally low (between 4% and 17%). This indicates that when the model does forecasts an anomaly, it occurs in 83-96% of the cases. A combination of a relatively high hit rate in combination with a relatively low false alarm rate, results in relatively high values of the Clayton Skill Score (CSS). The CSS is indicates positive skill when the chance of an event occurring is higher when it is forecasted, compared to when it is not forecasted.

Thirdly, the short rains generally exert higher skill, compared to the long rains. In the short rains, the linear model shows better skill when forecasting below-normal rainfall, with a hit rate of 40% and a false alarm rate of 4%. In the long rains, the LSTM model shows a substantial improvement over the linear model, especially when forecasting below-normal rainfall, with a hit rate of 42% and a false alarm rate of 8%.

The performance of the forecast by SEAS5 exerted no skill in regards to the 2x2 contingency table and was therefore not included in this analysis.

Table 5 Mean scores (Bias, Hit Rate, False Alarm Rate (F) and CSS skill score) over the best 7 stations for below-normal and above-normal conditions.

Season	Domain	Model	Bias	Hit Rate	F	CSS
Long Rains	below-normal	linear	1,04	0,28	0,17	0,17
		lstm	0,77	0,42	0,08	0,55
	above-normal	linear	0,63	0,07	0,12	-0,06
		lstm	0,4	0,12	0,05	0,29
Short Rains	below-normal	linear	0,5	0,4	0,03	0,7
		lstm	0,41	0,28	0,04	0,5
	above-normal	linear	0,6	0,45	0,04	0,67
		lstm	0,64	0,48	0,04	0,68

A detailed overview of the results for individual stations is provides in Appendix G. From this overview the variation in scores between stations can be observed. Although the mean performance of the LSTM is higher in the long rains (compared to the linear model), some stations exert very low skill for both models. The increased performance of the linear model in the short rains in below-normal seasons, is more consistent over all of the stations and the LSTM shows no skill at some of the stations.

Seasonal differences (4-month lead time)

From a more general perspective (not considering seasonal anomalies in particular), the performance differences between the seasons were assessed by means of the WNDI and the anomaly correlation. Whereas the WNDI is a measure of the error between observations and forecast, the anomaly correlation is a measure of linear correlation between the observations and forecasts with respect to the climatological mean. In this analysis, the forecasts of SEAS5 were included for comparison.

Figure 22 provides a comprehensive overview of the median scores, based on the 7 best performing stations. The wet bias of SEAS5, observed in Figure 19 is reflected here by relatively high values of WNDI. In both seasons, the linear model and LSTM show substantially lower error, with WNDI values ranging between 0.75 and 0.85. These values of WNDI correspond to an RMSE equal to roughly 6-7% of the mean annual rainfall. In terms of the anomaly correlation and with respect to SEAS5, performance is not lifted in the long rains. The anomaly correlation of the LSTM is 0.77. In the short rains, the anomaly correlation is increased by the linear model and the

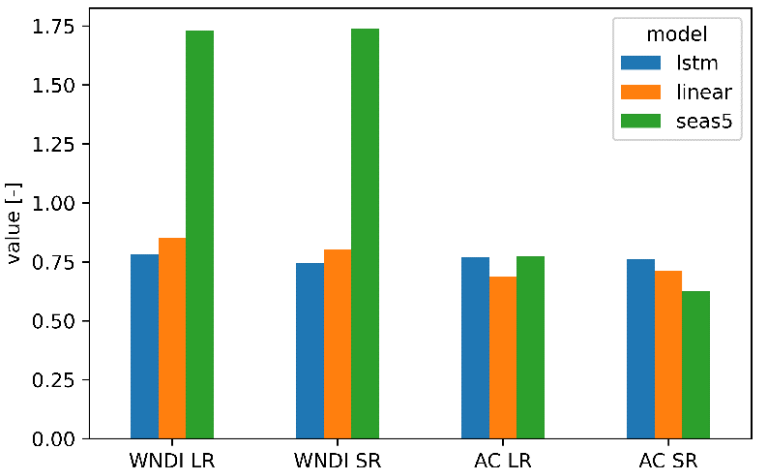


Figure 22 Median performance comparison between LSTM, Linear and SEAS5 in the long rains and short rains based on the Anomaly Correlation and WNDI at a lead time of 4 months

LSTM, with values of 0.71 and 0.76, respectively.

A detailed overview of the performance of individual stations, is presented in appendix H. the anomaly correlation seems to be relatively consistent between seasons, although at some stations the performance is relatively better in one of the seasons.

Lead time and performance

So far, the performance of the models at a lead time of 4 months has been discussed. Also the effect of lead time on performance was studied. Here, all months were considered and no differences between seasons were evaluated.

Figure 23 gives a comprehensive overview of the median performance of the different models in terms of WNDI and anomaly correlation, based on the 7 best performing stations and with respect to lead times ranging from 1 to 6 months.

The wet bias of SEA5, reflected by high values of WNDI, is also observed here. This wet bias is most pronounced at longer lead times. At a lead time of one month, the performance of SEAS5 is observed to drop quickly, in terms of both the WNDI and the anomaly correlation. Both the linear model and the LSTM are able to sustain a better performance at longer lead times, although after a lead time of 4 months, performance drops more rapidly. The LSTM generally shows slightly better performance in terms of both WNDI and LSTM.

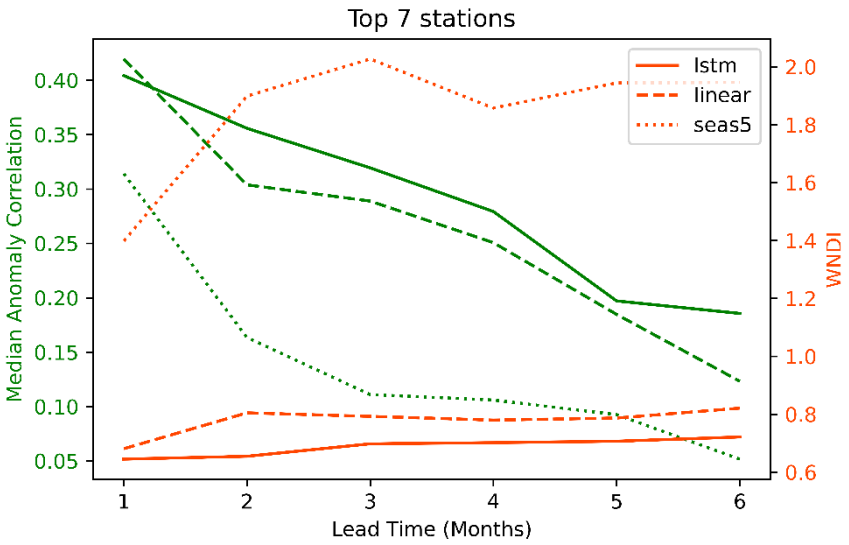


Figure 23 Performance as a function of lead time based on the top 7 stations.

Sensitivity analysis

To investigate the relative importance of forecasts of SEAS5 and the climate indices at different lead times, sensitivity analysis was carried out.

In the procedure, the LSTM model was first trained on only the rainfall lags and consequently run on the rainfall lags in combination with one feature (for example ENSO). Subsequently, the performance difference was evaluated. If performance was lifted by at least 1%, the feature was labeled to be relevant. This analysis was carried out for each feature and on all of the 17 stations and for all lead times.

Figure 24 provides an overview of the relevancy of different features at different lead times. This is expressed by counting the number of stations for which a specific feature (i.e. ENSO) increased performance. Figure 24 indicates that forecasts of SEAS5 (both precipitation and temperature) are especially relevant at a lead time of 1 month. At lead times longer than 1 month, this relevancy is observed to drop substantially, whereas the climate indices are observed to be more relevant. The combined results observed in Figure 23 and Figure 24 indicate that the sustained performance at longer lead times is due to a good coupling between the forecasts of SEAS5 and the climate indices.

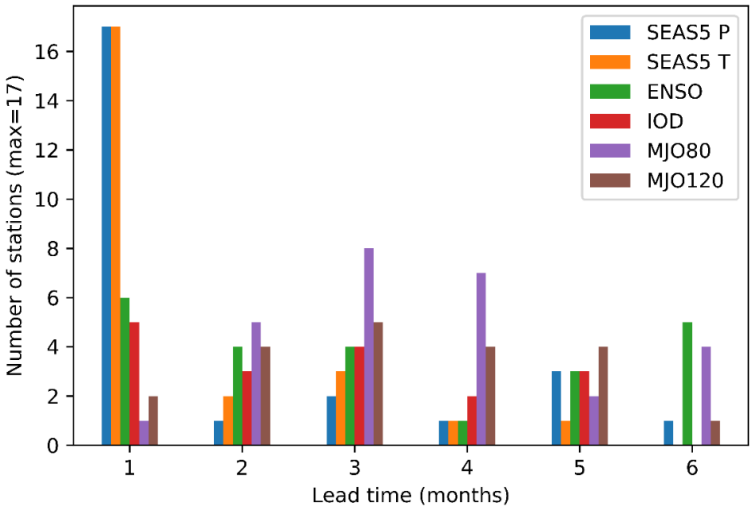


Figure 25 The number of stations at each lead time where a feature is increasing performance.

Figure 25 shows the relevancy of different climate indices for individual stations, this time by counting the lead times at which a feature is relevant. A substantial degree of variation can be observed from this figure. Some stations barely show any relevancy with any of the indices, whereas other stations show a stronger connection. For example, station 9 shows a strong connection to ENSO and MJO80. Likely due to this strong connection, performance at station 9 is observed to be one of the best in both the long rains and the short rains (Appendix G & Appendix H).

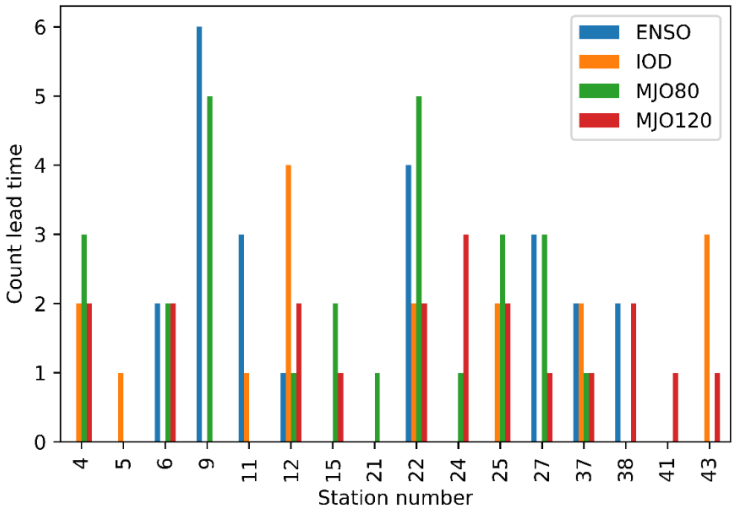


Figure 24 Count of lead times in which ENSO/ IOD or MJO was improving model performance (note: the count lead time does not correspond to the actual lead time).

The sensitivity analysis carried out above, aimed to gain an understanding in the relevancy of different features in the input dataset. Other sensitivity analyses were carried out to find the optimal lag times for different indices, to tune the hyperparameters of the LSTM and to test the robustness of the model. As these results have little physical relevance and mainly refer to the methodology, they are presented and described in Appendix I.

Discussion

Methodology and data

Data quantity and quality

In every machine learning problem, a sufficient amount of data has to be available for training and testing. When too little data is available, it will be difficult to reach an optimum with regard to the bias-variance trade-off problem. Essential patterns within the data might not be captured well enough and cause a high bias, or the model might not generalize well to unseen data and result in a high variance model. Therefore, the data quantity is an essential factor for a successful model.

The hindcasts as generated by the ECMWF SEAS5 model range from 1993 onward, setting the lower temporal limit of this study. The gauge data set the upper limit to the temporal range of this study, since rainfall data was only available until 2012. The original range of the data was therefore 1993-2011. The forecasts range was 1994-2011, due to the inclusion of lead and lag times. Thus, a total temporal coverage of 18 years was achieved. A similar study applying an ANN to a seasonal rainfall forecasting by Abbot and Marohasy (2012) used a comparable time period of data ranging from 1993 to 2009. Compared to a linear model, ANNs like LSTMs require more data due to their complexity (Brownlee, 2018). In this study, the LSTM generally outperformed the linear model, but differences were small. However, due to the limited data availability of 18 years, the LSTM might not have reached its full potential (over a linear model) for seasonal forecasts and performance is likely to increase when more data becomes available in the future.

Another aspect is the data quality. In this study the personally assembled gauge dataset of Professor Sharon S. Nicholson (Florida State University) was used and included data derived from meteorological and agricultural bulletins. This personally assembled dataset may have been an effort to account for the decline in available gauge data in East Africa since the 1980s (Camberlin & Phillipon, 2002). The quality of this data cannot be assessed but it is assumed to be a potential source of variability in the study results.

Seasonality, long-term trends and nonstationary relationships

Several pre-processing techniques were applied to modify the data set into a usable format for both models. No seasonal trends were removed from the rainfall data before training, based on the knowledge that the LSTM is able to learn the seasonal patterns. However, performance might have been lifted slightly (Géron, 2019).

Woodman (2020) performed a trend analysis of SPI06 in East Africa in a time period from 1980 to 2014. Significant drying trends are found in West-Uganda, Tanzania and South-Kenya. Significant wetting trends were found in Western Kenya (Figure 27).

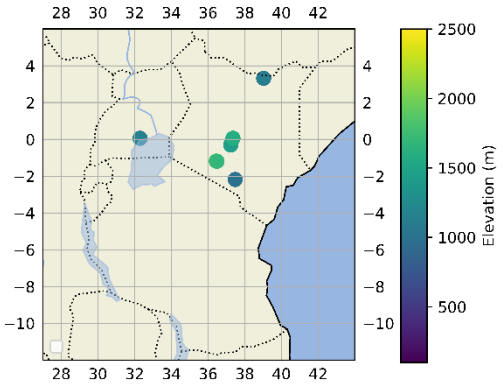


Figure 27 Geographical distribution of the top 7 stations (at a lead time of 4 months).

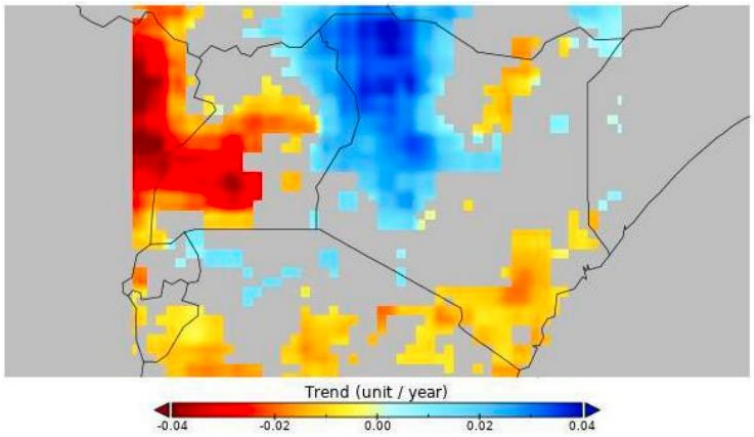


Figure 26 Drying and wetting trends in Equatorial East Africa (SPI06, 75% confidence level). Adopted from: Woodman (2020).

For the models, the presence of long-term trends may be problematic and the optimal methodology would include detrending the data before training. However, comparing Figure 26 and Figure 27, the best forecasts were made mostly in areas where no significant trend was present. Trend analysis was not a part of this study, but might be useful for the detection of trends at the stations and for detrending the data before training.

The non-stationarity between ENSO and IOD on the short rains is considered to be an important point of discussion. In the period between 1993 and 2011 under consideration in this study, the strength of the teleconnections between ENSO and IOD and the short rains was overall significant and relatively stationary, but lower before 1997 and lowest in 1995 (Figure 15). In the time range considered in this study, the varying strength of the teleconnections may have had an effect on the performance of the models. With regard to operational use of the models, the potential occurrence of a regime shift (as indicated in Figure 15) is likely to be detrimental to the quality of the forecasts.

Direct Multi-Step Forecasting

In this study a direct approach was used to make forecasts at multiple lead times, also referred to as direct multi-step forecasting. In this approach, a separate model is created for every lead time. Implicit to this methodology is the fact that all models act independently of each other and dependencies between lead times are not modelled. Alternative methods include recursive multi-step forecasting and multi-output strategies. In the recursive approach, one model is used multiple times and the prediction at a prior timestep is used in the input dataset for the prediction at a following time-step. In this way, the dependency is modelled but the forecast quality might quickly degrade when prediction errors accumulate (Brownlee, 2018). In the multi-output strategy (only applicable to LSTMs), the LSTM can be configured to predict not just one value one step ahead, but multiple values multiple steps ahead. Due to time constraints in this study, the multi-step approach could not be implemented but performance is expected to improve at longer lead times using a multi-step method.

Sensitivity analysis

The method applied to test the relative importance of individual features in the input dataset was relatively simple. In the procedure, the model was run on the rainfall lags and one of the inputs and the performance difference is evaluated. As this procedure only tests the importance of one feature at a time, it is not able to provide any information on the importance of interrelations between multiple inputs. This has been witnessed during the analysis at one station, where none of the individual features improved performance, but including all of the features lifted performance by 10%. The presence of some of these interrelations is known to exist. For example, during an El Niño, extreme events in the long and the short rains have been related to the phase of the MJO, increasing convection over eastern Africa (Pohl and Camberlin, 2011). From the literature, no alternative methods have been found for feature importance assessment for LSTM networks. However, if a method would be available to provide information on the interrelations between features it could potentially provide useful information to better understand these interrelations.

Performance evaluation

The quality of the forecasts produced in this study was assessed by means of the WNDI, anomaly correlation and several scores associated with the 2x2 contingency table, all of which aim to address different aspects of the forecast quality. First, the effect of lead time on performance will be addressed. Then, the differences between de seasons will be discussed in relation to the forecasts at a lead time of 4 months. Finally, the value of the models developed in this study for operational use will be discussed.

Lead time and performance

With respect to the performance of the models at different lead times, both the linear model and the LSTM have proven to be capable of increasing performance relative to the SEAS5 system and climatology. The performance of SEAS5 was shown to drop substantially after a lead time of 1 month, whereas the linear model and the LSTM are able to sustain a better performance at longer lead times. From the sensitivity analysis it was shown that the climate indices remain relevant, also at longer lead times. As a result, the increased performance at longer lead times is likely due to a good coupling by the linear model and the LSTM between the climate indices and the rainfall observed at the stations.

Differences between the seasons

Consistent with what is observed in the literature, the performance of the models was generally better in the short rains. Although the mean values of WNDI and anomaly correlation over the 7 best stations indicated rather similar performance in the seasons, the results obtained from the 2x2 contingency table indicate a substantial improvement in the short rains, when forecasting anomalous seasonal rainfall.

For the LSTM, mean values of WNDI and anomaly correlation were observed to range between 0.75-0.85 (equal to an RMSE of 6-7% of the mean annual rainfall) and 0.76-0.77 in both seasons. The linear model showed slightly worse values, although it is not clear whether or not these differences are statistically significant. With respect to the SEAS5 system, no improvement was made in the long rains in terms of the anomaly correlation.

With respect to quality of the forecasts in the context of anomalous total seasonal rainfall, the LSTM model showed a substantial improvement over the linear model in the long rains, especially when forecasting below-normal rainfall. Whereas the linear model exerted minimal to negative skill, the

mean hit rate of the LSTM was observed to be 42%, with a false alarm rate of 8% and a Clayton Skill Score of 0.55, in the long rains and in below-normal seasons. The most extensive study carried out in regard to the long rains in East Africa between 1968 and 1997, was carried out by Camberlin & Phillipon (2002). Their model predicted seasonal anomalies 70% of the time and with a correlation of 0.66, based on February predictors related to SSTs and atmospheric variables. The increased performance compared to this study may be due to the more extensive input dataset used specifically for the long rains, a shorter lead time and different time period.

The best scores were observed in the short rains, with hit rates between 40-50%, false alarm rates of 4% and Clayton Skill Scores of around 0.7. With regard to the below-normal seasons in the short-rains, the linear model performed better than the LSTM. Nicholson (2014) produced forecasts of October and November rainfall in the equatorial east African region for the period 1950-2005. She found a correlation of 0.80 between observed and predicted values. However, it is unclear how well the model forecasted anomalous rainfall over the region.

As known from the literature, ENSO and IOD primarily modulate the short rains. For this reason, the performance of the models is generally better in the short rains, compared to the long rains. However, performance in the short rains is still suboptimal, since a significant portion (50-60%) of the anomalous events is not captured accurately. Although ENSO and IOD play an important role in modulating the short rains, their effect is not always consistent. This was also found in the literature, where Nicholson & Kim (1997) showed that anomalously high rainfall was observed in 12/20 of the El Niño events, an anomalously low in 12/17 of the La Niña events.

During the long rains, the state of the ocean is rapidly changing and the ESNO index switches its sign. For this reason, the predictability of ENSO is limited to 2 months in the long rains, also known as the spring predictability barrier. Therefore, ENSO and IOD are limited sources of predictability during the long rains. The MJO, however, is shown to be an important source of rainfall variability in the long rains (Camberlin & Phillipon, 2002). This was also found in this study, for example at station 9, where the sensitivity analysis showed a strong association with the MJO80 index. As a result of this association, the performance in the long rains at this station was lifted substantially. However, the influence of any of the indices on rainfall was shown to vary between stations. Some stations showed a high degree of association, whereas at other station association was limited and other factors may play a more important role in the modulation of the rainfall (Figure 13).

Value of the forecasts for operational use

The scores associated with the 2x2 contingency table are presumably the most relevant when assessing the value of the forecasts for operational use. Especially the hit rate and the false alarm rate are intuitive to understand as they provide a straightforward measure of forecast quality. One of the main limitations of the models developed in this study is that they are observed to be underconfident when forecasting seasonal anomalies. In other words, a substantial amount (roughly 50-60%) of extreme events is not accurately forecasted by the models, on average. For this reason, the models developed in this study are considered to be not sophisticated enough for operational use. However, the observed false alarm rates were generally low (between 4-8%), indicating that in most cases, when the model forecasts an anomaly, it is also likely to occur. Therefore, the models may still provide valuable information to water-users near the stations, as the forecasts are relatively reliable.

Recommendations

Forecasting rainfall in eastern Africa remains a difficult task as processes driving rainfall (especially in the long rains) are inadequately understood (Nicholson, 2017). Several suggestions will be made that may increase the performance of the model developed in this study.

From SEAS5, only forecasts of temperature and precipitation were used in this study. However, forecasts of sea surface temperatures and sea level pressures are also available from the SEAS5 system. By implementing these variables, forecasts of ENSO and IOD can be constructed and used as input into the model. By doing this, the model not only has access to the observed values of ENSO and IOD in the past but also to forecasts of ENSO and IOD in the future. In addition to using more variables in the input dataset, including other forecasting systems may also be explored.

In addition to the use of sea surface temperature and pressure and the associated indices, the use of atmospheric variables such as zonal and meridional winds at different locations has been proven to be more successful than the use of sea surface temperature pressure in increasing forecasting skill in the short rainy season, especially at shorter lead times (Nicholson, 2014). For the long rains, Camberlin & Phillipon (2002) used February predictors of ENSO, an energy gradient from the East African highlands, geopotential height anomalies over the Near-East, and westerly winds from the Congo basin. Unfortunately, the study of Camberlin & Phillipon (2002) provides little value with respect to operational use due to the limited lead time.

Although multiple suggestions are made for the improvement of the forecasting model developed in this study, the lack of understanding of the processes driving rainfall in the eastern African region remain a limiting factor, especially in the long rains. Further investigation into the driving processes is required for the improvement of seasonal forecasts.

Conclusion

In this study, forecasts of monthly totals of rainfall at 17 rainfall stations in the equatorial east African region between 1994 and 2011 were produced, based on climate indices of ENSO, IOD and MJO and seasonal forecasts of temperature and precipitation from the ECMWF SEAS5 seasonal forecasting system. A long short-term memory (LSTM) model was developed to forecast rainfall at each individual station at lead times ranging from 1-6 months. The performance was evaluated with respect to the long rainy season and the short rainy season and with respect to different lead times. Benchmarks for performance evaluation included a simple multi-variate linear model, the precipitation forecasts of SEAS5 and climatology. Performance metrics included the anomaly correlation and the WNDI (a weighted index of RMSE to allow comparison between stations). The relevance of individual features in the input dataset were also evaluated between stations and lead times.

For all lead times, the LSTM outperformed both the linear model and SEAS5 in terms of both anomaly correlation and WNDI, although it is not clear whether or not this is statistically significant. Especially at longer lead times the LSTM shows improved performance, due to a good coupling with the climate indices. Both the LSTM and the linear model prove to be successful at correcting the wet bias of SEAS5 that is observed at all stations, as the WNDI generally improved with about 200%.

The main portion of this study focused on forecasts at a lead time of 4 months at the 7 best performing stations. In terms of WNDI and the anomaly correlation, differences between the long rains and the short rains were small. The LSTM model showed slightly better performance with a WNDI of 0.75-0.85 (equal to an RMSE of 6-7% of the mean annual rainfall) and an anomaly correlation of 0.76-0.77. In the context of anomalous seasonal rainfall, the performance in the short rains was substantially better, on average, with hit rates between 40-50%, false alarm rates of 4% and Clayton Skill scores of roughly 0.7. The linear model performed better at forecasting below-normal rainfall. The LSTM showed to be more valuable in the long rains, especially when forecasting below-normal rainfall, with a hit rate of 42%, false alarm rate of 8% and a Clayton Skill Score of 0.55, on average.

With regard to operational use, especially forecasts of anomalous rainfall are of interest, as they may be associated with the occurrence of floods and droughts. The models developed in this study are underconfident in forecasting these anomalies and are therefore not sophisticated enough for operational use. However, due to the observed low false alarm rates, they may still provide valuable information to farmers and water managers near the stations.

As the drivers for both interannual and intraseasonal variability of the rainy seasons in equatorial east Africa are still inadequately understood and since the quality of seasonal forecasts are rooted in this understanding, further research is necessary to improve seasonal forecasts in the future.

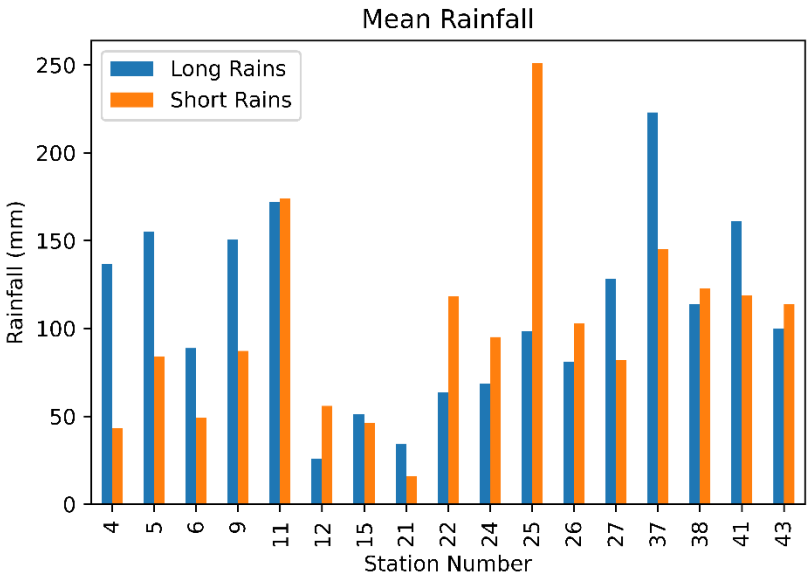
References

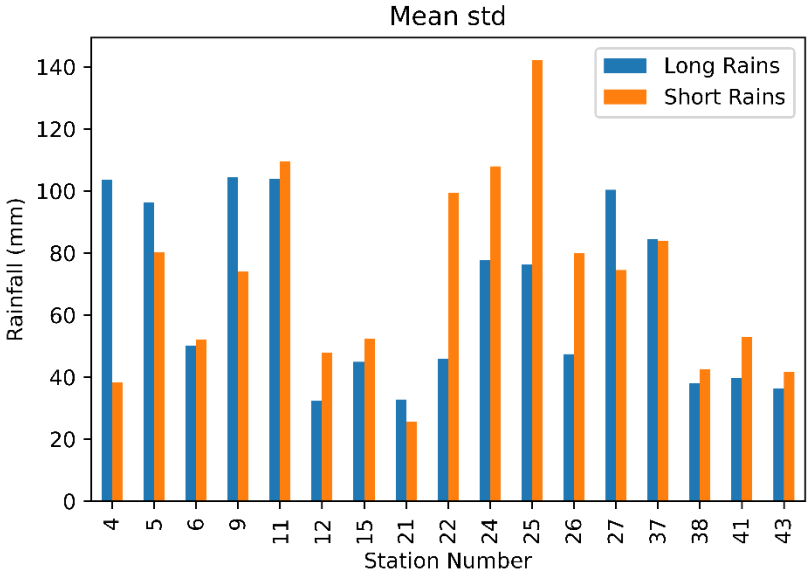
1. Abadi, Martín, et al. "Tensorflow: A system for large-scale machine learning." *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*. 2016.
2. Abbot, John, and Jennifer Marohasy. "Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks." Atmospheric Research 138 (2014): 166-178.
3. Bergonzini, Laurent, et al. "Zonal circulations over the Indian and Pacific Oceans and the level of Lakes Victoria and Tanganyika." International Journal of Climatology: A Journal of the Royal Meteorological Society 24.13 (2004): 1613-1624.
4. Berhane, Fisseha, and Benjamin Zaitchik. "Modulation of daily precipitation over East Africa by the Madden–Julian oscillation." Journal of climate 27.15 (2014): 6016-6034.
5. Black, E., J. M. Slingo, and K. R. Sperber (2003), An observational study of the relationship between excessively strong short rains in coastal East Africa and Indian Ocean SST, Mon. Weather Rev., 131(1), 74–94, doi:10.1175/1520-0493(2003)1312.0.CO;2.
6. Brownlee, Jason. Deep learning for time series forecasting: Predict the future with MLPs, CNNs and LSTMs in Python. Machine Learning Mastery, 2018.
7. Camberlin, Pierre, and Nathalie Philippon. "The East African March–May rainy season: Associated atmospheric dynamics and predictability over the 1968–97 period." Journal of Climate 15.9 (2002): 1002-1019.
8. Camberlin, Pierre, and Olivier Planchon. "Coastal precipitation regimes in Kenya." Geografiska Annaler: Series A, Physical Geography 79.1-2 (1997): 109-119.
9. Chen, Chia-Jeng, and Aris P. Georgakakos. "Seasonal prediction of East African rainfall." International Journal of Climatology 35.10 (2015): 2698-2723.
10. Clayton, H. Helm. "A method of verifying weather forecasts." *Bulletin of the American Meteorological Society* 8.10 (1927): 144-146.
11. Cohen, Judah, et al. "S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts." Wiley Interdisciplinary Reviews: Climate Change 10.2 (2019): e00567.
12. Den Daas, Jan. "Drought analysis in Tanzania using Markov chains." Student Undergraduate Research E-journal! 2 (2016).
13. Diro, G. T., D. I. F. Grimes, and E. Black (2011b), Teleconnections between Ethiopian summer rainfall and sea surface temperature: Part II. Seasonal forecasting, Clim. Dyn., 37(1), 121–131, doi:10.1007/s00382-010-0896-x
14. Dracup, John A., Kil Seong Lee, and Edwin G. Paulson Jr. "On the definition of droughts." Water resources research 16.2 (1980): 297-302.
15. Dutra, Emanuel, et al. "The 2010–2011 drought in the Horn of Africa in ECMWF reanalysis and seasonal forecast products." International Journal of Climatology 33.7 (2013): 1720-1729.
16. EM-DAT, 2020. The Emergency Events Database – Université catholique de Louvain (UCL) – CRED, D. Guha-Sapir. www.emdat.be (Brussels, Belgium retrieved at 2 October 2020).
17. FAO, 2014. Adapting to Climate Change through Land and Water Management in Eastern Africa Results of Pilot Projects in Ethiopia. (Kenya and Tanzania, Rome, Italy).
18. Funk, Chris, et al. "Warming of the Indian Ocean threatens eastern and southern African food security but could be mitigated by agricultural development." Proceedings of the national academy of sciences 105.32 (2008): 11081-11086.
19. Gebrechorkos, Solomon, Stephan Hülsmann, and Christian Bernhofer. "Evaluation of multiple climate data sources for managing environmental resources in East Africa." *Hydrology and Earth System Sciences* 22.8 (2018): 4547-4564.
20. Géron, Aurélien. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media, 2019.
21. Goddard, Lisa, and Nicholas E. Graham. "Importance of the Indian Ocean for simulating rainfall anomalies over eastern and southern Africa." Journal of Geophysical Research: Atmospheres 104.D16 (1999): 19099-19116.
22. Hagedorn, Renate, Francisco J. Doblas-Reyes, and Tim N. Palmer. "The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept." Tellus A: Dynamic Meteorology and Oceanography 57.3 (2005): 219-233.
23. Hansen, James W., et al. "Review of seasonal climate forecasting for agriculture in sub-Saharan Africa." (2011): 205-240.
24. Hao, Zengchao, Vijay P. Singh, and Youlong Xia. "Seasonal drought prediction: advances, challenges, and future prospects." Reviews of Geophysics 56.1 (2018): 108-141.
25. Hastenrath, S. (2007), Circulation mechanisms of climate anomalies in East Africa and the equatorial Indian Ocean, Dyn. Atmos. Ocean.,
26. Hastenrath, S., D. Polzin, and C. Mutai (2010), Diagnosing the droughts and floods in equatorial East Africa during boreal autumn 2005–08,
27. Hendon, Harry H., and Murry L. Salby. "The life cycle of the Madden–Julian oscillation." Journal of the Atmospheric Sciences 51.15 (1994): 2225-2237.
28. Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.
29. Hoell, Andrew, and Chris Funk. "The ENSO-related west Pacific sea surface temperature gradient." Journal of climate 26.23 (2013): 9545-9562.
30. Huschke, Ralph E. "Glossary of meteorology." (1959).
31. Johns, Tim C., et al. "The new Hadley Centre climate model (HadGEM1): Evaluation of coupled simulations." Journal of climate 19.7 (2006): 1327-1353.
32. Johnson, Stephanie J., et al. "SEAS5: The new ECMWF seasonal forecast system." *Geoscientific Model Development* 12.3 (2019).
33. Keijzer, T. "Drought Analysis of the Lake Manyara Catchment: Metorological Drought Occurrence, Influence of Atmospheric Teleconnections and Impact on Lake Manyara." MSc Thesis Utrecht University (2020).

34. Kijazi, A. L., and C. J. C. Reason. "Relationships between intraseasonal rainfall variability of coastal Tanzania and ENSO." Theoretical and applied climatology 82.3-4 (2005): 153-176.
35. Kiladis, George N., and Henry F. Diaz. "Global climatic anomalies associated with extremes in the Southern Oscillation." Journal of Climate 2.9 (1989): 1069-1090.
36. Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).
37. Kirtman, Ben, and Anna Pirani. "The state of the art of seasonal prediction: Outcomes and recommendations from the First World Climate Research Program Workshop on Seasonal Prediction." Bulletin of the American Meteorological Society 90.4 (2009): 455-458.
38. Krishnamurti, T. N., et al. "Improved weather and seasonal climate forecasts from multimodel superensemble." Science 285.5433 (1999): 1548-1550.
39. Lau, K. M., and S. Yang. "Walker circulation." Encyclopedia of atmospheric sciences (2003): 2505-2510.
40. Linsley Jr, Ray K., Max Adam Kohler, and Joseph LH Paulhus. "Hydrology for engineers." (1975).
41. Liwenga, Emma T., Lukas Kwezi, and Tamer Afifi. "Rainfall, food security, and human mobility: case study: Tanzania." (2012).
42. Lloyd-Hughes, Benjamin. "The impracticality of a universal drought definition." Theoretical and Applied Climatology 117.3-4 (2014): 607-611.
43. Lorenz, Edward N. "Deterministic nonperiodic flow." Journal of the atmospheric sciences 20.2 (1963): 130-141.
44. Lyon, Bradfield. "Seasonal drought in the Greater Horn of Africa and its recent increase during the March–May long rains." Journal of Climate 27.21 (2014): 7953-7975.
45. Madden, Roland A., and Paul R. Julian. "Observations of the 40–50-day tropical oscillation—A review." Monthly weather review 122.5 (1994): 814-837.
46. McKee, Thomas B., Nolan J. Doesken, and John Kleist. "The relationship of drought frequency and duration to time scales." *Proceedings of the 8th Conference on Applied Climatology*. Vol. 17. No. 22. 1993.
47. Miller, R. L., 1996: Intertropical convergence zone. Encyclopedia of Weather and Climate, S. H. Schneider, Ed., Vol. 1, Oxford University Press, 445–448.
48. Mölders, Nicole, and Gerhard Kramm. Lectures in meteorology. Cham/Heidelberg/New York/Dordrecht/London: Springer, 2014.
49. Mondiale, Banque. "World development report 2008: Agriculture for development." (2008).
50. Moron, Vincent, Pierre Camberlin, and Andrew W. Robertson. "Extracting subseasonal scenarios: an alternative method to analyze seasonal predictability of regional-scale tropical rainfall." *Journal of climate* 26.8 (2013): 2580-2600.
51. Moussiopoulos, N., and Th Flassak. "Two vectorized algorithms for the effective calculation of mass-consistent flow fields." Journal of climate and applied meteorology 25.6 (1986): 847-857.
52. Mwale, Davison, and Thian Yew Gan. "Wavelet analysis of variability, teleconnectivity, and predictability of the September–November East African rainfall." Journal of applied meteorology 44.2 (2005): 256-269.
53. Mwangi, E., F. Wetterhall, E. Dutra, F. Di Giuseppe, and F. Pappenberger (2014), Forecasting droughts in East Africa, Hydrol. Earth Syst. Sci., 18(2), 611–620, doi:10.5194/hess-18-611-2014.
54. Nayak, Deepak Ranjan, Amitav Mahapatra, and Pranati Mishra. "A survey on rainfall prediction using artificial neural network." International Journal of Computer Applications 72.16 (2013).
55. Nicholson, S. E. (2015a), Long-term variability of the East African "short rains" and its links to large-scale factors, Int. J. Climatol., 35(13).
56. Nicholson, S. E., and D. Entekhabi (1986), The quasi-periodic behavior of rainfall variability in Africa and its relationship to the Southern
57. Nicholson, Sharon E. "Climate and climatic variability of rainfall over eastern Africa." Reviews of Geophysics 55.3 (2017): 590-635.
58. Nicholson, Sharon E., and Jeeyoung Kim. "The relationship of the El Niño–Southern oscillation to African rainfall." *International Journal of Climatology: A Journal of the Royal Meteorological Society* 17.2 (1997): 117-135.
59. Nicholson, Sharon E. "The ITCZ and the seasonal cycle over equatorial Africa." Bulletin of the American Meteorological Society 99.2 (2018): 337-348.
60. Nicholson, Sharon E. "The nature of rainfall variability over Africa on time scales of decades to millenia." Global and planetary change 26.1-3 (2000): 137-158.
61. Nicholson, Sharon E., and Dara Entekhabi. "The quasi-periodic behavior of rainfall variability in Africa and its relationship to the Southern Oscillation." Archives for meteorology, geophysics, and bioclimatology, Series A 34.3-4 (1986): 311-348.
62. Ogallo, L. J. "Relationships between seasonal rainfall in East Africa and the Southern Oscillation." Journal of Climatology 8.1 (1988): 31-43.
63. Ogallo, Laban J., and Janowiak JE. "Teleconnection between seasonal rainfall over East Africa and global sea surface temperature anomalies." Journal of the Meteorological Society of Japan. Ser. II 66.6 (1988): 807-822.
64. Oliphant, Travis E. *A guide to NumPy*. Vol. 1. USA: Trelgol Publishing, 2006.
65. Oscillation, Arch. Meteorol. Geophys. Bioclimatol. Ser. A, 34(3–4), 311–348, doi:10.1007/BF02257765
66. Parmar, Aakash, Kinjal Mistree, and Mithila Sompura. "Machine learning techniques for rainfall prediction: A Review." International Conference on Innovations in information Embedded and Communication Systems. 2017.
67. Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." *the Journal of machine Learning research* 12 (2011): 2825-2830.
68. Persson, Anders. "User guide to ECMWF forecast products." (2001).
69. Pohl, Benjamin, and Pierre Camberlin. "Intraseasonal and interannual zonal circulations over the equatorial Indian Ocean." Theoretical and applied climatology 104.1-2 (2011): 175-191.
70. Python Software Foundation. Python Language Reference, version 3.7 (2020) Available at http://www.python.org
71. Risbey, James S., et al. "On the remote drivers of rainfall variability in Australia." Monthly Weather Review 137.10 (2009): 3233-3253.

72. Rockström, Johan, et al. "Managing water in rainfed agriculture—The need for a paradigm shift." Agricultural Water Management 97.4 (2010): 543-550.
73. Rodhe, H., and H. Virji. "Trends and periodicities in East African rainfall data." Monthly Weather Review 104.3 (1976): 307-315.
74. Ropelewski, Chester F., and Michael S. Halpert. "North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO)." Monthly Weather Review 114.12 (1986): 2352-2362.
75. Saji, N. H., et al. "A dipole mode in the tropical Indian Ocean." Nature 401.6751 (1999): 360-363.
76. Sarle, Warren S. "Neural networks and statistical models." (1994).
77. Segele, Z. T., P. J. Lamb, and L. M. Leslie (2009a), Large-scale atmospheric circulation and global sea surface temperature associations with horn of Africa June–September rainfall, Int. J. Climatol., 29(8), 1075–1100, doi:10.1002/joc.1751.
78. Sheffield, Justin, et al. "A drought monitoring and forecasting system for sub-Sahara African water resources and food security." Bulletin of the American Meteorological Society 95.6 (2014): 861-882.
79. Sikder, Safat, et al. "Are general circulation models ready for operational streamflow forecasting for water management in the Ganges and Brahmaputra River basins?." Journal of Hydrometeorology 17.1 (2016): 195-210.
80. Sivakumar, Mannava VK, et al. "High level meeting on national drought policy: summary and major outcomes." Weather and climate Extremes 3 (2014): 126-132.
81. Svoboda, M., M. Hayes, and D. Wood. "Standardized precipitation index user guide." *World Meteorological Organization Geneva, Switzerland* (2012).
82. US Geological Survey. "Digital elevation models, data users guide 5." *Reston, Virginia, US Geological Survey* (1993): 1-50.
83. Van Aalst, Maarten, Molly Hellmuth, and Daniele Ponzi. Working Paper 89-Come Rain or Shine-Integrating Climate Risk Management into African Development Bank Operations. No. 224. 2007.
84. Van Loon, Anne F. "Hydrological drought explained." *Wiley Interdisciplinary Reviews: Water* 2.4 (2015): 359-392.
85. Van Loon, Anne F., et al. "Drought in the Anthropocene." Nature Geoscience 9.2 (2016): 89.
86. Wanders, Niko, and Eric F. Wood. "Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations." Environmental Research Letters 11.9 (2016): 094007.
87. Wang, Zhi-liang, and Hui-hua Sheng. "Rainfall prediction using generalized regression neural network: case study Zhengzhou." 2010 International conference on computational and information sciences. IEEE, 2010.
88. Webster, Peter J., et al. "Coupled ocean–atmosphere dynamics in the Indian Ocean during 1997–98." Nature 401.6751 (1999): 356-360.
89. Wilks, Daniel S. Statistical methods in the atmospheric sciences. Vol. 100. Academic press, 2006.
90. Willmott, Cort J., et al. "Statistics for the evaluation and comparison of models." Journal of Geophysical Research: Oceans 90.C5 (1985): 8995-9005.
91. Woodman, B. "Spatio-temporal drought variability in Equatorial East Africa from 1980 to 2014" MSc Thesis University of Amsterdam (2020).
92. Zhou, Lei, et al. "A Central Indian Ocean mode and heavy precipitation during the Indian summer monsoon." Journal of Climate 30.6 (2017): 2055-2067.
93. Zorita, Eduardo, and Faustine F. Tilya. "Rainfall variability in Northern Tanzania in the March-May season (long rains) and its links to large-scale climate forcing." Climate Research 20.1 (2002): 31-40.

Appendix A: Basic statistics at all rainfall stations

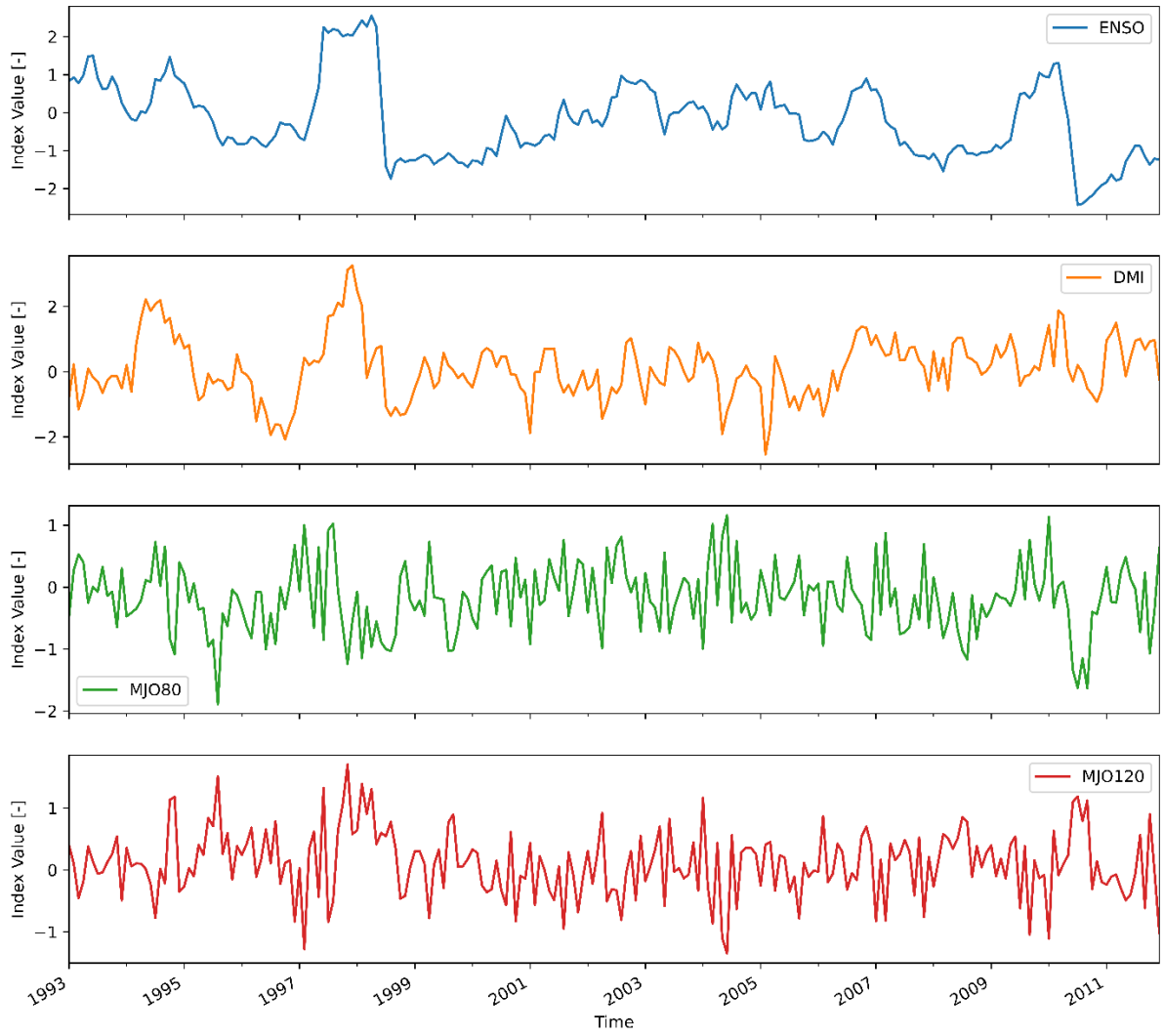




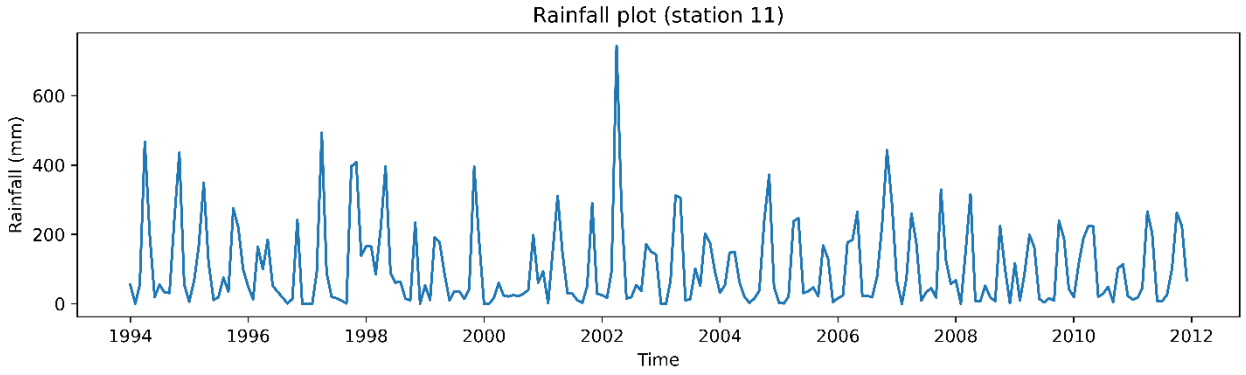
Mean statistics over all stations:

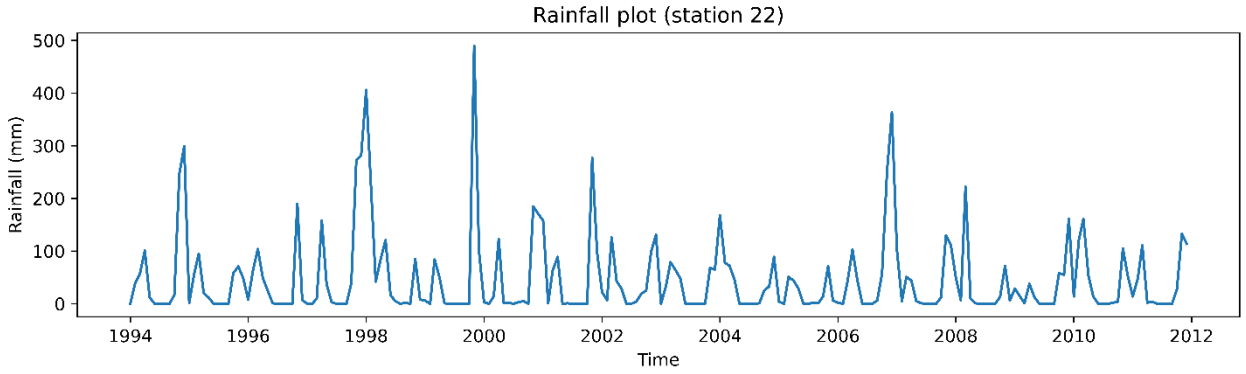
	Long Rains	Short Rains
Mean rainfall (mm)	108	100
Mean std (mm)	66	71

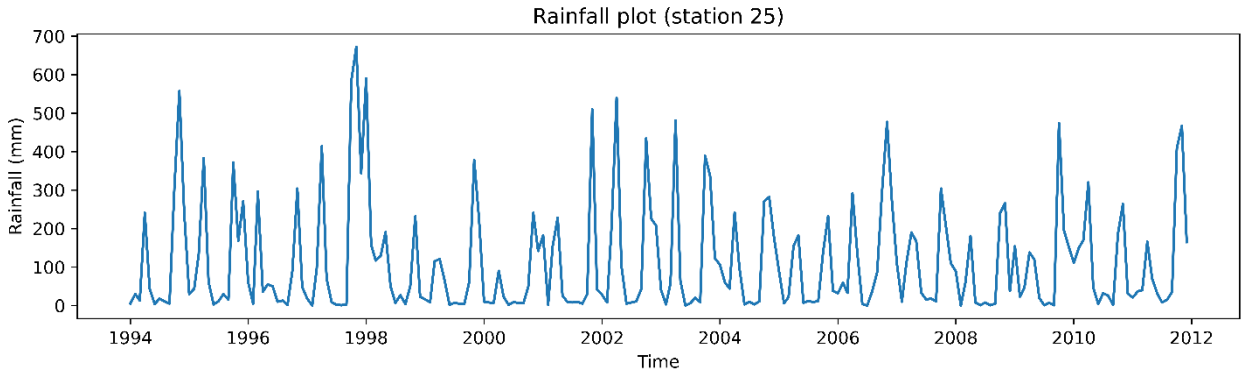
Appendix B: Overview climate indices

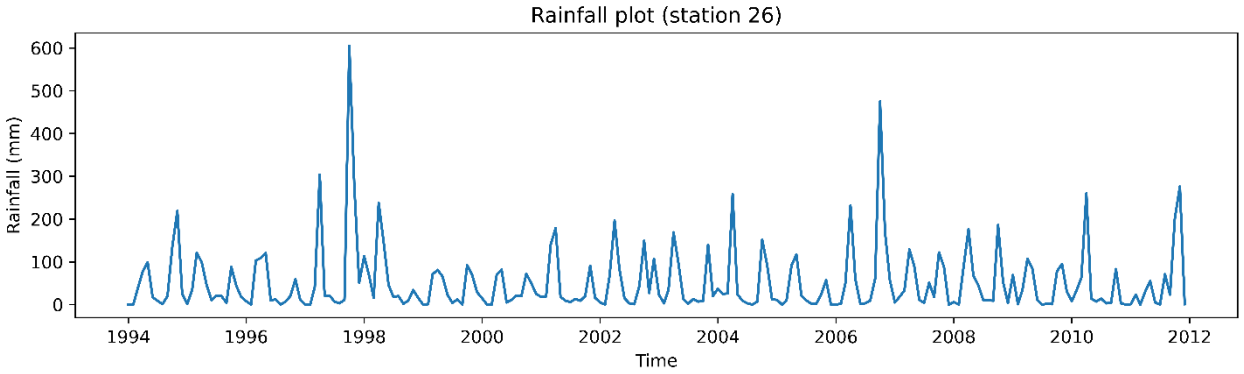


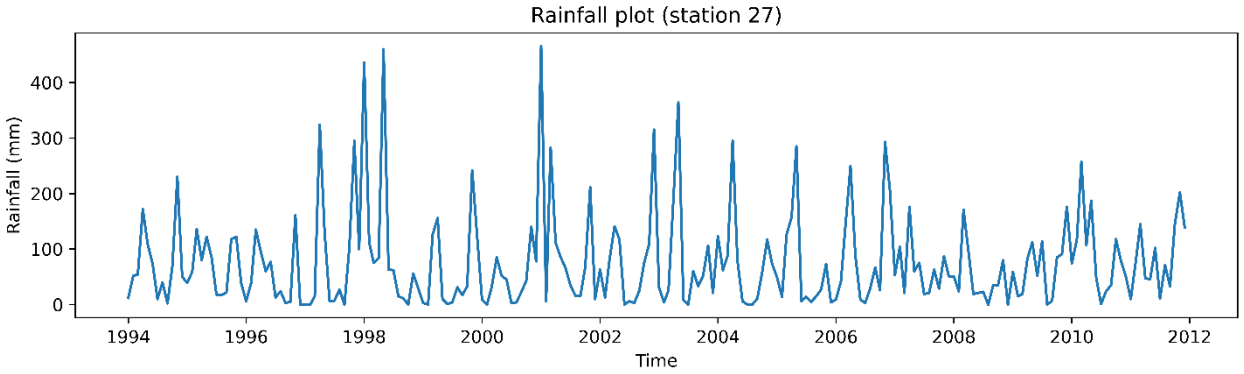
Appendix C: Rainfall records at additional stations

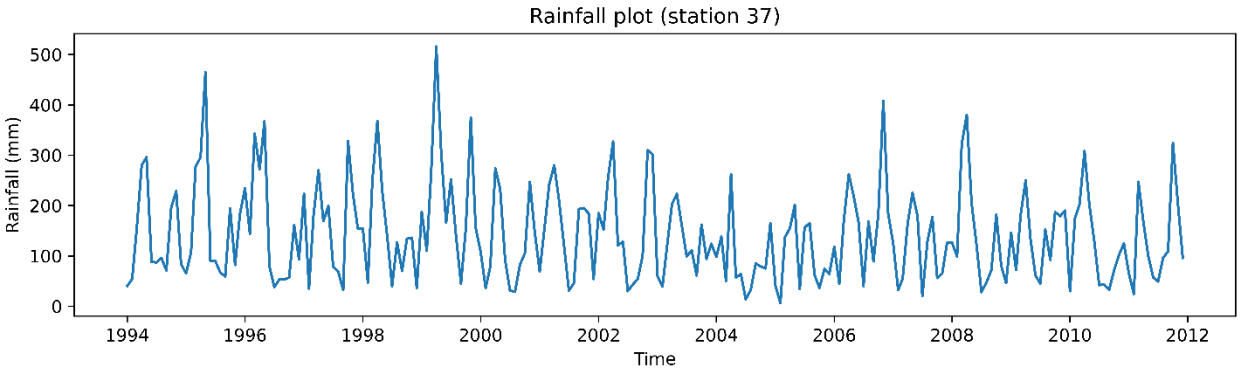




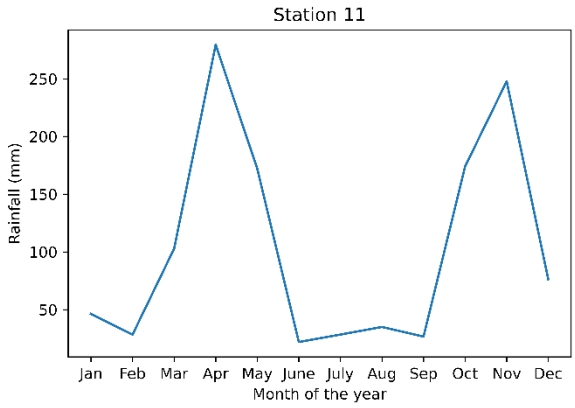


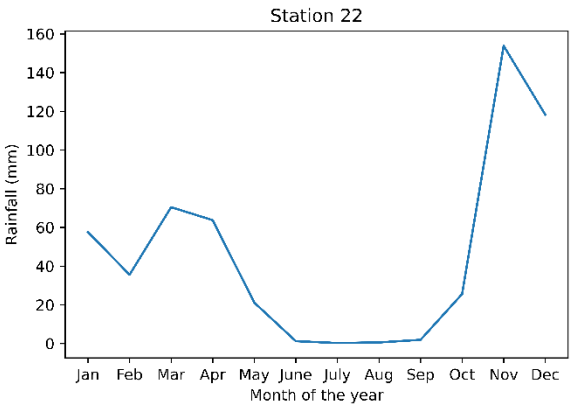

Rainfall plot (station 26)
Rainfall (mm)
0
100
200
300
400
500
600
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
Time

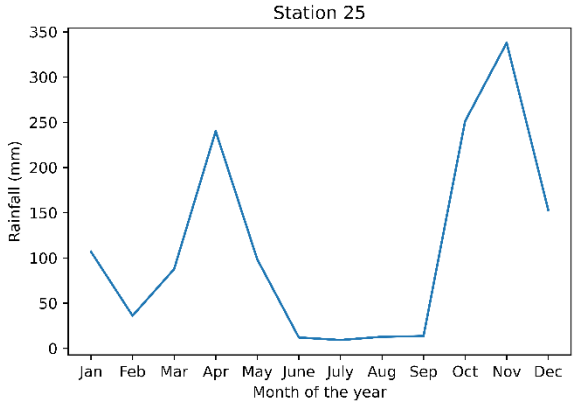

Rainfall plot (station 27)
Rainfall (mm)
0
100
200
300
400
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
Time

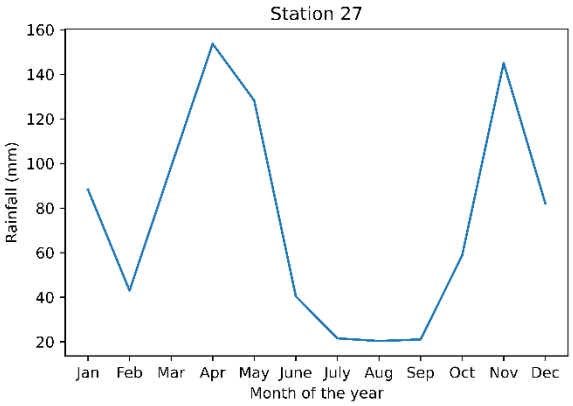

Rainfall plot (station 37)
Rainfall (mm)
0
100
200
300
400
500
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
Time

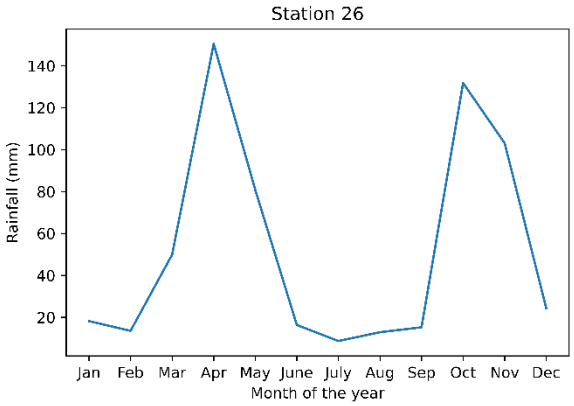
Appendix D: Mean rainfall distributions at additional stations

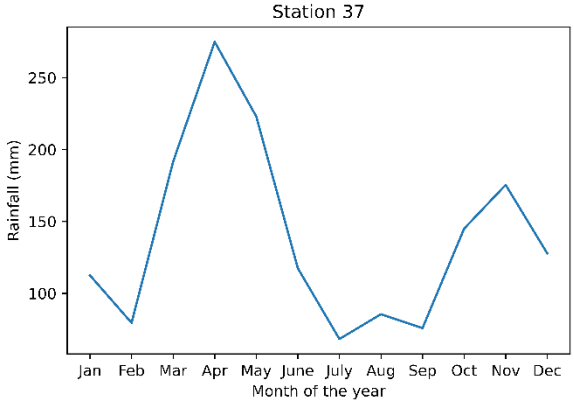




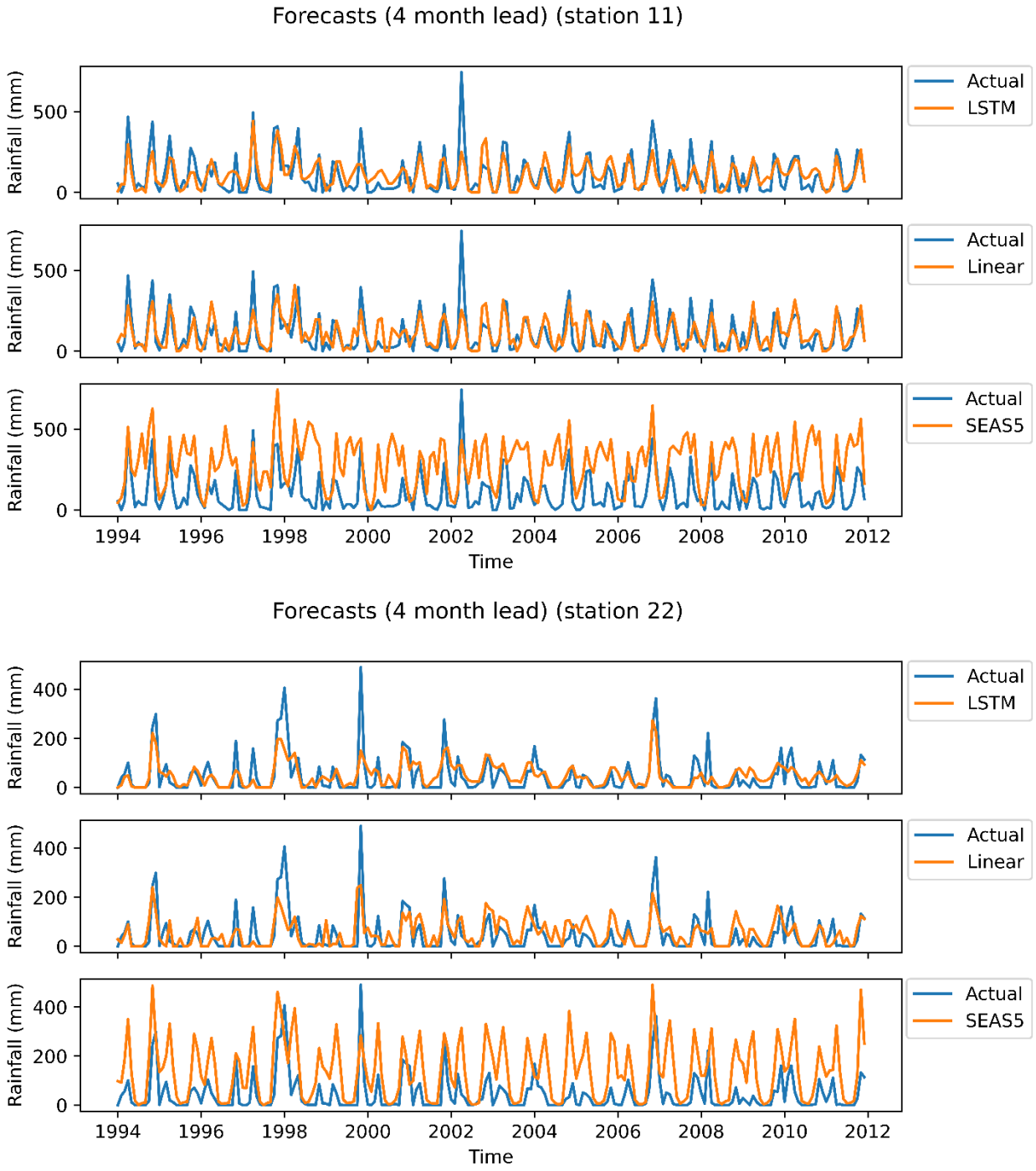


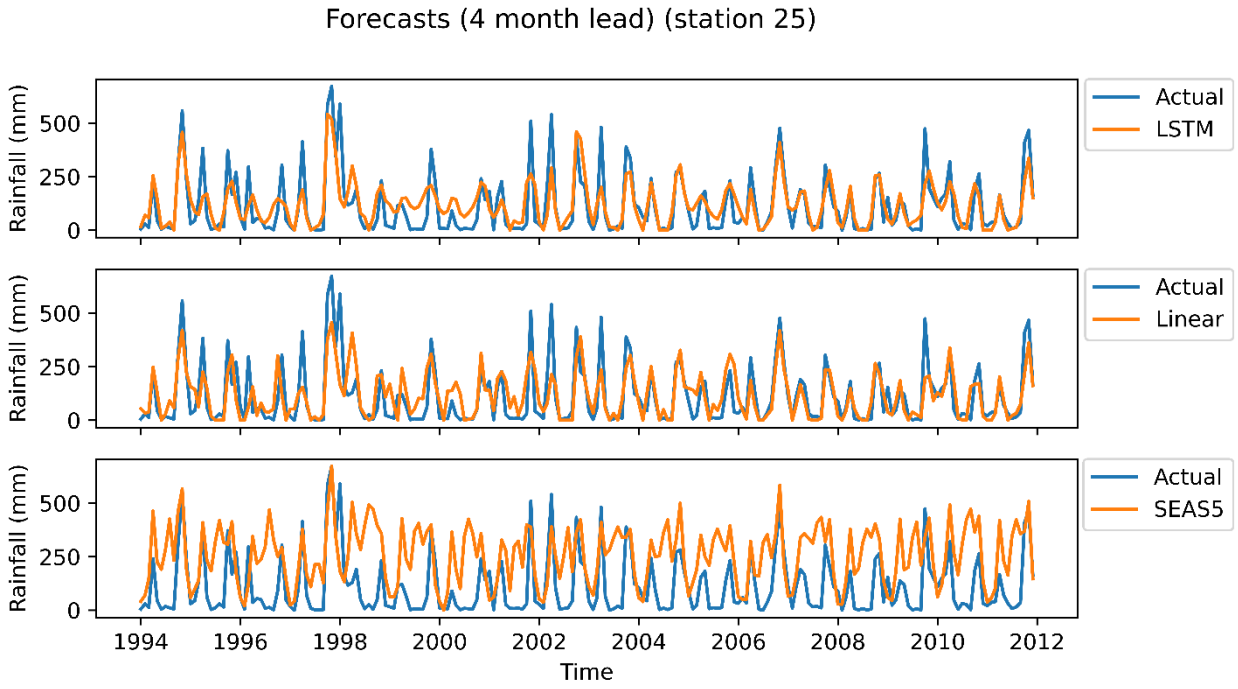


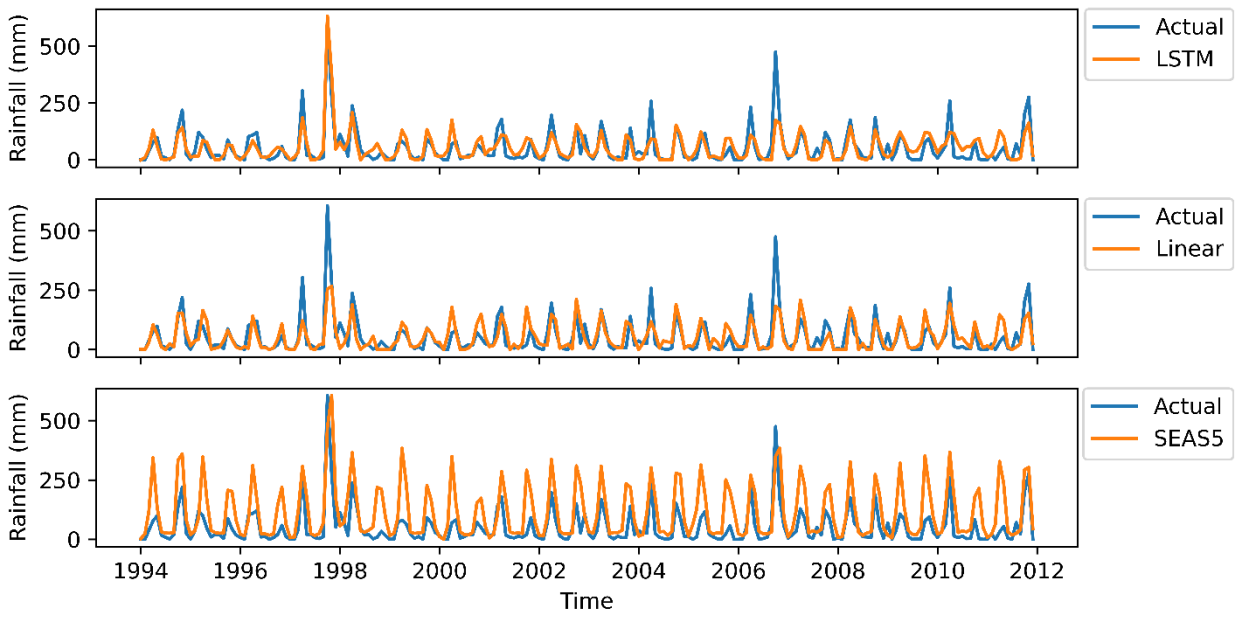


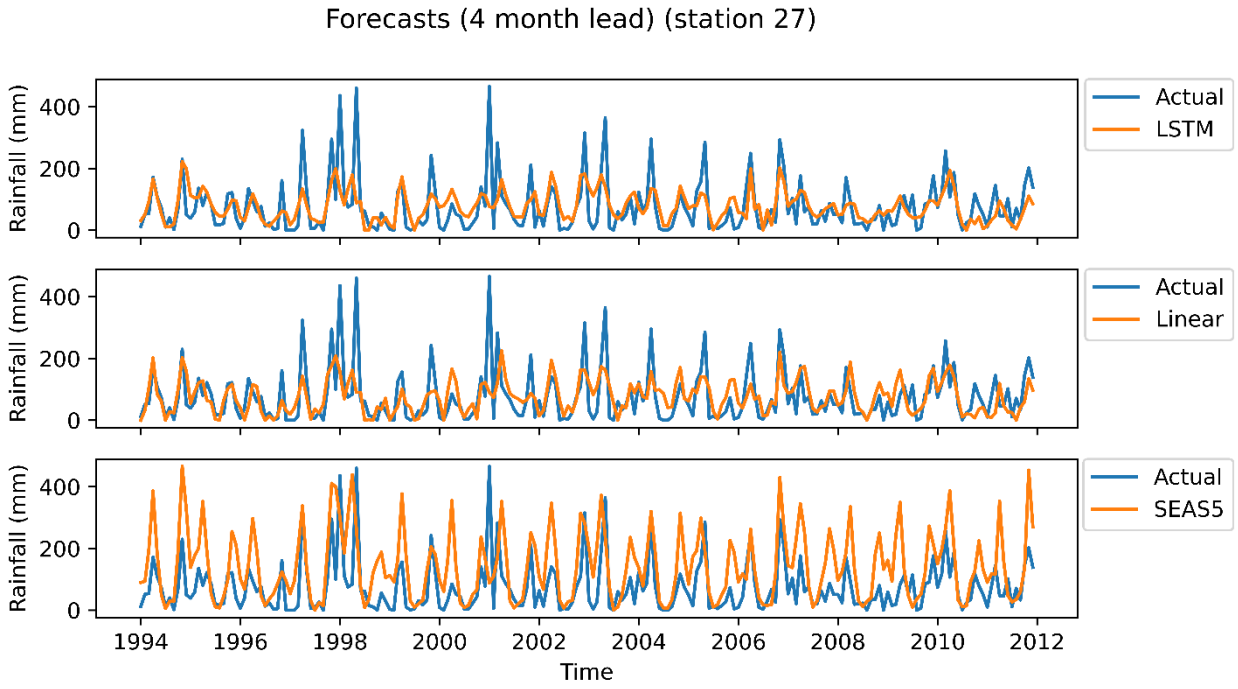


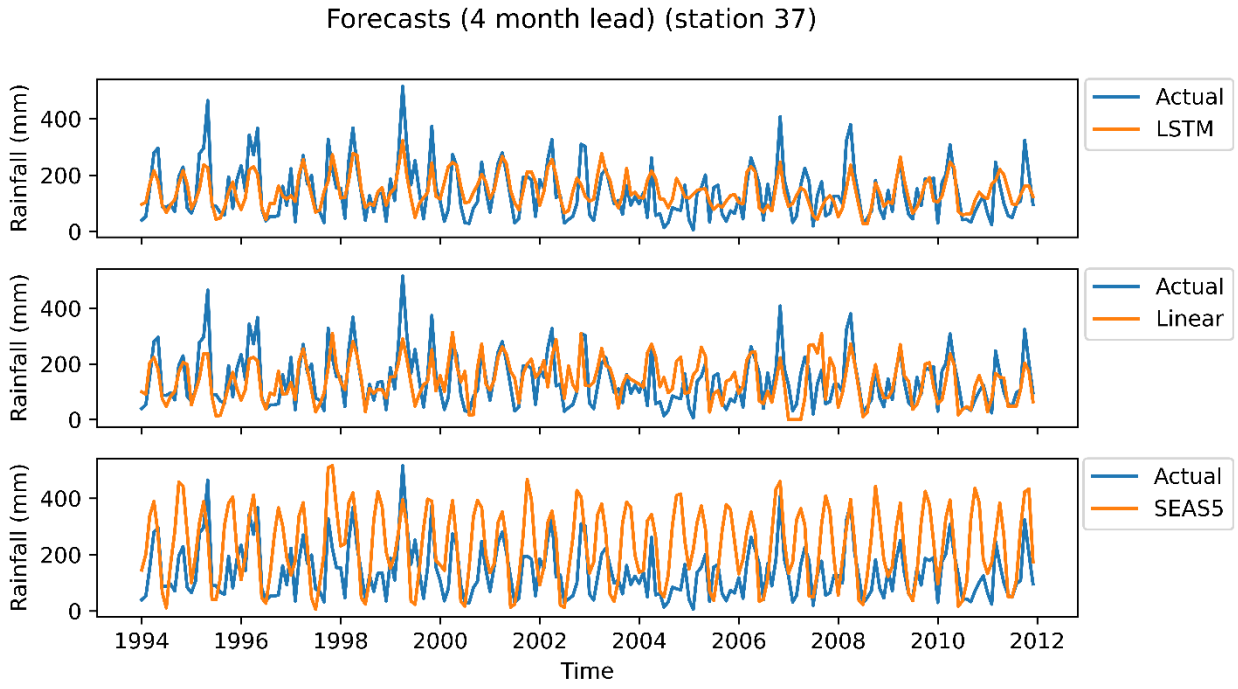
Appendix E: Forecasts at additional stations



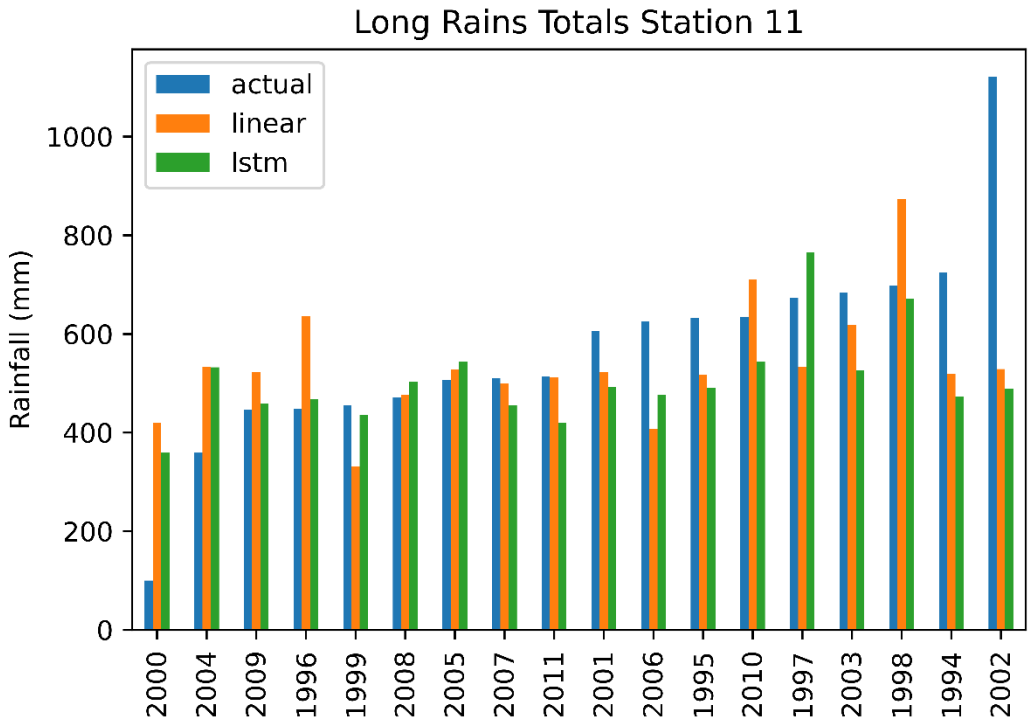

Forecasts (4 month lead) (station 25)
Rainfall (mm)
0
250
500
Actual
LSTM
Actual
Linear
Actual
SEAS5
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
Time

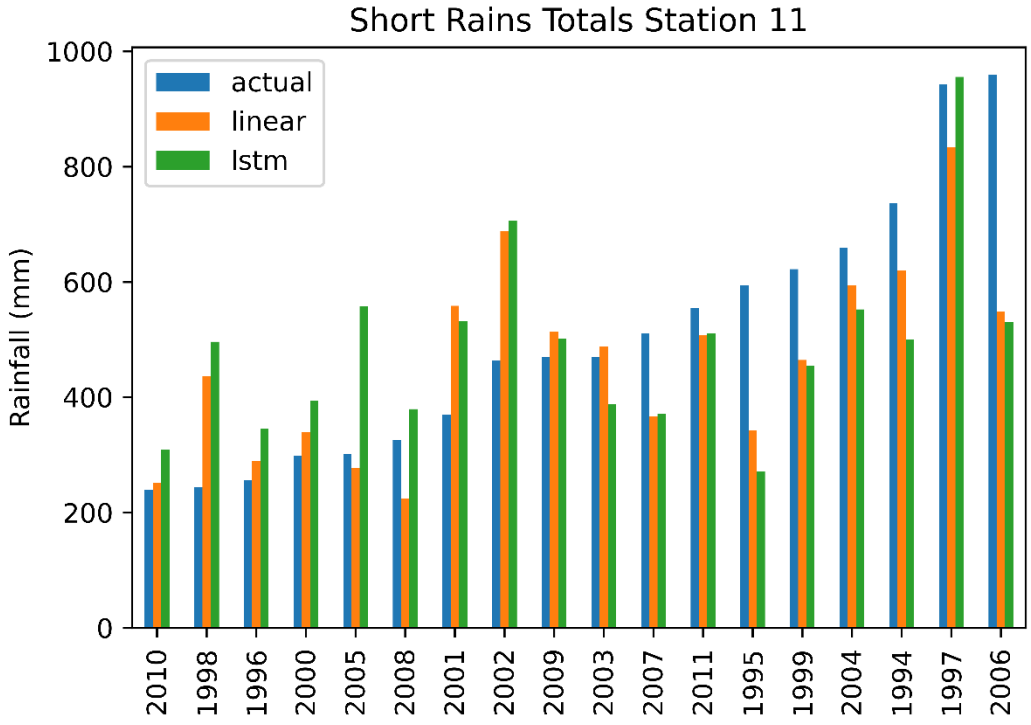

Forecasts (4 month lead) (station 26)
Rainfall (mm)
0
250
500
Actual
LSTM
Actual
Linear
Actual
SEAS5
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
Time

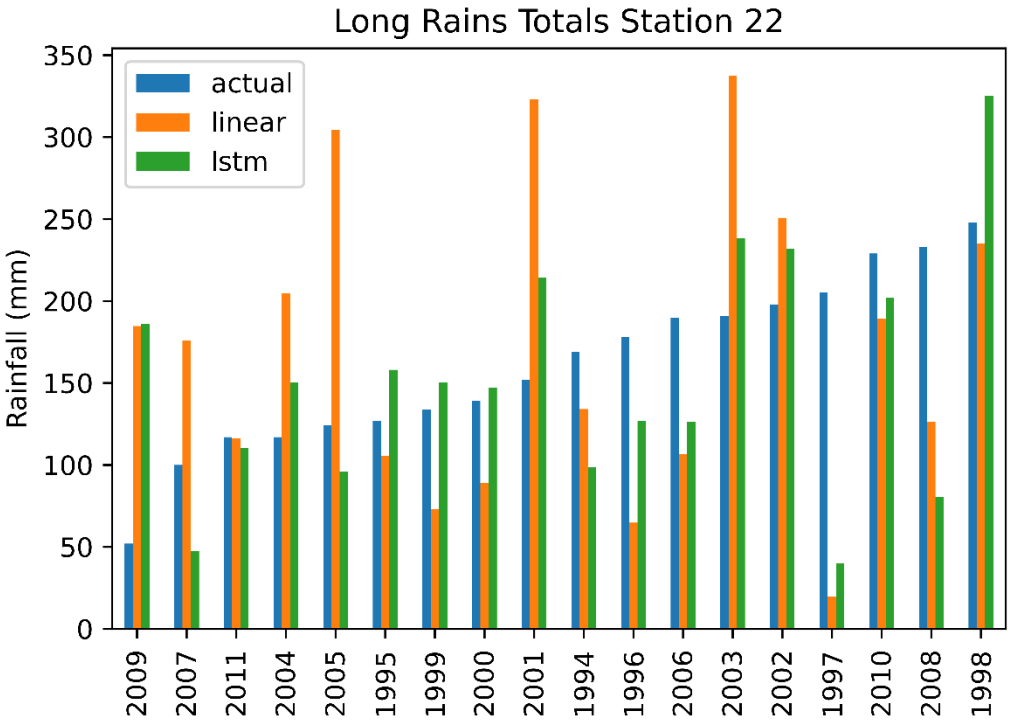

Forecasts (4 month lead) (station 27)
Rainfall (mm)
0
200
400
Actual
LSTM
Actual
Linear
Actual
SEAS5
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
Time

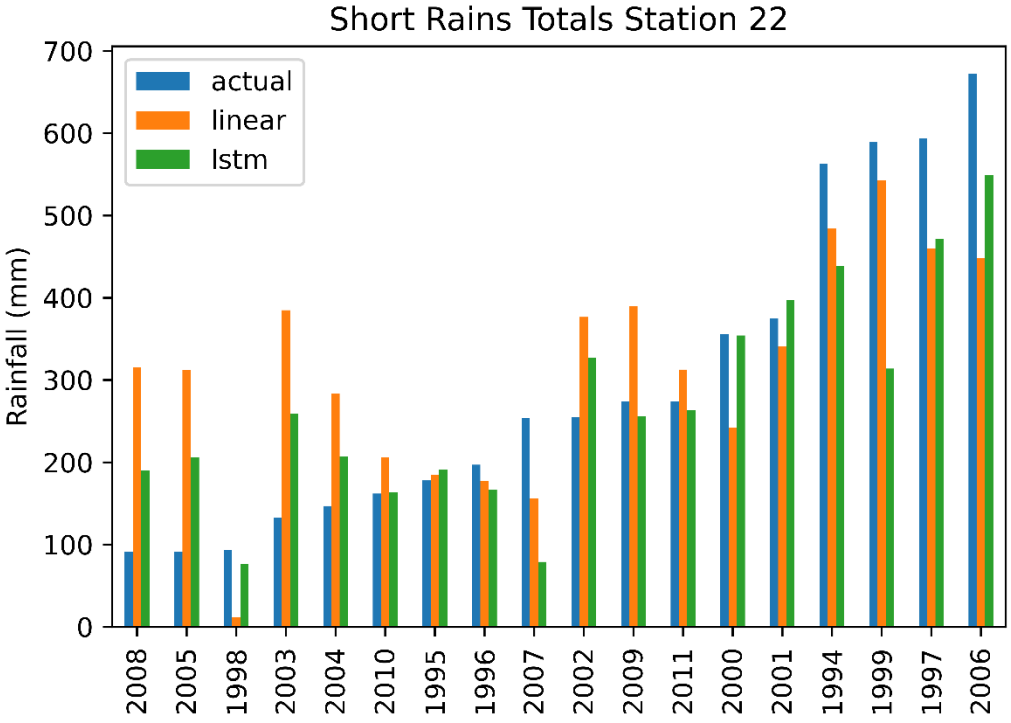

Forecasts (4 month lead) (station 37)
Rainfall (mm)
0
200
400
Actual
LSTM
Actual
Linear
Actual
SEAS5
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
Time

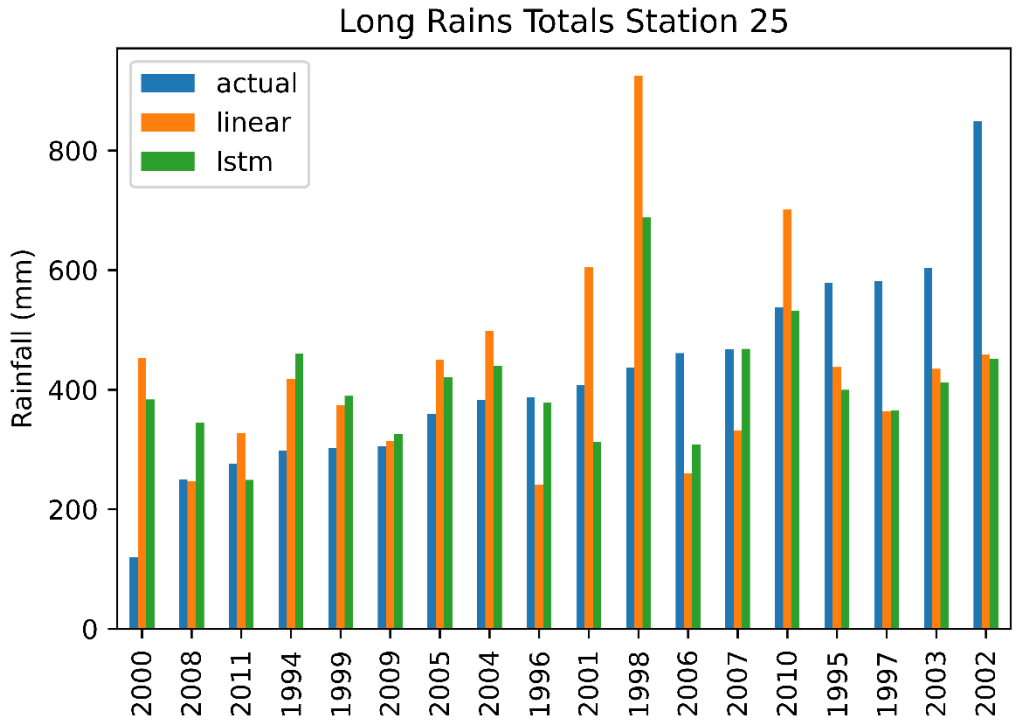
Appendix F: Seasonal rainfall totals at additional stations

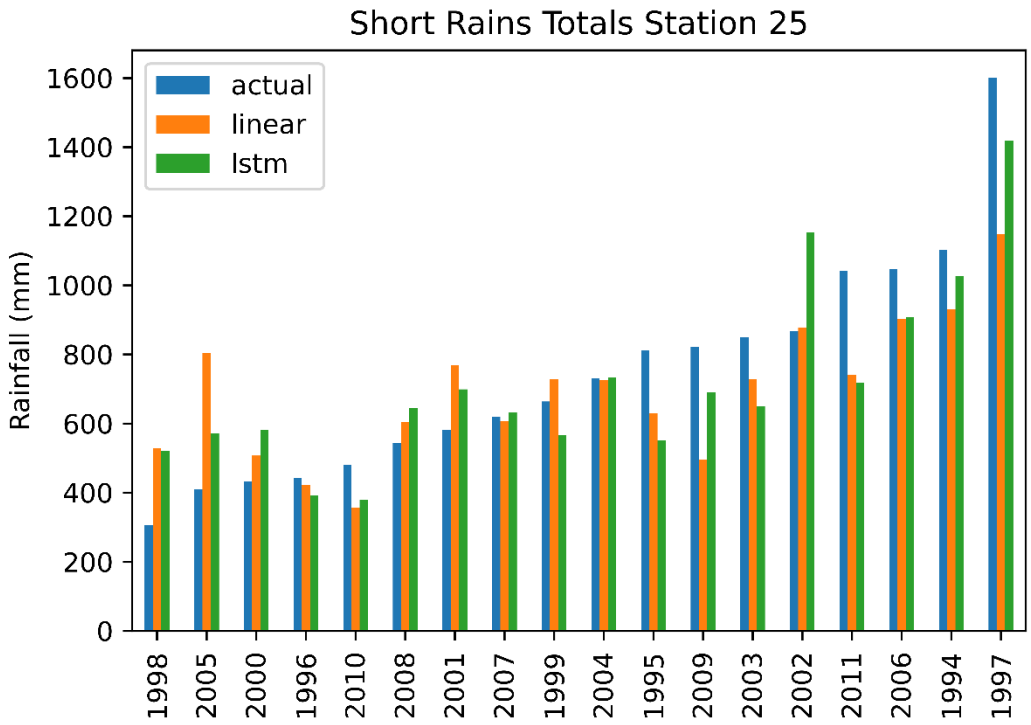


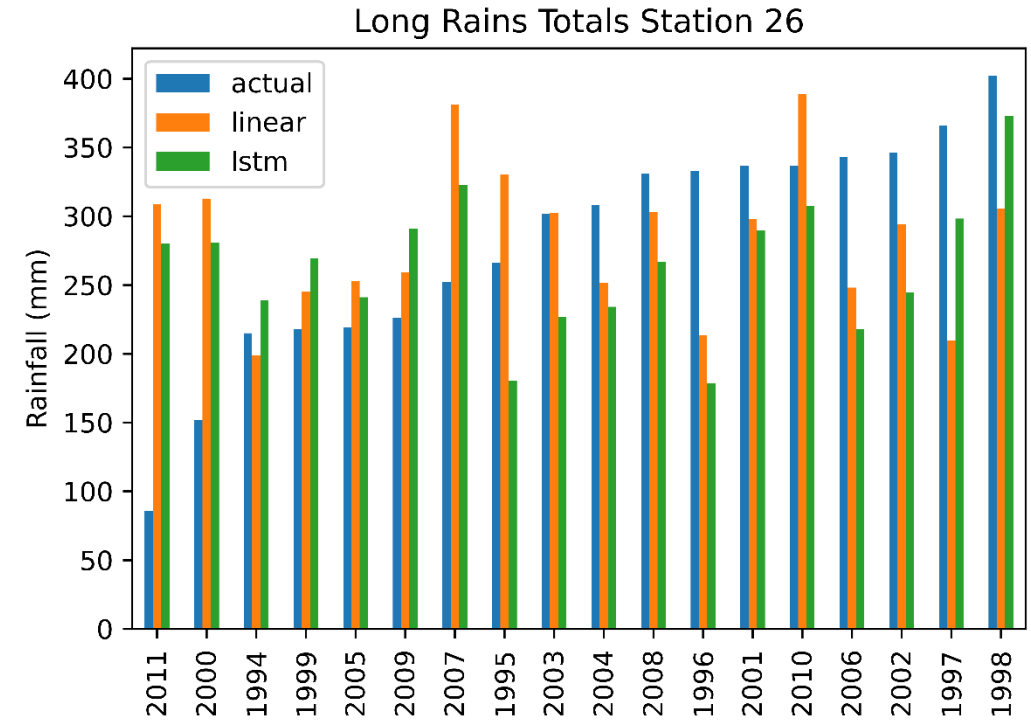


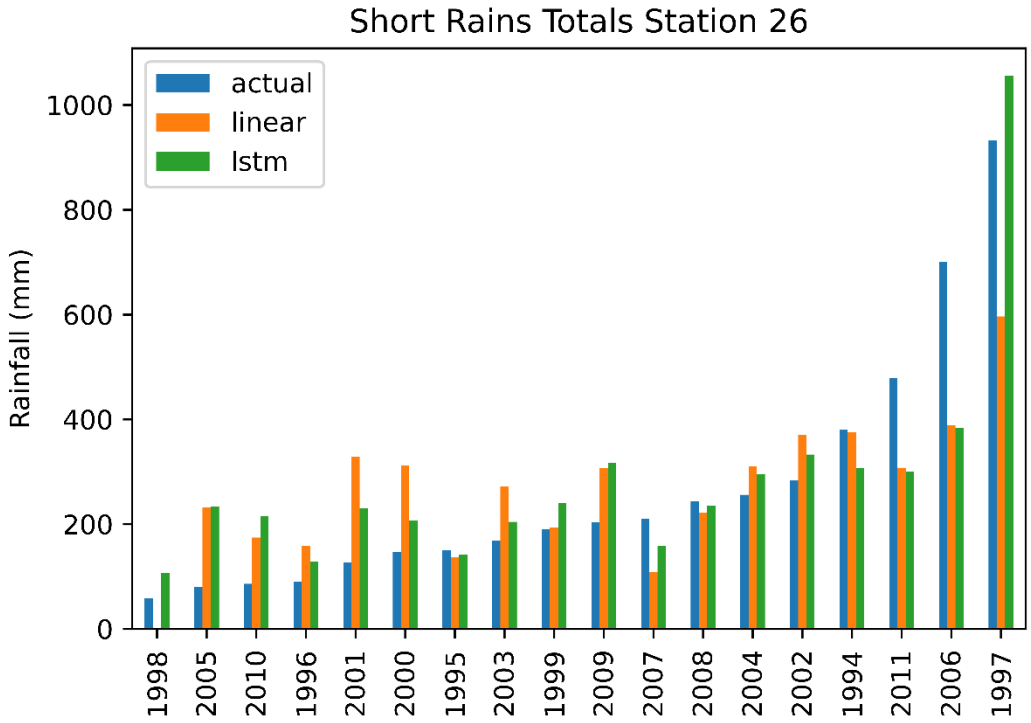

Long Rains Totals Station 22
actual
linear
lstm
Rainfall (mm)
0
50
100
150
200
250
300
350
2009
2007
2011
2004
2005
1995
1999
2000
2001
1994
1996
2006
2003
2002
1997
2010
2008
1998

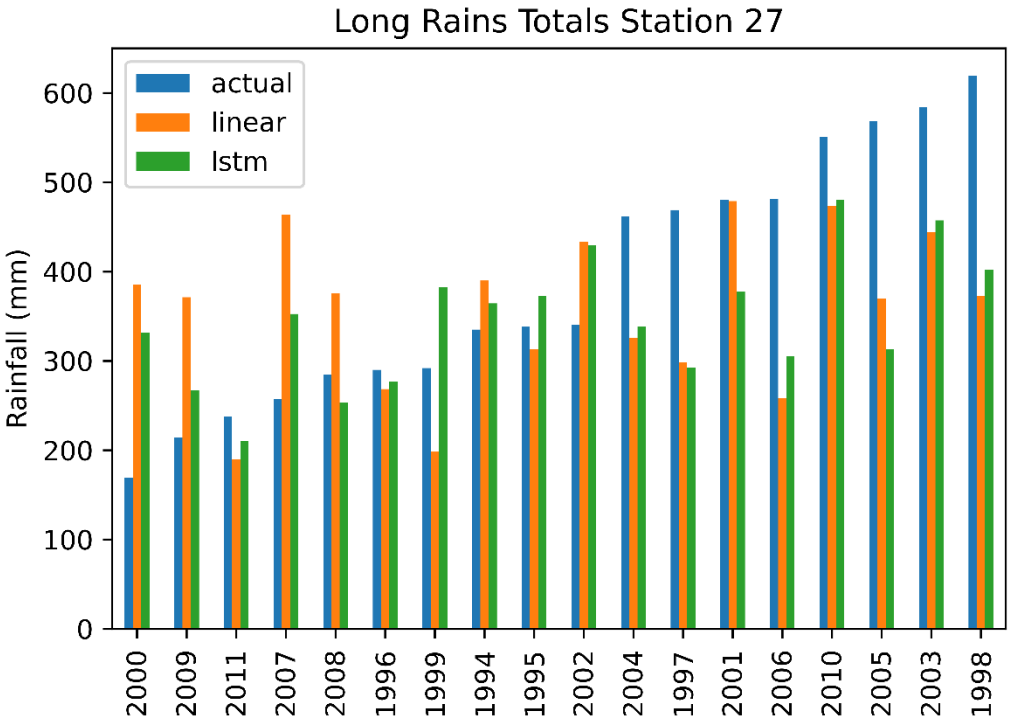

Short Rains Totals Station 22
actual
linear
lstm
Rainfall (mm)
0
100
200
300
400
500
600
700
2008
2005
1998
2003
2004
2010
1995
1996
2007
2002
2009
2011
2000
2001
1994
1999
1997
2006

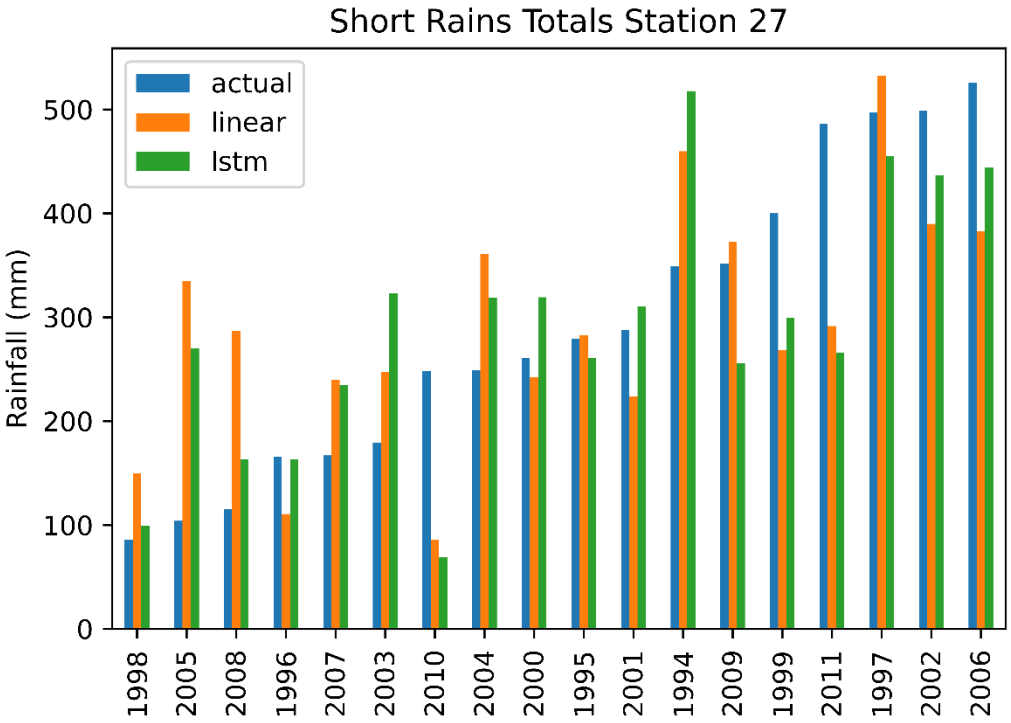

Long Rains Totals Station 25
actual
linear
lstm
Rainfall (mm)
0
200
400
600
800
2000
2008
2011
1994
1999
2009
2005
2004
1996
2001
1998
2006
2007
2010
1995
1997
2003
2002

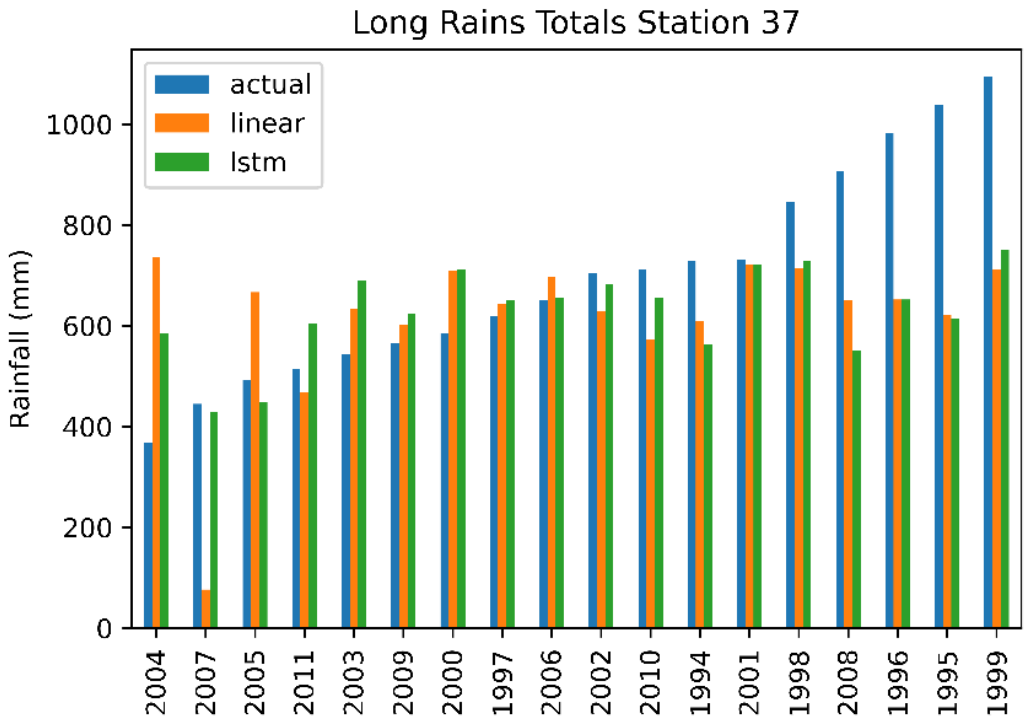

Short Rains Totals Station 25
actual
linear
lstm
Rainfall (mm)
0
200
400
600
800
1000
1200
1400
1600
1998
2005
2000
1996
2010
2008
2001
2007
1999
2004
1995
2009
2003
2002
2011
2006
1994
1997

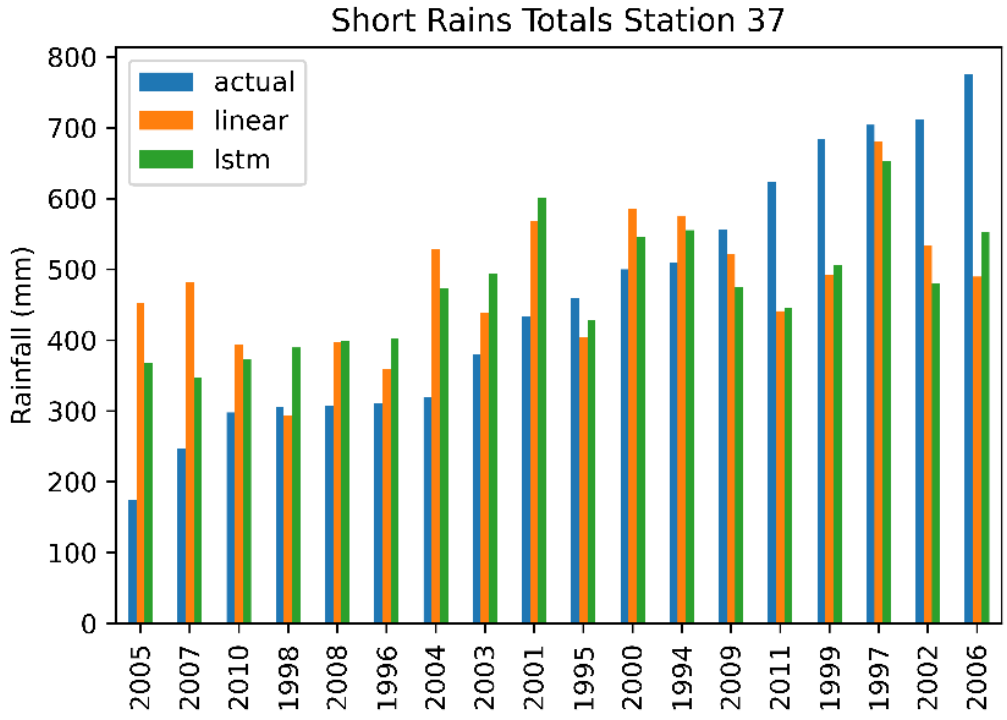

Long Rains Totals Station 26
actual
linear
lstm
Rainfall (mm)
0
50
100
150
200
250
300
350
400
2011
2000
1994
1999
2005
2009
2007
1995
2003
2004
2008
1996
2001
2010
2006
2002
1997
1998


Short Rains Totals Station 26
actual
linear
lstm
Rainfall (mm)
0
200
400
600
800
1000
1998
2005
2010
1996
2001
2000
1995
2003
1999
2009
2007
2008
2004
2002
1994
2011
2006
1997


Long Rains Totals Station 27
actual
linear
lstm
Rainfall (mm)
0
100
200
300
400
500
600
2000
2009
2011
2007
2008
1996
1999
1994
1995
2002
2004
1997
2001
2006
2010
2005
2003
1998


Short Rains Totals Station 27
actual
linear
lstm
Rainfall (mm)
0
100
200
300
400
500
1998
2005
2008
1996
2007
2003
2010
2004
2000
1995
2001
1994
2009
1999
2011
1997
2002
2006


Long Rains Totals Station 37
actual
linear
lstm
Rainfall (mm)
0
200
400
600
800
1000
2004
2007
2005
2011
2003
2009
2000
1997
2006
2002
2010
1994
2001
1998
2008
1996
1995
1999


Short Rains Totals Station 37
actual
linear
lstm
Rainfall (mm)
0
100
200
300
400
500
600
700
800
2005
2007
2010
1998
2008
1996
2004
2003
2001
1995
2000
1994
2009
2011
1999
1997
2002
2006

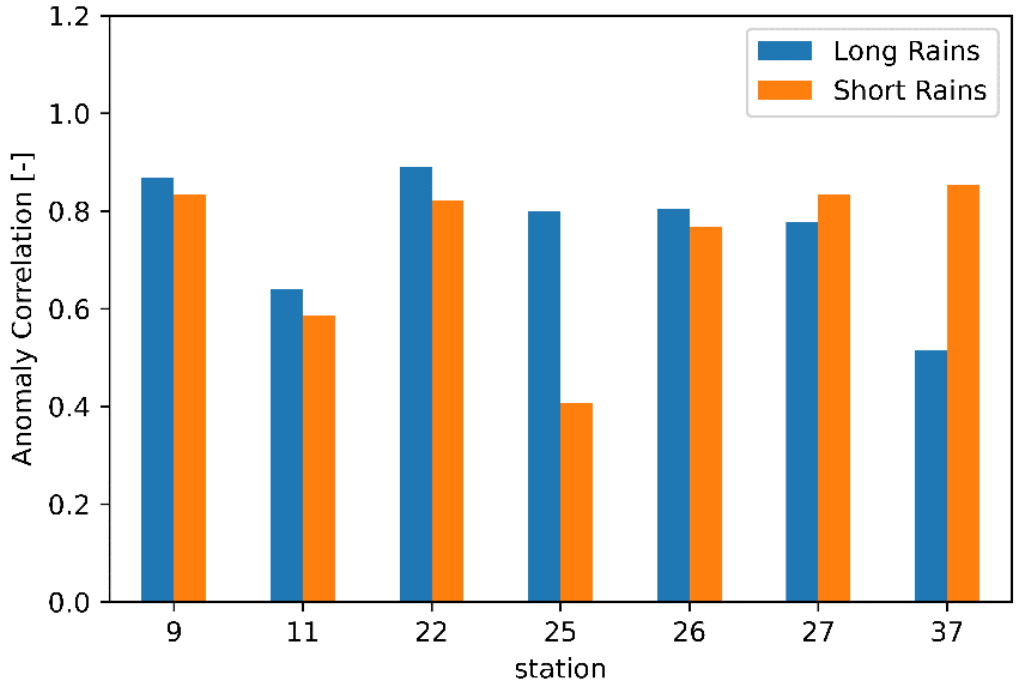
Appendix G: 2x2 contingency tables and scores for individual stations

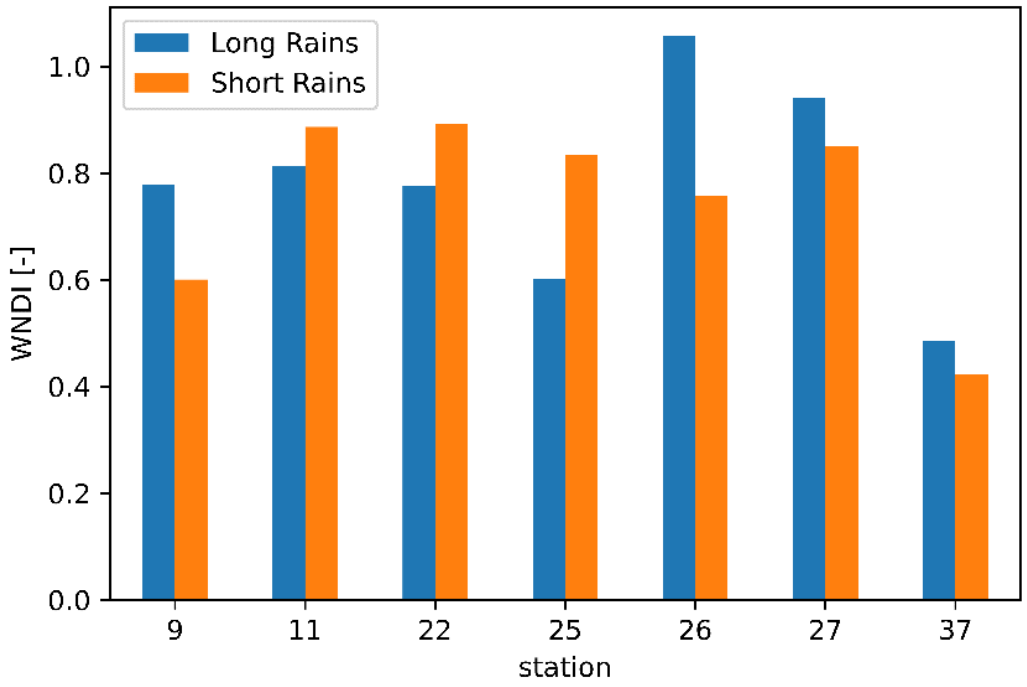
Season	Station	Domain	Model	Bias	Hit Rate	F	CSS
Long Rains	9	below-normal	linear	1	0,6	0,15	0,45
			lstm	0,8	0,6	0,08	0,61
	11	below-normal	linear	0,5	0	0,06	-0,12
			lstm	0,5	0,5	0	0,94
	22	below-normal	linear	1,75	0,25	0,43	-0,13
			lstm	1,5	0,5	0,29	0,17
	25	below-normal	linear	1	0,33	0,13	0,2
			lstm	0,33	0,33	0	0,88
	26	below-normal	linear	1,5	0	0,19	-0,13
			lstm	1	0	0,12	-0,12
	27	below-normal	linear	1	0,25	0,21	0,04
			lstm	0,75	0,5	0,07	0,53
	37	below-normal	linear	0,5	0,5	0	0,88
			lstm	0,5	0,5	0	0,88
Short Rains	9	below-normal	linear	0,75	0,5	0,07	0,53
			lstm	0,75	0,5	0,07	0,53
	11	below-normal	linear	0,8	0,6	0,08	0,61
			lstm	0,4	0,2	0,08	0,25
	22	below-normal	linear	0,25	0,25	0	0,82
			lstm	0,5	0,25	0,07	0,31
	25	below-normal	linear	0,4	0,4	0	0,81
			lstm	0,4	0,4	0	0,81
	26	below-normal	linear	0,5	0,5	0	0,94
			lstm	0	0	0	
	27	below-normal	linear	0,6	0,4	0,08	0,47
			lstm	0,8	0,6	0,08	0,61
	37	below-normal	linear	0,17	0,17	0	0,71
			lstm	0	0	0	

Season	Station	Domain	Model	Bias	Hit Rate	F	CSS
Long Rains	9	normal	linear	1,22	0,78	0,44	0,35
			lstm	1,44	1	0,44	0,69
	11	normal	linear	1,07	0,87	1	-0,19
			lstm	1,07	0,93	0,67	0,38
	22	normal	linear	0,5	0,1	0,5	-0,49
			lstm	0,8	0,5	0,38	0,12
	25	normal	linear	1,09	0,55	0,86	-0,33
			lstm	1,45	0,91	0,86	0,12
	26	normal	linear	1	0,69	0,8	-0,11
			lstm	1,15	0,85	0,8	0,07
	27	normal	linear	1,4	0,7	0,88	-0,25
			lstm	1,5	0,9	0,75	0,27
	37	normal	linear	1,6	1	0,75	0,62
			lstm	1,6	1	0,75	0,62
Short Rains	9	normal	linear	1	0,7	0,38	0,32
			lstm	1,1	0,8	0,38	0,44
	11	normal	linear	1,2	0,8	0,5	0,33
			lstm	1,4	0,8	0,75	0,07
	22	normal	linear	1,3	1	0,38	0,77
			lstm	1,4	0,9	0,62	0,39
	25	normal	linear	1,67	1	0,67	0,6
			lstm	1,44	0,89	0,56	0,42
	26	normal	linear	1,23	1	0,6	0,81
			lstm	1,31	1	0,8	0,76
	27	normal	linear	1,44	0,78	0,67	0,14
			lstm	1,11	0,78	0,33	0,45
	37	normal	linear	2,29	1	0,82	0,44
			lstm	2,43	1	0,91	0,41

Season	Station	Domain	Model	Bias	Hit Rate	F	CSS
Long Rains	9	above-normal	linear	0,5	0,25	0,07	0,31
			lstm	0,25	0,25	0	0,82
	11	above-normal	linear	1	0	0,06	-0,06
			lstm	1	0	0,06	-0,06
	22	above-normal	linear	1,5	0,25	0,36	-0,08
			lstm	1	0,25	0,21	0,04
	25	above-normal	linear	0,75	0	0,21	-0,27
			lstm	0,25	0	0,07	-0,24
	26	above-normal	linear	0,67	0	0,13	-0,19
			lstm	0,33	0,33	0	0,88
	27	above-normal	linear	0	0	0	
			lstm	0	0	0	
	37	above-normal	linear	0	0	0	
			lstm	0	0	0	
Short Rains	9	above-normal	linear	1,25	0,75	0,14	0,52
			lstm	1	0,75	0,07	0,68
	11	above-normal	linear	0,67	0,33	0,07	0,38
			lstm	0,67	0,33	0,07	0,38
	22	above-normal	linear	1	1	0	1
			lstm	0,5	0,5	0	0,88
	25	above-normal	linear	0,25	0,25	0	0,82
			lstm	0,75	0,5	0,07	0,53
	26	above-normal	linear	0,33	0,33	0	0,88
			lstm	0,33	0,33	0	0,88
	27	above-normal	linear	0,5	0,25	0,07	0,31
			lstm	1	0,75	0,07	0,68
	37	above-normal	linear	0,2	0,2	0	0,76
			lstm	0,2	0,2	0	0,76

Appendix H: performance at individual stations



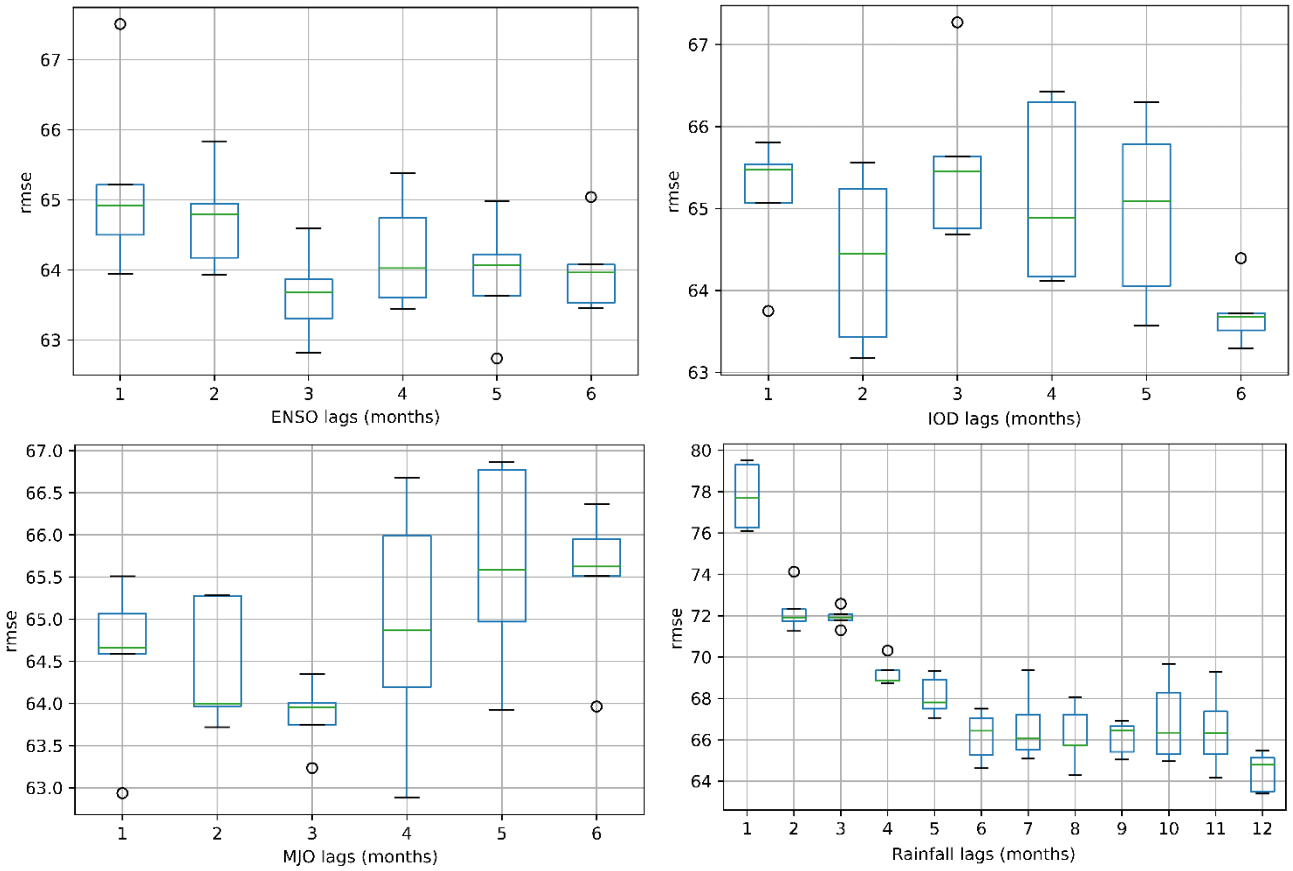


Appendix I: model calibration and robustness

Optimal lag times for rainfall, ENSO, IOD and MJO

The input dataset used in this study consisted of lagged monthly rainfall, lagged climate indices and seasonal forecasts produced by the ECMWFs SEAS5. Multiple repeated experiments were carried out to find the optimal lag times of these features. Note that an optimal lag time of, for example, 4 months means that all smaller lag times are also included (1, 2 and 3 months).

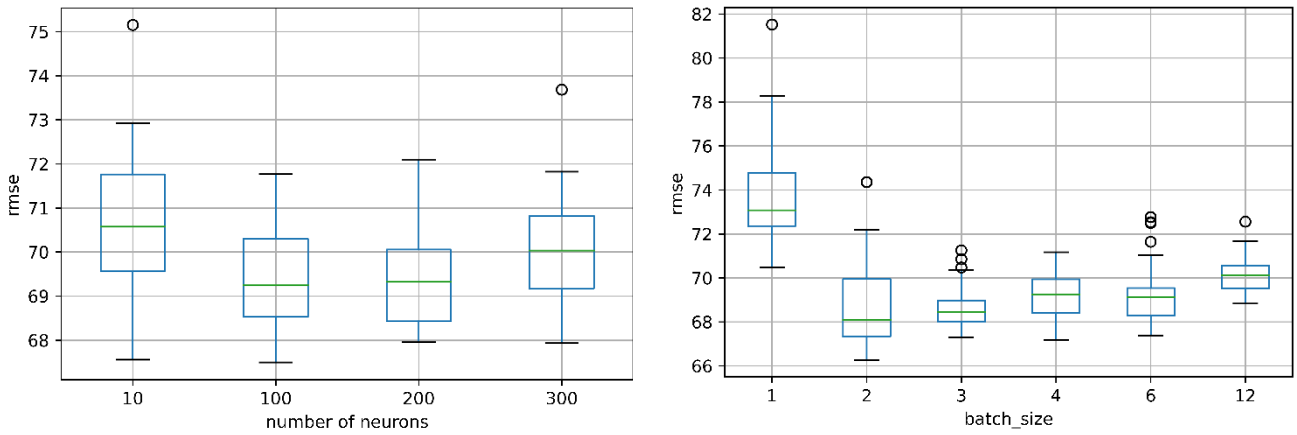
The figure below provides an overview of the results. Optimal lag times for ENSO were found between 3 and 6 months. For DMI, the optimal setting was found at 6 months. Combining the results of the repeated experiments with a physical understanding of the ENSO and IOD phenomena, the lag time for both ENSO and IOD was set to 6 months. MJO lags were set to 3 months and the lags of precipitation to 12 months as those settings resulted in the lowest mean/spread.



Boxplots of RMSE for several lag times for ENSO, IOD, MJO and Rainfall (in months).

Hyperparameter tuning

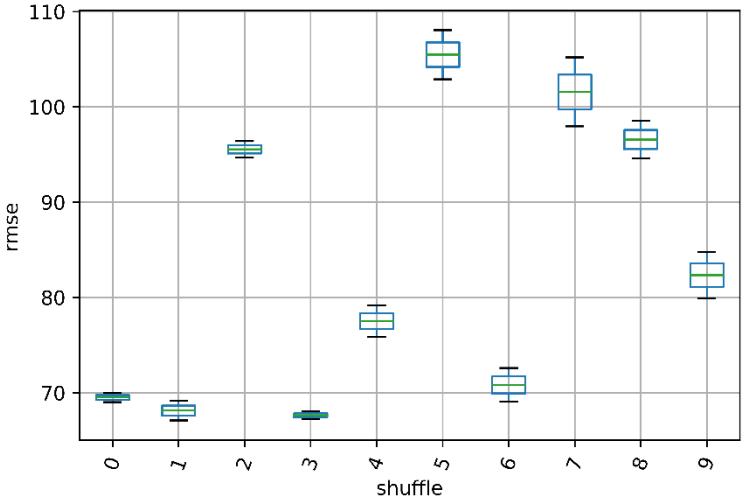
The Long Short-Term Memory (LSTM) model was tuned specifically for the data used in this study. Optimal settings for 2 hyperparameters (number of neurons & batch size) were found based on another repeated experiment. Based on the results in the figure below, the number of neurons was set to 100 and the batch size was set to 3. Additional experimentation with smaller and larger values of the number of neurons, indicated an optimum somewhere between 100 and 200 neurons. As higher numbers of neuron may increase the risk of overfitting, the number of neurons was set to 100.



Boxplots of RMSE for different settings for LSTM hyperparameters: number of neurons (left) and batch size (right).

Robustness experiment

The final repeated experiment was carried out to evaluate the robustness of the model when the input dataset is shuffled before training. In the procedure, only the years are shuffled and the months within the years are kept in chronological order (Figure 11 in the main text). The figure below shows the results of the repeated experiment. Shuffle 0 represents the original chronological input dataset (so without any shuffling applied) and shuffle 1-9 indicate randomly shuffled versions of the input dataset. The range of median RMSE values observed throughout the experiment is about (65-105) 40 mm. 5/10 experiments have a median RMSE below 80 and the largest variance is observed for the largest median values. The original chronological dataset is observed to result in the 3rd best performance.



Boxplots of RMSE as a function of the shuffle (shuffle=0 represents the chronological dataset, shuffle 1-9 are randomized)