

Modeling Seasonal and Annual Precipitation using long-term Climate Records and Topography

Master's Thesis

Faculty of Science

University of Bern

presented by

Noemi Imfeld

2016

Supervisor:

Prof. Dr. Stefan Brönnimann
Institute of Geography, Climatology Group
Oeschger Centre for Climate Change Research

Advisor:

Dr. Hanspeter Liniger
Institute of Geography
Center for Development and Environment

Abstract

Rainfall around Mount Kenya is of high importance for the people in its surrounding. The mountain is the water tower for a wide-spread region. People on the slopes of Mt. Kenya mainly depend on rain-fed agriculture, whereas for people in the semi-arid lowlands the river water is important for pastoralism. The high amounts of precipitation around the mountain due to orographic enhancement recharge the main river and provide a livelihood for both upstream farmers and downstream pastoralists. The characteristic topography shows a major influence on local to regional rainfall variation. The study area located in the tropics exhibits a bimodal annual rainfall cycle with areas comparatively dry for their latitude. Prolonged dry seasons or failing rains trigger conflicts between upstream and downstream users and highlight the need to develop water resource allocation strategies.

Continuous spatial information of the rainfall pattern may support decisions in water resource allocations between upstream and downstream users or may improve hydrological studies. In order to provide such continuous spatial information a reliable and sufficiently dense network of rain data is necessary. For the study area in Central Kenya such a dense network of rain gauge with reliable long-term records is available.

The aim of this study is to estimate precipitation at unsampled locations using topography and to produce gridded precipitation maps of seasonal and annual precipitation for a period from 1977 to 2014. A multiple linear regression model is evaluated for the seasonal and annual precipitation totals for geographic and topographic predictors. In addition, residual precipitation is interpolated using ordinary kriging and is added to the maps predicted by the multiple linear regression model. Topographic predictors, such as slope, aspect, altitude or combinations of these variables, are calculated at different scales, as topography might influence precipitation distinctively at different scales. Further, a predictor to account for the barrier effect of Mt. Kenya is constructed. It consists of an angle around the mountain and a distance function. The angle is used to predict on what expositions of the mountain highest precipitation occurs. The distance function accounts for a decreasing effect of the mountain on precipitation with distance from the mountain and for decreasing precipitation at high altitudes. Geographic predictors are latitude and longitude. They are used in order to model the gradient between the semi-arid north and the humid southern part of the study area.

The best regression model consists of predictors of slope, a combination of slope and aspect, two variables at different frequencies to account for Mt. Kenya, and of the geographic predictors latitude and longitude. The regression model is the same for the two rainy seasons and annual total precipitation. Slope calculated at a scale of 27 km shows higher correlations with precipitation totals than calculated at lower scales. This scale is therefore used in the regression model. Altitude is not used, as a lot of years show unrealistic negative correlations with precipitation. However, correlations with altitude are higher calculated for lower scales and for annual totals compared to the two rainy seasons. The impact of the Mt. Kenya variables on predictive R^2 shows that constructing such a predictor is a suitable way to model the precipitation pattern around the mountain. Mean best predictive skills for the entire time period are found for the boreal autumn rains with a predictive R^2 of 0.656. Values for boreal spring rain totals and annual totals are lower and of 0.464, respectively of 0.454. This indicates that there is an influence of topography on the spatial pattern of precipitation and that this influence is more pronounced for the boreal autumn rains, than for the boreal spring rains and annual precipitation totals.

Ordinary Kriging of the residual precipitation allows to add additional valuable information that the multiple linear regression model is not able to predict. Thus, this combined approach of multiple linear regression and Ordinary Kriging seems a reasonable way to model the precipitation pattern in the area around Mt. Kenya.

Acknowledgements

First of all, I would like to thank to my supervisor Stefan Brönnimann for sharing his useful ideas and for his support. Further, I would like to thank my advisor Hanspeter Liniger for his help and for the two fascinating, instructive and active weeks at the beginning of our fieldwork in Kenya. I would like to thank to my driver Dennis, who was ready to drive everywhere I asked for, and Irene for her company during our numerous field trips in Kenya.

A special thank to Fabienne Kaufmann and Danielle Huser for the time we shared in Nanyuki, for having enlightening discussions and pleasant evenings.

And another special thank to my family for their support throughout my studies, to my flatmates for motivating words and to my friends that were ready to do a fast proof-reading.

Contents

Abstract	iii
List of Figures	ix
List of Tables	xiii
List of Abbreviations	xv
1 Introduction	1
1.1 Motivation	1
1.2 State of Research	2
1.3 Research Objectives	3
2 Study Area	5
2.1 Location and Topography	5
2.2 Climatology	6
2.2.1 Spatial Pattern of Precipitation	6
2.2.2 Annual Cycle of Precipitation	8
2.2.3 Interannual Variability of Rainy Seasons	9
3 Data	11
3.1 Manual Rain Gauge Data	11
3.1.1 Description of Network	11
3.1.2 Quality Assurance	12
3.1.3 Quality Control	13
3.2 Automatic Weather Stations	14
3.2.1 Network Description	14
3.2.2 Quality Control	14
3.3 Global Precipitation Mission Satellite Data	15
3.4 Digital Elevation Model	15
4 Methodology	17
4.1 Regression Kriging	17
4.2 Multiple Linear Regression	18
4.2.1 Explanatory Variables	18
4.2.2 Model Selection	21
4.2.3 Scale Selection	22
4.2.4 Cross-Validation of the MLR Estimation	22
4.3 Ordinary Kriging of Residuals	22
4.3.1 Theoretical Basis	22
4.3.2 Implementation	23
4.4 Comparison with Independent Data	24
4.4.1 Comparison with IMERG Data	24
4.4.2 Comparison with AWS Data	24

4.5	Trend Calculations for Total Precipitation	24
5	Results	25
5.1	Quality of Manual Rain Gauge Network	25
5.1.1	Quality Classes	25
5.1.2	Quality Assurance	26
5.1.3	Quality Control	28
5.2	Model and Scale Selection	29
5.2.1	Correlation of Total Precipitation with Predictors	29
5.2.2	MAM Model Selection	31
5.2.3	OND Model Selection	31
5.2.4	Annual Model Selection	32
5.3	Prediction Maps	33
5.3.1	Seasonal Precipitation	33
5.3.2	Annual Precipitation	41
5.3.3	Cross-Validation for Individual Stations	43
5.4	Comparison with Independent Data	44
5.4.1	Comparison with IMERG Data	44
5.4.2	Comparison with AWS Data	45
5.5	Trend Calculation	46
6	Discussion	47
6.1	Rain Gauge Network	47
6.2	Multiple Linear Regression	48
6.2.1	Model Selection	48
6.2.2	Predictors	49
6.3	Final Estimation	51
6.4	Comparison with Independent Data	53
6.5	Trends	54
7	Conclusion	55
A	Data Overview	57
B	Rainfall Maps	61
C	Predictive Skills	75
	Bibliography	77

List of Figures

2.1	(a) Location of study area in East Africa (grey square). Source: Schmocker et al. (2015). (b) Mean annual precipitation for all stations in the study area for the period from 1981- 2010. The two peaks show the two main rainy seasons in boreal spring and boreal autumn.	6
2.2	Mean annual rainfall for the period 1981- 2010 with topography of the study area (400 m equidistance). Colors and size of bubbles denote the amount of rainfall. . . .	7
2.3	Total precipitation along vertical transects for different expositions of Mt. Kenya and a schematic view of moisture advection, circulation, cloudiness and precipitation. Source: Barry (2008).	7
2.4	Seasonal wind at 10-m (vectors) and divergent (red) and convergent (blue) areas over East Africa from ERA-Interim for the period 1979- 2009. The red line depicts the smoothed 1000-m contour. (a) January-February, (b) March-April-May, (c) July-June-August-September, (d) October-November-December. Source: Yang et al. (2014a)	8
2.5	Schematic view of the zonal circulation along the equator leading to subsidence over East Africa and ascending motion over Indonesia. Source: Lau and Yang (2002) . . .	9
3.1	(a) Evolution of the measurement network since the beginning of observations. All years succeeding the dotted blue line are used for the estimations. (b) Example of an ordinary manual rain gauge used in the study area. Ol Maisor Farm, 2015 (own illustration).	12
3.2	(a) Map of manual and automatic weather stations on a digital elevation model of the area. IS denotes automatic weather stations in densely sampled parts of the study area and OS in data sparse areas. (b) Percentage distribution of stations (dashed line) and grid points for slope at a 45 cell window size. (c) Percentage distribution of stations (dashed line) and grid points for altitude at a 9 cell window size.	13
4.1	Visualization of equation 4.5: Blue to green colours show positive phases of the harmonic oscillations, orange to red show negative phases. The coefficients c_1 to c_4 are set to 1.	20
5.1	Quality classes of all rain gauges including former and newly assessed quality ratings, CETRAD Database, Nanyuki.	26
5.2	(a) Smoothed running correlations calculated for daily records of Satima Farm and Nicholson Farm for 2010 - 2012; as date of reading and as date of event. (b) Rounding of daily data and missing low values starts around 2003, Shamata rain gauge. . . .	27
5.3	Correlation between predictors and total precipitation for MAM, OND and annual precipitation. The upper row shows histograms for correlations with slope at S45 (blue) and at S9 (red) for MAM (a), OND (b) and annual precipitation (c). The lower row shows correlation with altitude at A157 (blue) and A19 (red) for MAM (a), OND (b) and annual precipitation (c).	30
5.4	Seasonal precipitation 1977: (a) MLR MAM, (b) MLR OND, (c) final estimation MAM, (d) final estimation OND. The filled squares represent the measured value at the station.	35

5.5	Seasonal residual kriging 1977: (a) MAM , (b) OND	35
5.6	Seasonal Precipitation 1989: (a) MLR MAM, (b) MLR OND, (c) final estimation MAM, (d) final estimation OND. The filled squares represent the measured value at the station.	36
5.7	Seasonal residual kriging 1989: (a) MAM, (b) OND.	36
5.8	Seasonal precipitation 2005: (a) MLR MAM, (b) MLR OND, (c) final estimation MAM, (d) final estimation OND. The filled squares represent the measured value at the station. The dashed white line represents the area where negative precipitation is estimated.	38
5.9	Seasonal residual kriging 2005: (a) MAM, (b) OND.	38
5.10	Seasonal precipitation 2006: (a) MLR MAM, (b) MLR OND, (c) final estimation MAM, (d) final estimation OND. The filled squares represent the measured value at the station.	39
5.11	Seasonal residual kriging 2006: (a) MAM, (b) OND.	39
5.12	Semivariogram for seasonal residuals 1977 and 1989: (a) MAM 1977, (b) OND 1977, (c) MAM 1989, (d) OND 1989.	40
5.13	Semivariogram for seasonal residuals 2005 and 2006: (a) MAM 2005, (b) OND 2005, (c) MAM 2006, (d) OND 2006.	40
5.14	Semivariogram for annual residual precipitation 2005 and 2006: (a) 2005, (b) 2006.	41
5.15	Annual precipitation 2005 and 2006: (a) MLR 2005, (b) MLR 2006, (c) final estimation 2005, (d) final estimation 2006. The filled squares represent the measured value at the station.	42
5.16	Annual residual kriging: (a) 2005, (b) 2006.	42
5.17	R^2 between observed value and predicted values of a Leave-One-Out Cross-Validation for every station between 1977 to 2014: (a) MAM, (b) OND, (c) annual.	43
5.18	Comparison of OND 2014 estimation with the IMERG satellite product, rescaled to 0.1° . (a) IMERG OND, (b) final estimation OND, (c) difference between IMERG and final estimation in mm (right colour scale). Black triangles denote manual rain gauges, crosses denote AWS in the densely sampled area and black squares denote AWS in less dense sample area.	44
5.19	Difference in precipitation between AWS data and model estimation for OND 2007 to 2014. Green dots show stations located in data dense region ("inside"), orange dots denote data in extrapolation areas ("outside").	45
5.20	Trends for total precipitation in % per decade for the period 1979 to 2011 and relative to a reference period from 1981 to 2010. Crosses denote all stations that have been used for the estimated maps over the entire period. (a) MAM (b) OND, (c) Annual.	46
B.1	Final estimations of the MAM rainy season from 1977 to 1988. Dots represent the observations used for the estimation.	62
B.2	Final estimations of the MAM rainy season from 1989 to 2000. Dots represent the observations used for the estimation.	63
B.3	Final estimations of the MAM rainy season from 2001 to 2012. Dots represent the observations used for the estimation.	64
B.4	Final estimations of the MAM rainy season from 2012 to 2014. Dots represent the observations used for the estimation.	65
B.5	Final estimations of the OND rainy season from 1977 to 1988. Dots represent the observations used for the estimation.	66
B.6	Final estimations of the OND rainy season from 1989 to 2000. Dots represent the observations used for the estimation.	67
B.7	Final estimations of the OND rainy season from 2001 to 2012. Dots represent the observations used for the estimation.	68

B.8	Final estimations of the OND rainy season from 2013 to 2014. Dots represent the observations used for the estimation.	69
B.9	Final estimations of annual precipitation for 1977 to 1988. Dots represent the observations used for the estimation.	70
B.10	Final estimations of annual precipitation for 1989 to 2000. Dots represent the observations used for the estimation.	71
B.11	Final estimations of annual precipitation for 2001 to 2012. Dots represent the observations used for the estimation.	72
B.12	Final estimations of annual precipitation for 2012 to 2014. Dots represent the observations used for the estimation.	73

List of Tables

5.1	Changes of daily data after quality control. Value indicates number of adjusted/set to missing days.	28
5.2	Mean correlation for total precipitation and the explanatory variables at different windowsizes for the period of 1977 - 2014. The numbers indicates the window size used for the calculation. Al is the abbreviation for altitude, S for slope, Lat for Latitude and Long for Longitude.	31
5.3	Model Selection for the different periods MAM, OND and annual and a choice of 7 different models. The highlighted model is the final model for every period. Abbreviations of predictors are: Al = Altitude, S=Slope, SAs = Slope · Aspect, LL = Latitude and Longitude, MK = Mt.Kenya Variable using one harmonic oscillation, 2MK = Mt. Kenya Variable using two harmonic oscillations. MAE is given in [mm].	32
5.4	Pearson-Correlation of OND total precipitation and explanatory variables at different windowsizes for the years 1977, 1989, 2005 and 2006. The numbers indicated the size of the window used for the calculation. Al is the abbreviation for altitude, S for slope. Italic values are significant at a 0.01 level.	33
5.5	Adjusted R^2 , predictive R^2 and MAE for the years 1977, 1989, 2005 and 2006. For annual estimations only the years 2005 and 2006 are shown.	34
5.6	Simple Linear Regression between final model estimation, IMERG and AWS for OND 2014. Model vs. satellite value is calculated for a resampled model to a 0.1 ° resolution. The comparison with AWS data is based on 14 observations and the resp. extractions from model or satellite map.	44
5.7	RMSE of AWS observations with final estimations for the years 2007 to 2014. Values in brackets denote RMSE calculated for all stations.	45
A.1	Automatic weather stations with coordinates (UTM Zone 37 South) and period of measurements.	57
C.1	Adjusted and predictive R^2 for all years and periods.	76

List of Abbreviations

AICc	Akaike Information Criterion with correction for small sample size
AIC	Akaike Information Criterion
Al	Altitude
AWS	Automatic Weather Station
CETRAD	CEnter for Training and integrated Research in ASAL Development
CV	Cross Validation
DEM	Digital Elevation Model
GLS	Generalized Least Squares
GPM	Global Precipitation Climatology Centre
GPCC	Global Precipitation Mission
ENSO	El Niño Southern Oscillation
IDW	Inverse Distance Weighting
IMERG	Integrated Multisatellite Retrieval for GPM
ITCZ	Inner Tropical Convergence Zone
KFS	Kenyan Forest Service
KMD	Kenyan Meteorological Department
KWS	Kenyan Wildlife Service
LOO-CV	Leave One Out Cross- Validation
LL	Latitude Longitude
MAM	March April May
MDS	Maximum Downhill Slope
MAE	Mean Absolute Error
MK	Mount Kenya Variable
MLR	Multiple Linear Regression
OK	Ordinary Kriging
OLS	Ordinary Least Squares
OND	October November December
PCA	Principal Component Analysis
PRISM	Parameter-elevation Relationship on Independent Slope Model
RK	Regression Kriging
RMSE	Root Mean Square Error
S	Slope
SAs	Slope Aspect
SST	Sea Surface Temperature
WLS	Weighted Least Squares
WMO	World Meteorological Organisation

Dedicated to my parents and my tupper ware group.

Chapter 1

Introduction

1.1 Motivation

Interest in East African rainfall has been rising since the beginning of this century, mainly because of the consequences of failing rains, followed by droughts, and at times leading to famines, or on the contrary heavy rains and devastating floods (Gamoyo et al., 2015). Rains in East Africa are mainly of a bimodal regime with rains in boreal spring and autumn, related to the movement of the Inter Tropical Convergence Zone (ITCZ). The failing of a rainy season implicates prolonged dry seasons and exerts considerable pressure on people and their livelihoods. Such temporal water scarcity or the occurring floods are especially hitting regions where economies and food security highly depend on rain-fed agriculture, such as in Central Kenya.

In the region around Mt. Kenya, the high amounts of rain due to orographic enhancement (1500 mm/year on western slopes) are important for people in its surrounding as well as for people on the adjacent Laikipia plateau and in the lowlands further north and west. These regions are rain sparse, reaching 350 mm/year (Fig. 2.2). People depend on river water from catchments up the mountains. Changes in this vital river water discharge are attributed to increased water abstractions upstream. The amount of river water abstraction increased two to eightfold in the years 1960 to 2004. Small scale and subsistence farmers have become more dependent on river water due to changes in their agricultural practices and in their crop variety (Liniger et al., 2005). When the rains set in late, or when a dry period is especially pronounced, the risk of dried out river beds becomes imminent. People and livestock pushing upstream are intensifying the conflict between upstream and downstream users.

In order to limit consequences of floods and drought and to reduce conflicts between upstream and downstream users a better understanding of precipitation pattern is necessary. On the one hand, a temporal understanding is important, as water scarcity in East Africa is mainly of temporal manner. On the other hand, a spatial picture is important. A spatial map of rainfall in this region may be supportive for decisions in agriculture, for policy makers, for water user associations and for hydrological studies. Spatial information is available on larger scales, but regional to local studies are sparse (Funk et al., 2015). Hence, this thesis aims at providing rainfall maps on a local to regional scale by exploiting the relation between topography and precipitation.

The topography of the study area is rather unique with the single mountain peak Mt. Kenya followed by a high plateau. The relation between topography and precipitation is used manifold for the estimation of precipitation at unsampled locations (Goovaerts, 2000; Daly et al., 2008; Marquinez et al., 2003; Masson and Frei, 2014). Topographic information is easily available, highly resolved and continuously in space. However, it has not been used as explanatory variable for the precipitation pattern around Mt. Kenya. Local expertise about rainfall patterns can be integrated to refine

these relations. Along with the purpose of estimation, the rain-topography relation is worth exploring, as it might improve understanding of precipitation mechanisms and provide a tool for more accurate climatologies in regions with few observational data.

Often, in mountainous and especially in developing regions, observational networks are sparse. This is not the case around Mt. Kenya. A comparatively dense network, comprising data of different governmental institutions and private efforts exists. This thesis profits of a good network, a characteristic topography and available local expertise, in order to produce rainfall maps on a regional scale for seasonal and annual sums and to explore the relation between precipitation and topography.

1.2 State of Research

Devastating floods mainly occur in the boreal autumn rainy season, whereas failing rains appear in both rainy seasons, revealing distinct mechanisms for both rainy seasons (Hastenrath et al. (2011), Gamoyo et al. (2015)). Consequently, a lot of research has been focusing on identifying the mechanisms of the two main rainy seasons, on their interannual variability and its link to the large-scale (Hastenrath et al., 2011; Camberlin et al., 2009; Mutai and Ward, 2000; Nicholson, 2015). For the boreal autumn rainy season, a tight connection to the zonal circulation cell of the Indian Ocean has been identified, with high correlations for the equatorial surface westerlies. This connection is much weaker for the boreal spring rains. The surface westerlies are accelerated during positive ENSO phases due to an increased pressure gradient between East Africa and Indonesia (Hastenrath et al., 1993). Recent studies indicate different relationships between the zonal winds and wet or dry boreal autumn rains. The zonal winds seem to be more important for wet boreal autumn seasons, than for dry seasons (Nicholson, 2015). Less is known about the boreal spring rainy season and less distinctive connections have been found. Still, they are important, because they are the main rainy season for most parts of Kenya (Camberlin et al., 2009) and more so, because trends indicate a decrease in boreal spring rains. A study based on the rain gauge network around Mt. Kenya shows the decreasing trend for boreal spring rains (Schmocker et al., 2015). This is further confirmed by Funk et al. (2015) with trend calculations based on a newly aggregated gridded precipitation dataset, CenTrends for whole East Africa. This publicly available dataset is a result of pooling different station archives and kriging of seasonal anomalies.

For the study area around Mt. Kenya, a spatial interpolation of rainfall on monthly and annual time scales for the period 1967 to 1997 has been conducted by Sturm (2002) using regularized splines with tension, but without incorporating secondary information. These rainfall maps show a sector of increased precipitation for the leeward western side of Mt. Kenya (location of Nanyuki is found in Fig. 2.2). For a region comprising Kenya, Uganda and Tansania, the spatial pattern of rains and its interactions with topographical and geographical parameters has been analysed on a basis of monthly means (Oettli and Camberlin, 2005; Hession and Moore, 2011). Oettli and Camberlin (2005) performed a regression analyses for East African rainfall using topographic predictors, such as altitude, slope and principal components of topography at different scales. They find exposition of slopes being an important predictor and show that east-facing stations are in general wetter for short rains, whereas southern rain gauges show more precipitation throughout the year. Hession and Moore (2011) use spatial regression with stepwise forward selection to find adequate predictors. They find altitude to be relevant for boreal summer months on small scales and for boreal spring and autumn on very large scales.

The relation between altitude and precipitation has been reported manifold to be relevant for predictions of precipitation and comparison studies have been conducted for distinct regions and different interpolation schemes (e.g. Goovaerts, 2000; Bostan et al., 2012). In general, precipitation is

increasing with altitude up to the point where the free atmosphere is reached (Barry, 2008). The PRISM (Parameter-elevation Relationship on Independent Slopes Model) is an example of a model for climatological interpolation of temperature and precipitation. As the name reveals, it is based on the relation of elevation with temperature and with precipitation. PRISM further accounts for topographic factors by weighting rain gauges for example considering the atmospheric layer a rain gauge is located in or its topographic similarity (Daly et al., 2008).

Other than altitude, precipitation can exhibit a complex pattern entailed by various causes such as barrier effects of mountains, exposition of slopes relative to prevailing wind directions or the distance to a moisture source (Barry, 2008). For example, Kyriakidis et al. (2001) build an additional rainfall predictor composed of wind directions and specific humidity from an assimilated data product in order to account for moisture being subject to orographic lifting. They compare different interpolation schemes, some incorporating secondary information and some only relying on the spatial relation of observations.

1.3 Research Objectives

The importance of Mt. Kenya on the precipitation pattern has been recognized severally (Berger, 1989, Notter et al., 2007, Sturm, 2002). However, up to now, no spatial prediction using topographic information has been performed. Spatial information of rainfall is valuable as it provides practical information to local communities and exploring the rain-topography relation may improve understanding of local mechanisms between Mt. Kenya, prevailing winds and the subsequent rainfall pattern. Thus, the four main objectives of this thesis are composed of field work in the Mt. Kenya region and the subsequent analysis of the rainfall data. They are described briefly in the following:

1. Update of the rainfall database and assessment of quality of all still operating rain gauges.
2. Estimation of gridded seasonal and annual total precipitation maps for the period 1977 to 2014 using topographic variables. The explanatory variables are chosen based on what is known from the Mt. Kenya region and the relation between topography and precipitation:
 - Mt. Kenya acts as a barrier to prevailing winds and moisture transport, enhancing orographic precipitation on windward sides of the mountain.
 - The precipitation pattern is related to altitude, aspect and slope.
 - Geographic variables, such as latitude and longitude may help to describe the occurring gradients with a semiarid north and a humid southern zone in the study area.
3. Comparison of the estimated rainfall maps with two independent datasets:
 - Integrated multi-satellite Retrieval for GPM (IMERG) for the short rains 2014
 - Automatic weather station data
4. Improvement of understanding of the mechanism influencing the precipitation pattern

The thesis is structured as followed: Chapter 2 will give a more indeep introduction to the study area and its peculiarities, in Chapter 3 origin and quality of the data is described, Chapter 4 will explain the method of rainfall estimation, Chapter 5 and Chapter 6 will show results and discuss them. A conclusion including ideas for further studies is provided in Chapter 7.

Chapter 2

Study Area

The study area located in the tropics exhibits a bimodal annual cycle in precipitation with areas comparatively dry for their latitude. The characteristic topography for the area shows a major influence on local to regional rainfall variation due to its effects on low-level atmospheric circulation (Yang et al., 2014a). East African rains are subject to large-scale forcing, such as the ITCZ and the pressure gradient over the Indian Ocean. In the following chapter, this characteristic topography, the annual cycle of precipitation and the connection of East African Rainfall to the zonal circulation are described briefly in order to provide a general understanding of the climatology in the study area.

2.1 Location and Topography

The study area is located in the central part of Kenya, reaching from 36.2° E to 38° E and 0.7° S to 1.3° N (Fig. 2.1, a). It is situated to the east of the rift valley, which cuts Eastern Africa in a north to south-eastern direction. To the east of the study area, Kenya borders the Indian Ocean, which is of significance for the area as a moisture source during the rainy seasons. As part of the eastern edge of the rift valley, the study area exhibits a peculiar topography. The highest peak, Mt. Kenya reaches an altitude of 5199 m a.s.l. To the east, altitude decreases to sea level at a distance of around 350 km. North and west, the slopes of Mt. Kenya end on the high plateau of Laikipia. In the south, this high plateau ends with the Aberdare Mountain range, reaching altitudes of around 4500 m a.s.l. The Laikipia plateau itself is a gently undulating semi-arid plateau with altitudes of 1700 – 1900 m a.s.l., interrupted to the north by the Loldaiga Hills and further north and east, rapidly decreasing into the arid lowlands with altitudes of around 800 m a.s.l (Liniger et al., 2005). In the lowlands further mountain chains can be found such as the Matthew Ranges extending in a north-south direction with altitudes up to 2800 m a.s.l.

Ecologically, the area exhibits steep changes from semi-arid zones in the northern part to humid zones along the mountain slopes spanning only a short distance of 100 km (Liniger et al., 2005). These steep changes are attributed on the one side, to the water balance with high evaporation and low precipitation in the northern, lower areas and on the other side, to topography enhancing orographic precipitation (Winiger, 1981).

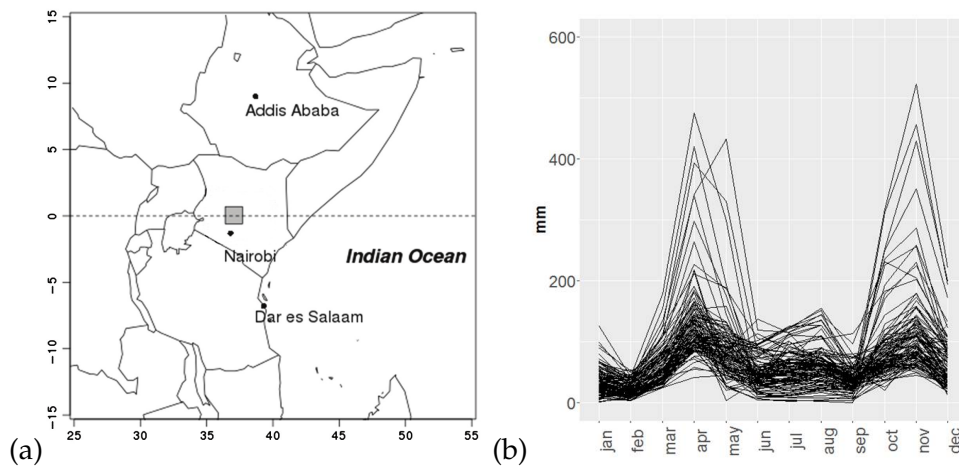


Figure 2.1: (a) Location of study area in East Africa (grey square). Source: Schmocker et al. (2015). (b) Mean annual precipitation for all stations in the study area for the period from 1981- 2010. The two peaks show the two main rainy seasons in boreal spring and boreal autumn.

2.2 Climatology

2.2.1 Spatial Pattern of Precipitation

The distribution of rainfall strongly follows topographical features. Areas on the slopes of the Mt. Kenya can receive up to 3000 mm per year, whereas areas located only 200 km north, at altitudes of 800 m a.s.l receive around 350 mm per year (Fig. 2.2). The rainy seasons contribute differently to the total amount of precipitation for distinct regions. North-eastern areas receive more rainfall during the boreal autumn rainy season in October-November-December. Southern areas and the central part of the Laikipia plateau receive more rain during the boreal spring, i.e. the March-April-May rainy season. In boreal spring, rains are in general more abundant and boreal spring is the main agricultural period for most of Kenya (Gamoyo et al., 2015). Western parts of the area receive a third rainy season from June to August related to the eastern branch of the West African Monsoon (Yang et al., 2014a; Gichuki et al., 1998). Rain gauges in these areas due not exhibit the distinct bimodal rainfall regime (Fig. 2.1, b).

Rainfall amounts around Mt. Kenya differ depending on the exposition of the mountain. South-eastern slopes of Mt. Kenya receive highest amount of precipitation during the boreal spring rainy season with a peak of precipitation at an altitude of 2600 m a.s.l., based on measurements for a period from 1951- 1952 (Fig. 2.3). The western slopes show a maximum occurring at an altitude above 3000 m asl. According to Winiger (1981), precipitation maxima occur as a mixture of slope heating and moisture advection around the mountain and secondary precipitation maxima resulting from a combination of these two processes exist. North-western slopes are driest and receive most of their rainfall during the boreal autumn rains. On eastern slopes, precipitation is decreasing with increasing altitudes with a maximum of 1800 mm at the foot slopes (Barry, 2008).

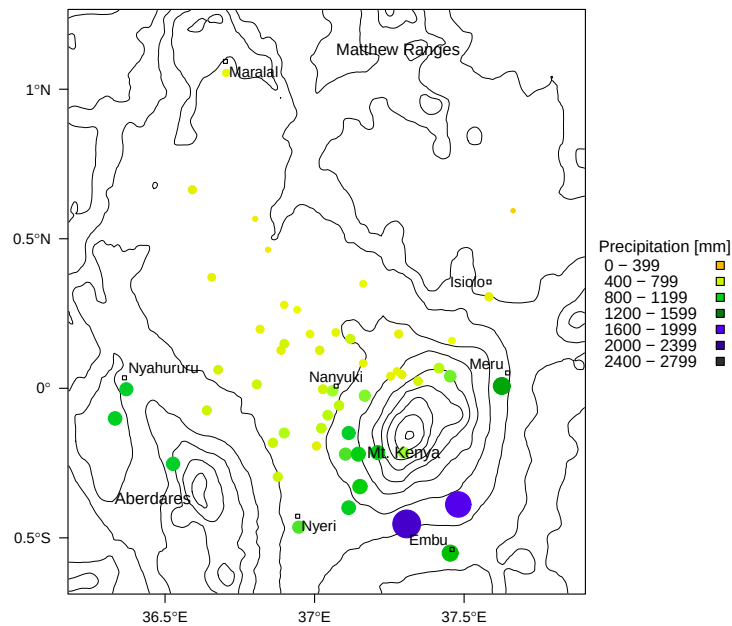


Figure 2.2: Mean annual rainfall for the period 1981- 2010 with topography of the study area (400 m equidistance). Colors and size of bubbles denote the amount of rainfall.

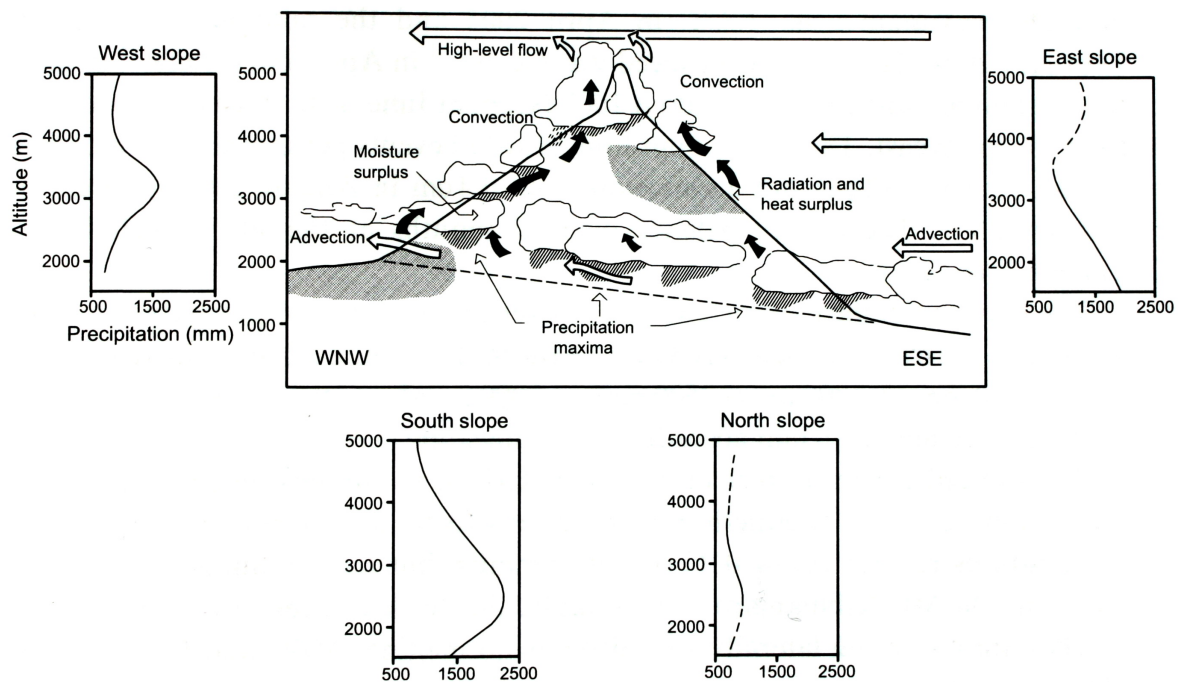


Figure 2.3: Total precipitation along vertical transects for different expositions of Mt. Kenya and a schematic view of moisture advection, circulation, cloudiness and precipitation. Source: Barry (2008).

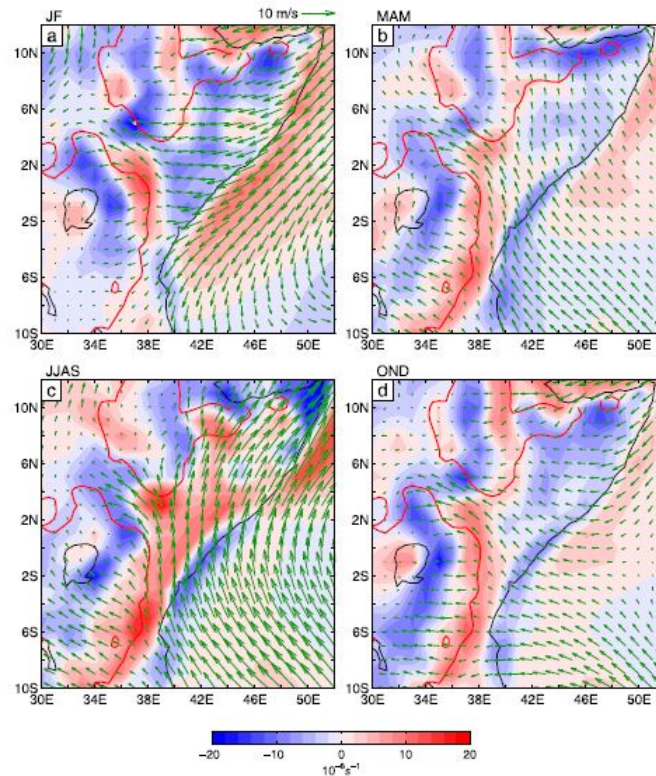


Figure 2.4: Seasonal wind at 10-m (vectors) and divergent (red) and convergent (blue) areas over East Africa from ERA-Interim for the period 1979-2009. The red line depicts the smoothed 1000-m contour. (a) January-February, (b) March-April-May, (c) July-June-August-September, (d) October-November-December. Source: Yang et al. (2014a)

2.2.2 Annual Cycle of Precipitation

The bimodal rainfall regime with peaks in boreal spring and boreal autumn is related to the movement of the Intertropical Convergence Zone (ITCZ) throughout the year. Along the ITCZ, convection cells are formed, transporting water vapour and heat upwards and bringing rain to the underlying areas. The rising air in the convection cells is replaced by trade winds blowing into the low pressure area from north and south east. Throughout the year, the position of the ITCZ changes following the zenith of the sun and the convection belt shifts its location from over the equator to a northern position in boreal summer and a southern position in boreal winter (Berger, 1989). The trade winds, the ITCZ and a counter current in the upper air with sinking air form together the schematic Hadley Cell, which is one of the schematic components of tropical rainfall.

In boreal summer, the ITCZ with its low pressure system reaches over land its northernmost position over India with the resulting Indian Summer Monsoon and south-western winds over the Indian Ocean and East Africa. In the two transition phases between winter and summer monsoon with the ITCZ positioned at the equator the two rainy seasons are found in East Africa (Hastenrath et al., 2011). During the Indian Winter Monsoon, north-easterly winds with dry and cold air are prevailing over the study area (Fig. 2.4, a). They are in general convergent, but not able to bring precipitation, mostly due to the stable atmosphere and shallowness of the convergence layer (Yang et al., 2014a). In boreal spring, the Long Rains (MAM) set in. The dominant wind directions changes to south-east bringing moist and warm air from the south-western Indian Ocean (Fig. 2.4, b). SST of the adjacent Indian Ocean reach their annual maximum at that time and the winds are able to take up and transport high amounts of moisture.

In boreal summer, the south-eastern wind prevail, but turn east after crossing the equator (Fig. 2.4, c). SSTs off the coast of East Africa show their lowest values in the course of the annual cycle. Winds are correspondingly cold and dry and stable air is advected over East Africa. At that time, the West African Monsoon reaches its easternmost extension, bringing rain to the western and north-western part of the study area (Yang et al., 2014a).

In the monsoon transition phase from summer to winter monsoon, meaning in boreal autumn, the Short Rains (OND) set in. The windfield becomes again similar to boreal spring with south-easterly winds, however stronger north-easterlies prevail in the northern parts of Kenya and East Africa (Fig. 2.4, d). Precipitation is in general lower than for the boreal spring Long Rains, possibly due to cooler SSTs at that time (Yang et al., 2014a).

2.2.3 Interannual Variability of Rainy Seasons

The two main rainy seasons have rather disparate characteristics and are lead by distinct mechanisms (Camberlin et al., 2009; Gamoyo et al., 2015). The rains occurring in boreal autumn, in the following termed as OND, receive in general less precipitation, compared to the rain occurring in boreal spring (Gamoyo et al., 2015). However, floods (1961, 1994, 1997, 2006) and droughts (2005, 2008, 2011) tend to be more frequent during boreal autumn compared to boreal spring rains. Correspondingly, the OND rains show a higher interannual variability than the rains in MAM (Hastenrath et al., 2011). This high interannual variability is related to changes in the equatorial zonal circulation cell that develops over Indonesia and East Africa and is thus related to El Niño Southern Oscillation (ENSO) and the pressure gradient between the East and West Pacific (Fig. 2.5). The equatorial zonal circulation cell consists of ascending air over Indonesia, subsidence over East Africa, and related westward motion in the upper troposphere and eastward winds at the surface (Hastenrath et al., 2011). Highest correlation for the OND rains is found for the equatorial zonal winds over the Indian ocean, however this relation is recently found to be time dependent, and changes for wet and dry years (Nicholson, 2015). In comparison, the MAM rains are not correlated with the equatorial westerlies. No equatorial zonal circulation does develop during boreal spring, although the strong surface westerlies are found for boreal spring over the equatorial Indian ocean. According to Hastenrath et al. (2011), no zonal circulation cell develops, because upper tropospheric easterly flow is impeded by a low stand of upper tropospheric topography related to the prevenient boreal winter.

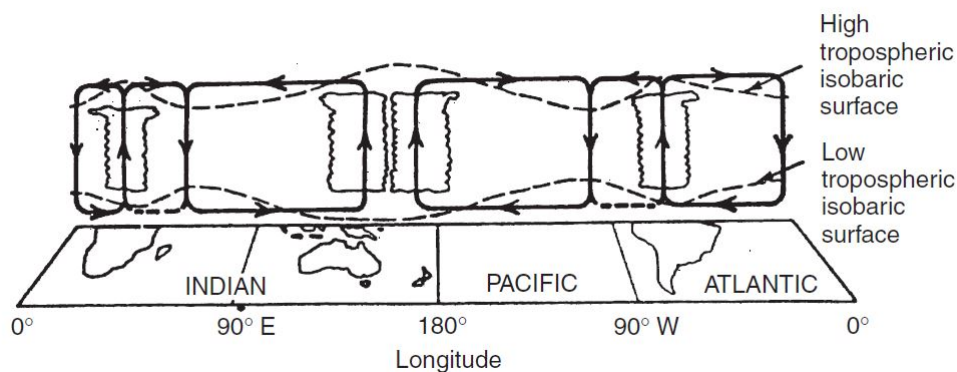


Figure 2.5: Schematic view of the zonal circulation along the equator leading to subsidence over East Africa and ascending motion over Indonesia. Source: Lau and Yang (2002)

Chapter 3

Data

Rainfall estimation in this study is conducted based on a dataset of ordinary manual rain gauges. The dataset is not split into a testing and a validation dataset, as observations are already rather sparse. A comparison is performed on independent datasets; a set of automatic weather stations and the newly available satellite product IMERG from GPM. These datasets, the quality assessment of the rain gauge stations and quality control of the data are described in the following section.

3.1 Manual Rain Gauge Data

3.1.1 Description of Network

A database for daily precipitation measurements has been established and maintained by the research institution CETRAD in Nanyuki. This database comprises data collected by governmental institutions, such as the Kenyan Forest Service (KFS), Kenyan Meteorological Department (KMD) and private farms. In total, the database contains data of 164 manual ordinary rain gauges, all of them exhibiting different length of time series and covering different time periods. The longest time series dates back to 1934 and covers a period of 81 years.

Since the beginning of measurements, the number of stations in the area has been changing (Fig. 3.1, a). This change of the network over time will exert influence on estimations of precipitation. To model rainfall amounts around the mountain, a certain number of stations covering different orientations of the mountain are necessary. Similarly, only very few stations in the northern area, in the western, or at higher altitude will influence the outcome of precipitation estimation. A major amount of observations are taken around the 1990's by CETRAD, Nanyuki. These rain gauges cover a short period of up to 5 years only, but are still used in the analysis. The reduction of stations in the last decade is due to a loss of interest in the KFS network. Broken rain gauges have not been replaced and no measurements are taken anymore. The rain gauges are mainly of a Cassella type, but are slowly being replaced by similar types of rain gauges made of plastic (Fig. 3.1, b). Very few rain gauges are autographic, converting accumulated precipitation from weight to mm. An overview of the measurement stations used for analysis is listed in Appendix A. For this study, a period from 1977 to 2014 is chosen, as for this period in every year at least 40 rain gauges with measurements are available. This is a very conservative choice, led by the number of rain gauges and the amount of predictors used for the regression models. In the view of obtaining climatological long-term fields and the fact that the first satellite products for rainfall are appearing shortly after 1977, extending the maps further back would be desirable.

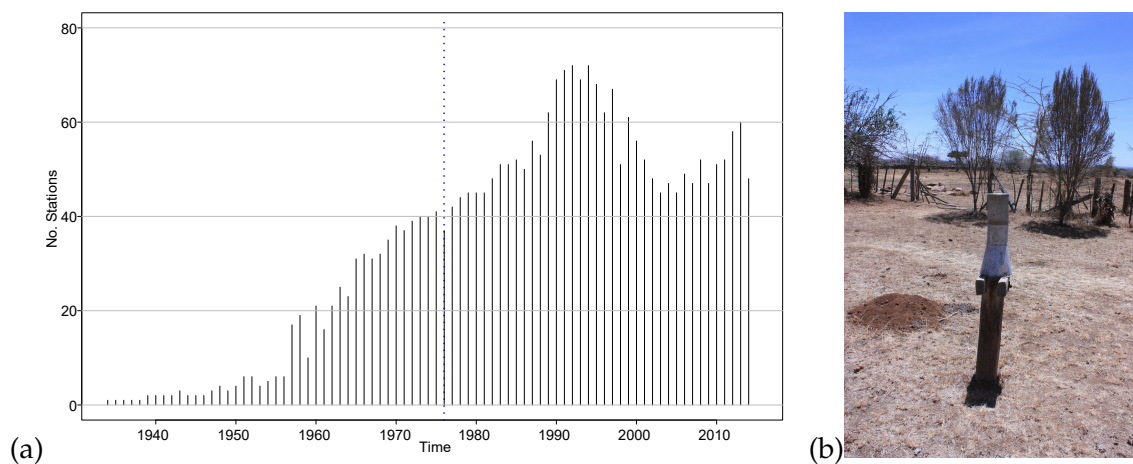


Figure 3.1: (a) Evolution of the measurement network since the beginning of observations. All years succeeding the dotted blue line are used for the estimations. (b) Example of an ordinary manual rain gauge used in the study area. Ol Maisor Farm, 2015 (own illustration).

The measurement stations in the study area are also distributed rather unevenly in space (Fig. 3.2, a). On the Laikipia plain and along the northern and western foot slopes of Mt. Kenya, a lot of stations are present, whereas east of Mt. Kenya and in the northern part of the area the network is much less dense.

In addition to this geographically inhomogeneous distribution, the rain gauges do not entirely represent topography. More rain gauges are located in areas with slopes of around 3.5 - 4 %, whereas the study area exhibits less gridcells of these slope values (Fig. 3.2, b). Altitudes of 1500 to 2000 m a.s.l. are overrepresented by rain gauges in comparison to the study area (Fig. 3.2, c). Thus, estimations are biased towards steeper slopes and medium altitudes. On average, the station density is around 1 station per 727 km², which is a lot higher than the reported density of 1 station per 10'483 km² in the CenTrends dataset (Funk et al., 2015). Hence, the herein used dataset is suitable for a local to regional analysis.

3.1.2 Quality Assurance

Every few years, the operational stations are visited in order to check if the rain gauges are still operating, to collect rain data, to assess the station's set up and the reading behaviour of the station's supervisor. During the last assessment, 52 rain gauges were operational and their data had been updated to 2015.

According to the WMO guide (WMO, 2008), frequently occurring measurement errors for non-recording ordinary rain gauges are errors (a) due to systematic wind field deformation above the rain gauge, (b) due to wetting loss on internal walls of container, (c) due to evaporation from container, (d) due to blowing and drifting snow, (e) due to in- and out-splashing of water and (f) random observational and instrumental errors. An assessment of the rain gauge, the location and of observation practices is made in order to assess the occurrence of errors, such as error source (f). A visit of the measurement site can help to detect, whether errors (a) or (e) may affect measurement, but no quantification is possible. Assessing reading behaviour is important, because insufficient knowledge may lead to mistakes when digitizing the data and eventually to asynchronous time series. Precipitation measurements are particularly sensitive to exposure, topography and wind

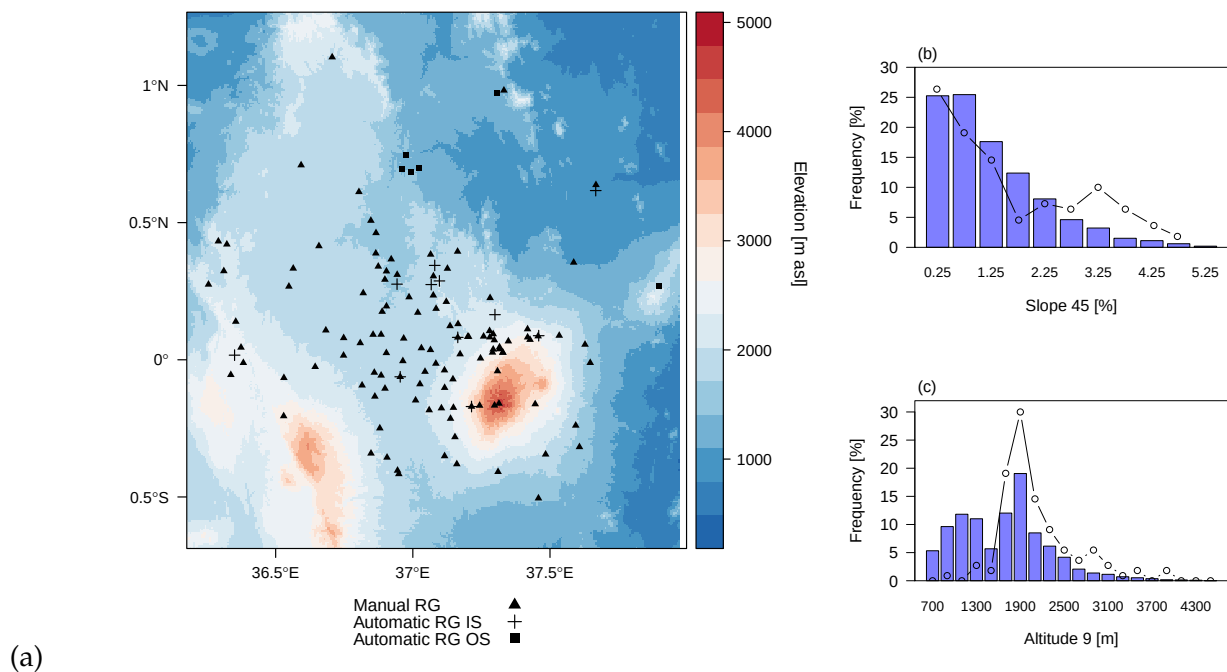


Figure 3.2: (a) Map of manual and automatic weather stations on a digital elevation model of the area. IS denotes automatic weather stations in densely sampled parts of the study area and OS in data sparse areas. (b) Percentage distribution of stations (dashed line) and grid points for slope at a 45 cell window size. (c) Percentage distribution of stations (dashed line) and grid points for altitude at a 9 cell window size.

(WMO, 2008). Therefore, metadata is particularly important and in addition to the rain data, information such as relocations of rain gauges and measurements device is collected.

The quality assurance of the stations is composed of three categories:

- **Condition of the rain gauge:** The rain gauge itself is checked. Is the funnel still intact? Is a correct measuring cylinder used? Has the rain gauge been relocated since the last visit? Has the type of rain gauge changed?
- **Environment of the rain gauge and set up:** What are the closeby obstacles, how far away and in what angle to the rain gauge are they located? How high are they?
- **Reading behaviour:** What time of the day is the measurement taken? Is it taken every day? Are measurements taken on weekends? Is the measurement noted on the day of reading or is it noted for the past day?

In addition to the quality assessment of the field trip, assessments performed in Sturm (2002) and Schmocker (2013) and the amount of daily gaps in each time serie are used for a final quality attribution.

3.1.3 Quality Control

Up to now, the complete database has been used for analysis only sporadically. Single rain gauges are used for hydrological analysis and two major studies have been conducted using a large part of the database. Sturm (2002) interpolated monthly and yearly precipitation and Schmocker (2013)

analysed trends in long-term time series. For this purpose, 37 time series were quality checked and tested for homogeneity by Schmocker (2013).

For this study, rain gauges exhibiting shorter time series are used likewise. The complete dataset is quality controlled in order to detect erroneous values. Daily data is controlled and compared to the nine closest rain gauges exhibiting similar recording periods for:

- Unrealistically high values with an absolute threshold of 300 mm per day
- Unusually high values
- Detection of row of constant non-zero values
- Unusual zero values (where neighbouring stations show high values)

Monthly data is compared to the nine closest rain gauges exhibiting similar recording periods for:

- Unusual monthly zero values
- Number of wet days
- The ratio of total monthly precipitation

Ratios are transformed and flagged as suspicious, if a ratio is below or above four standard deviations of the mean. A monthly value is set to NA if one daily values is missing. A visual control identifies further problems of the dataset, such as excessive rounding, missing or rounded low values, weekly cycles or double values in precipitation occurrence due to missing weekend readings. In the course of regression analysis, further outlying values have been excluded from the dataset.

3.2 Automatic Weather Stations

3.2.1 Network Description

Since 2007, CETRAD and the Syngenta Foundation have operated automatic weather stations (AWS) in the study area (Fig. 3.2). Due to server maintenance at the moment, most of the data is only available as daily sums for a time period from 00:00 to 23:59. Manual data is measured at 8 or 9 am. For this study, this poses only a smaller concern because seasonal and annual values are used. In addition to the manual rain gauges, Ol Jogi Farm set up some automatic rain gauges starting in 2007. These automatic rain gauges are used together with the AWS data. In total, 15 rain gauges are used for comparison with the final rainfall maps. In the first five years around four to seven automatic rain gauges are operational. For 2012 to 2014 around 10 to 14 automatic rain gauges can be used for comparison per year. Because the AWS data exhibits a lot of missing values, all of the stations are used for comparison with the rainfall maps, even though some locations are very close. Six automatic rain gauges are located off the dense rain gauge network and can therefore assess the ability of extrapolation of the estimations. These automatic rain gauges are denoted "OS" indicating that they are located in the sparsely sampled part of the study area. "IS" denotes AWS located in the densely sampled area.

3.2.2 Quality Control

A brief quality control is done with the automatic data on a daily basis. The data is screened for unrealistically high daily values, compared to neighbouring stations and visually checked for discrepancies. No monthly quality control is conducted, because measurement periods are rather

short. No quality assessment of the AWS (assessment of the station's set up) has been conducted, as they are visited and checked regularly by CETRAD or Syngenta employees.

3.3 Global Precipitation Mission Satellite Data

The Global Precipitation Mission (GPM) is a recent project with the GPM Core Observatory and operational satellites from GPM partners that provides a high temporal and spatial resolution of precipitation observation globally from 65° S to 65° N.

The IMERG data are rainfall estimates integrating all passive-microwave instruments in the GPM constellation. The estimates are inter-calibrated with the GPM Combined Instrument and with microwave-calibrated infra-red satellite estimates. The final run is further adjusted to monthly precipitation gauge data and where available other highly resolved precipitation estimates. The IMERG dataset has a resolution of 0.1° for a monthly, daily and 30 minutes time resolution.

As it is a very new dataset, a complete dataset for comparison with the seasonal precipitation estimates of this study is only available since mid of March 2014 (Huffman et al., 2015). Thus, the final run is used for October, November and December.

3.4 Digital Elevation Model

The digital elevation model for this study is retrieved from the ASTER instrument of the Terra Satellite with a 30 m resolution. The DEM is filtered with a focal mean function of a 3x3 neighbourhood in order to remove single missing values or peaks. Further, it is resampled to a grid with a cellsize of 600 m. All calculations in this study are performed on this 600 m cellsize grid (Fig. 3.2, a).

Chapter 4

Methodology

In this study, the method for creating seasonal and annual rainfall maps is a combination of estimating precipitation totals with multiple linear regression (MLR) and improving the estimation with ordinary kriging (OK) of the residuals. This combination of both can be termed a short-cut form of regression kriging (RK). In the following, the coupled method of regression kriging, the model building for MLR with the used predictors and residual OK are described.

4.1 Regression Kriging

Combining the two methods of MLR and OK, we obtain a model of a deterministic (first component, right hand side) and a stochastic component (second component, right hand side):

$$\hat{z}(s_0) = \hat{m}(s_0) + \hat{e}(s_0) = \sum_{k=0}^p \hat{\beta}_k q_k(s_0) + \sum_{i=1}^n \lambda_i e(s_i) \quad (4.1)$$

$\hat{m}(s_0)$ is the proportion of rain explained through topographic variables and $\hat{e}(s_0)$ the unexplained variability of rainfall. In the deterministic part, $\hat{\beta}_k$ are the coefficients estimated in the regression model and q_k are the explanatory variables at the location of interest. Likewise, the kriging part is in the form of a linear predictor, where λ_i are the calculated weights minimizing the kriging system and $e(s_i)$ is the residual precipitation at the observed locations. The second part is termed stochastic, because it is assumed that the generating process is random (Hengl et al., 2007).

Odeh et al. (1995) describe different types of RK. The herein chosen approach is equivalent to their mentioned short-cut form of RK, denoted as Type C. The β - coefficients of the MLR are estimated using ordinary least squares (OLS) and OK is performed on the remaining residuals. It is a short-cut insofar that instead of estimating the coefficients and determining the covariance structure of the residuals to re-estimate the coefficients until they stabilize, only one iteration using OLS is performed.

Splitting regression and kriging into two separate parts, allows to individually assess possible structures of regression models and facilitates to formulate (more complex) forms of models and predictors. Mathematically, methods such as Universal Kriging are equivalent to RK, but they determine the kriging weights directly using the secondary variables (Hengl et al., 2007). Statistically, several drawbacks exist, using a short-cut version of RK. They are not discussed in the scope of this study, but can be found in Schabenberg and Gotway (2005).

4.2 Multiple Linear Regression

In spatial statistics, MLR has a well established position as it allows for estimating values at unsampled locations, for inference about the influence of an explanatory variable on the dependent variable and it may deliver better results for extrapolations, if the relation between the variables holds for the extrapolated area. The basic formulation in matrix form for a spatial MLR model is:

$$\hat{Y}(s) = \beta X(s) + \epsilon(s) \quad (4.2)$$

β is a vector of coefficients, $\hat{Y}(s)$ a vector of the estimations, $X(s)$ is a matrix of explanatory variables and $\epsilon(s)$ is a vector of the residuals at locations s (Burnham and Anderson, 2002).

The main goal of this analysis is to calculate a rainfall map of the study area for every time period individually. Thus, regression analysis is run for every year and season independently and individual β_i coefficients are estimated. When estimating every year and season individually, we can benefit from stations with short periods and include them into the individual regression models. Therefore, every estimation is based on a slightly different station network and the estimated coefficients are a result of the distribution of the stations and the climatic factors at a given time. Explanatory variables for the MLR are selected according to present knowledge of interaction between precipitation and topography. They are described in the following section.

4.2.1 Explanatory Variables

The explanatory variables can be divided into geographic or topographic predictors and an individual predictor variable considering the effect of Mt. Kenya on the precipitation pattern.

Geographic Predictors

To address the gradient between the humid, tropic area around Mt. Kenya and the semiarid areas in the northernmost part of the study area, latitude and longitude based on the 600 m grid are used as predictors. Latitude is somewhat also representing altitudinal changes, because the terrain is decreasing towards north. Longitude may describe the distance to the main moisture source (Indian Ocean), or as well the distance towards the rising of the highland plateau. The relation of latitude and longitude with precipitation may be linear, or as well of a higher order polynomial of both predictors.

Topographic Predictors

Mechanisms of precipitation may operate at different spatial scales. For the study area of around 48'000 km², topographic scales of 9, 21, 33, 39, 45, and 57 gridcells are considered. This is equal to a window size of 5.4 km to 34.2 km. Higher scales are not considered in this study, but their relation to precipitation and the corresponding mechanism might still be important. The following variables are calculated on the 600 m gridsize DEM:

- **Altitude** is derived from the resampled 600 m DEM for the six aforementioned window sizes using a focal mean function. Altitude mainly acts as a surrogate for temperature assuming a constant lapse rate (atmospheric temperature gradient) (Kyriakidis et al., 2001).
- **Aspect** is calculated for the same window sizes. The calculation follows the approach used by Oettli and Camberlin (2005). The area of the respective window size around the target variable is divided into 9 equally sized gridcells, whereof the mean elevations are calculated. On this newly defined 3x3 neighbourhood the aspect algorithm is applied, as described in De By et al. (2001). Aspect is a circular variable. In order to be able to add it in a regression function, it can be written in a linear combination of sine and cosine.

- **Slope** is calculated on different scales following the same approach as for aspect. Additionally, to the 3x3 neighbourhood method for slope calculations described in De By et al. (2001), the method of maximum downhill slope (MDS) discussed in Dunn and Hickey (1998) is used to calculate slope values. According to Dunn and Hickey, 1998, this method better retains local variability of slopes. The MDS method calculates the slope between the target gridcell and the cell of the 3x3 neighbourhood, which exhibits the highest deviation in elevation compared to the target gridcell. For both methods, calculating a 3x3 neighbourhood and averaging elevation eventually leads to averaging of all values. For a window size of 9 gridcells higher slope values are calculated than for a window size of 57 gridcells.
- **Combinations of topographic predictors** are possible, because a predictor may exert influence not in itself but in a combination with another predictor (interaction term). Combinations for slope* aspect, altitude*aspect, latitude*aspect, latitude*slope have been evaluated during model selection. These seem to be the most promising, considering topography and climatological background. An example of such an interaction is that aspect only imposes on the target variable, if a sufficiently high obstacle has to be overcome.

Mt. Kenya Variable

A peculiarity of the region is Mt. Kenya as a massive with one single peak followed by a high-plateau. The mountain seems to have a major topographic influence on precipitation as it diverts low-level atmospheric motion and moisture flow around the mountain or up the plateau and leads to precipitation maxima on the windward eastern slopes of the mountain (2.2). The study of Sturm (2002) shows a second precipitation maximum on the western slopes of the mountain, possibly due to converging air flow on the leeward side of the mountain. A predictor consisting of an angle around the mountain and an exponential function of the distance from the highest point of the mountain is built to account for the effect. Maximum precipitation dependent on the angular position from the mountain can be described with a harmonic oscillation:

$$y(x) = c_1 \cdot \sin(x) + c_2 \cdot \cos(x) \quad (4.3)$$

c_1 denotes the first coefficient, c_2 the second coefficient and x the angle around the mountain. A second harmonic of higher frequency can be used to describe the second maxima at the western slopes of the mountain. This would lead to the addition of another sine and cosine term of $2x$ to equation 4.3. To account for decreasing precipitation, when a certain altitude is reached and for decreasing influence of the mountain further away from the mountain the exponential function in equation 4.4 is used:

$$f(x) = \frac{x + a}{b} \cdot e^{\frac{x+a}{b}} \quad (4.4)$$

The value x denotes the distance to the highest point of the mountain. The values a and b are chosen manually, in a manner to best reflect the altitudinal behaviour of precipitation. According to Gichuki et al. (1998) on western slopes, maximum precipitation of 1600 mm/year occurs at an altitude between 3000 m a.s.l. and 3500 m a.s.l and decreases to around 800 mm/year on top of the mountain. The distance function, as its name says does not depend on elevation, but on the distance to the highest point. Decreasing precipitation is however, rather related to elevation. With $a = 8000$ and $b = 0.3 \cdot 10^4$ the distance function reaches its maximum at around 12.5 km and decreases to a value of half the maximum at the top of the mountain. The 12.5 km are equivalent to an

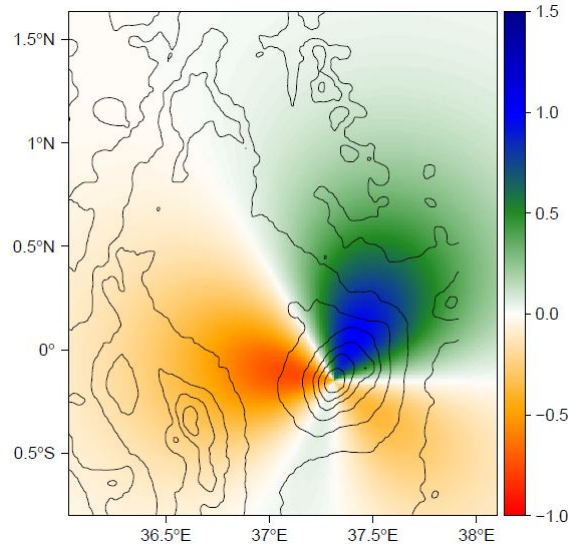


Figure 4.1: Visualization of equation 4.5: Blue to green colours show positive phases of the harmonic oscillations, orange to red show negative phases. The coefficients c_1 to c_4 are set to 1.

altitude of around 3000 m. It is a simplification and ignores the fact that the 3000 m contour line is reached at some points earlier, or later, depending on the shape of the mountain. Assuming, maximum precipitation occurs around the mountain at the same elevation is a further simplification, as combined effects of orographic precipitation and convection on the differently oriented slopes may lead to different maxima or as well altitudinal secondary maxima. The predictor can not account for such possible local effects.

The harmonic oscillation and the distance function are used together to build a distance dependent harmonic oscillation around the mountain, as denoted in equation 4.5. In order to account for two distinct precipitation maxima, in equation 4.5 a second sine and cosine term of $2x$ is added to equation 4.3. d describes the distance to the mountain, x the angle around the mountain. Solving for the regression coefficients, c_1 to c_4 are determined.

$$y(x) = c_1 \cdot d \cdot \sin(x) + c_2 \cdot d \cdot \cos(x) + c_3 \cdot d \cdot \sin(2x) + c_4 \cdot d \cdot \cos(2x) \quad (4.5)$$

The Mt. Kenya variable is thus able to predict different maxima and minima depending on the angle around the mountain and decays with distance from the mountain following the above defined distance function (Fig. 4.1).

4.2.2 Model Selection

Based on the above described explanatory variables, a total of 2^{11} possible models exists, considering only one scale at once and ignoring interactions between variables. A common approach for selecting an appropriate model is stepwise forward or backward selection to evaluate the combination of variables explaining most of the variability in the model, in combination with tolerance controls to avoid collinearity of predictors (e.g. in Marquinez et al., 2003 or Oettli and Camberlin, 2005). Burnham and Anderson (2002) suggest not to test through all possible model combinations, but to formulate a set of plausible models. Here, an a priori model is formulated, integrating the available local expertise and the general understanding of relation between topography and precipitation. A full model with the above described predictors has the structure seen in equation 4.6. β_1 to β_{13} are the coefficients estimated from OLS.

$$\begin{aligned} v = & \beta_1 + \beta_2 \cdot x + \beta_3 \cdot x^2 + \beta_4 \cdot y + \beta_5 \cdot y^2 + \beta_6 \cdot z + \beta_7 \cdot slp \\ & + \beta_8 \cdot d \cdot \sin(a) + \beta_9 \cdot d \cdot \cos(a) + \beta_{10} \cdot d \cdot \sin(2a) + \beta_{11} \cdot d \cdot \cos(2a) \\ & + \beta_{12} \cdot slp \cdot \sin(asp) + \beta_{13} \cdot slp \cdot \cos(asp) \end{aligned} \quad (4.6)$$

with:

- v = Estimation
- x = Longitude
- y = Latitude
- z = Altitude
- slp = Slope
- asp = Aspect
- d = Distance to Mt. Kenya
- a = Angle around Mt. Kenya

A final model has to fulfil several criteria: (1) the explanatory variables should be physically adequate, i.e the coefficient should point in the right direction. For example, decreasing precipitation with altitude or increasing precipitation with latitude would be suspicious. (2) The model should not be more complex than necessary. For a measure of relative complexity the AICc is used, because it corrects for small sample sizes. The calculation of AICc is based on AIC, sample size and number of predictors.

$$AICc = AIC + \frac{2k(k+1)}{n-k-1} \quad (4.7)$$

where k denotes the number of predictor and n the sample size (Burnham and Anderson, 2002). The sample size for this study varies between 40 and 85 rain gauges per year. (3) In order to make predictions, all models need to be valid; they need to fulfil the assumptions a multiple linear regression model is based on. Diagnostic plots and tests are applied for all the individual models to check for patterns in the residuals, e.g. heteroscedasticity or autocorrelation. Further, we can check how much a model overfits calculating the predictive R^2 , which is equivalent to a Leave-One-Out Cross-Validation (LOO-CV). The number of predictors is high compared to the number of observations (i.e. degrees of freedom are low) and the defined secondary variables are not completely

independent. Thus, quantitative interpretation of the coefficients is disadvised. However, these two concerns do not impair the final goal of prediction.

Rainfall in OND and MAM originates from rather different mechanisms, and in addition the continental rains contribute to annual rainfall. Thus, the optimal model is compared for the three time periods individually. Different combinations of the above described full model are tested and the one with the lowest AICc is preferred.

4.2.3 Scale Selection

Pearson correlation between total precipitation and topographic variables at all window sizes and between geographic variables is calculated in order to have an idea of the relation between the variables and precipitation. For topographic variables, the scale with highest correlation value is used in the model, but it is compared to models incorporating other scales. Slope or Altitude can exert influence on precipitation on larger and on smaller scales, and thus a combination of lower and higher scales is tested as well. Correlations with geographic variables give a first idea, whether the expected relation between precipitation and the respective variable is present.

4.2.4 Cross-Validation of the MLR Estimation

In order to obtain a measure of overall accuracy of the estimation, a Leave-One-Out-Cross-Validation is performed for every year. With the predicted values and the observed values a simple linear regression is calculated for every rain gauge individually. The R^2 is a measure of how well the predictions explain the variability in the observations. The slope is a measure of general under- or overestimation during the period from 1977 - 2014. This information can be used to assess the performance of the models at specific locations. No prediction intervals are calculated, because of collinearity between predictors and subsequent inflated variance.

4.3 Ordinary Kriging of Residuals

4.3.1 Theoretical Basis

In order to improve the MLR prediction, the information stored in the residuals can be used. They are interpolated using ordinary kriging (OK) and added to the MLR estimation. The kriging prediction formula is the second part of equation 4.1 and is basically a weighted sum of the residuals:

$$\hat{Y}(x_0) = \sum_{i=1}^n \lambda_i Y(x_i) \quad (4.8)$$

For this linear prediction system, the crucial point is to obtain the weightings λ_i . $\hat{Y}(x_0)$ is the unknown residual value at the unsampled location and $Y(x_i)$ are the sample residuals. The weights are derived by minimizing the kriging prediction variance and will lead to the following matrix equation:

$$\lambda_0 = M^{-1} \gamma_0 \quad (4.9)$$

where λ_0 is a vector of the weights and M a symmetric matrix of the semivariance at different distances (Webster and Oliver, 2007). γ_0 is a vector of semivariances between the cell of interest and the available observation. The derivation of equation 4.9 can be found in Webster and Oliver (2007) and is not further discussed here. Solving equation 4.9 depends on knowing the semivariances $\gamma(h)$ between two points separated at distance h . This is found in the semivariogram and is in practice not given. Thus, a theoretical semivariogram is fitted to an empirical semivariogram calculated from the residual precipitation in order to have a semivariance at every possible lag. The empirical semivariogram uses all calculated semivariances of all point pairs and calculates means of semivariances for point pairs at similar lags (or also called bins). The semivariogram is described by a function of three main variables, the nugget, sill and range and has to be in the form of an authorized models. Description of authorized models for kriging can be found in (Webster and Oliver, 2007). The range describes the distance, at which the semivariance does not further increase. The nugget is the expected semivariance at lag 0 and the sill represents the semivariance when the range is reached. The semivariance for the point pairs is calculated using equation 4.10:

$$\hat{\gamma}(h) = \frac{1}{2} \frac{1}{n(h)} \sum_{i=1}^{n(h)} (z(x_i) - z(x_i + h))^2$$

with (4.10)

$\hat{\gamma}(h)$ = semivariance

$n(h)$ = number of point pairs at lag h

z = variable of interest

We further assume that our process is isotropic. Thus, we can use the same semivariogram independent of direction of the distance.

4.3.2 Implementation

Estimating residual precipitation with ordinary kriging includes four steps: (1) Calculating an empirical semivariogram using equation 4.10, (2) estimating the theoretical semivariogram, (3) calculating the weights according to equation 4.9 (4) and estimating residual precipitation.

In order to avoid fitting every semivariogram for every year and period individually an automated procedure is used (automap package in R). This allows to estimate semivariograms automatically using initial values for nugget, sill and range. These values are calculated following empirical findings (Hiemstra, 2013). The autofitting function iterates over all possible authorized models or the authorized models provided by the user. The model with the lowest residual sum of squares for the empirical semivariogram is returned and used to determine the weights λ . The semivariogram is fitted using weighted least squares with weights according to the number of point pairs in a bin, divided by the squared semivariance of the bin (N_j/h_j^2). j denotes the bin (Hiemstra, 2013). Further, for autofitting a fixed cut-off (maximum separating distance of two residual observations) value of 60 km is selected.

Automatic fits exhibiting errors above 50 mm^4 are fitted manually, determining nugget, sill and range from the empirical semivariogram visually. This procedure is called variography (Webster and Oliver, 2007). It allows to adjust the width of bins (distance between lags) and the cut-off, if necessary. If the new model shows better cross-validation results, a lower error and a reasonable prediction, the new model is used instead of the automatic fit. For years and seasons that do not exhibit any spatial relation inverse distance weighting (IDW) is applied.

After having obtained a semivariogram for every year, the weights for the linear kriging predictor (Eq. 4.8) are determined and the kriging prediction are calculated. This is done in the package *gstat* in R using the *krige* function and will not be further discussed.

4.4 Comparison with Independent Data

The dataset is not divided into a training and testing set, because of the low number of observations. In order to be able to assess the predictions of the model, the data is compared to the above described satellite and AWS data. No comparison can be performed for the years prior to 2007.

4.4.1 Comparison with IMERG Data

The OND 2014 map is resampled to a resolution of 0.1° and compared to the OND IMERG satellite image. Because there is only one image available, no extensive comparison is performed. Differences are calculated in order to see where the model estimations and IMERG product differ. In addition, a simple regression is performed in order to assess the general accordance and possible over- or underestimation.

Several other spatial precipitation datasets for comparison exist for East Africa, such as the before mentioned CenTrends dataset based on interpolated rain gauge data (Funk et al., 2015), or the TRMM dataset based on satellites estimates. These datasets cover longer time periods and would allow for an extended comparison. In the scope of this master thesis, this is resigned. Comparison is focused on the spatially higher resolved IMERG dataset, integrating rain gauge data in order to see additionally how this new product performs for the study area in the tropics.

4.4.2 Comparison with AWS Data

The maps are compared to AWS data starting in 2007 using the root mean squared error RMSE and R^2 . The AWS network evolved rapidly in 2011 with a lot of new stations located at the edge of the study area. Thus, comparing these values e.g. between 2007 and 2014 is disadvised, because the first four years rather tell us about the validity of the model inside the study area, whereas the last 4 years also comprise stations located in extrapolation areas beyond the given data. To compare the AWS and the gridded precipitation dataset, the closest value of the gridcell to the AWS location is chosen. The AWS data for OND is also compared to the IMERG OND using the same method, i.e. comparing the value of the closest gridcell with the observation. Methods, using e.g. weighted averages of the gridcells around could be recommended, but are not applied in this thesis. RMSE is calculated for the group of data present from 2007 to 2014 and separately for all AWS comprising the newly AWS set up starting in 2011.

4.5 Trend Calculations for Total Precipitation

In the masters thesis of Schmocker (2013), a spatial pattern for the trends of total seasonal and annual precipitation has been detected. To be able to further explore this pattern, trends for the precipitation maps are calculated for the period 1979 - 2011. The trends are calculated for annual totals, MAM and OND relative to the climatological mean for the period 1981 - 2010 derived from the precipitation maps.

Chapter 5

Results

The results of this study comprise the quality assessment performed during the fieldtrip and the subsequent quality control, model selection, precipitation maps, comparison with independent data and a short section on the calculated trends. For all three periods, MAM, OND and annual, model selection results are shown and correlation of topographic variables on different window-size are described. For seasonal and annual precipitation, specific examples of estimated years are chosen and described concerning their MLR and OK estimation.

5.1 Quality of Manual Rain Gauge Network

5.1.1 Quality Classes

Following the quality assessment described in Chapter 3, every rain gauge is assigned a quality class of "good", "rather good", "medium" and "rather poor" (Fig. 5.1). Missing observations in days compared to total number of days of observations are used as a further criteria of quality. Thus, stations with a lot of missing information may receive a worse quality eventhough they show a reliable reading behaviour and a good station setting. For some stations, only the quality control performed in 1998 by Sturm, 2002 and the amount of gaps can be used for the quality rating. Appendix A lists all the stations with corresponding quality, location, time period of measurements and gaps. These quality classes are used to determine, if a station should be excluded. Stations are excluded, if in addition to a bad quality assessment, the quality control of the observations shows high inconsistencies with neighbouring stations. Six stations have been excluded entirely and 107 rain gauges remain for the analysis. It has been noted in Chapter 2 that the database consists of 164 rain gauges. The remaining rain gauges do not have sufficient information of the coordinate reference system and thus, can not be used.

Seven of the used time series exhibit gaps of more than 20 %. These rain gauges are located in areas, where observations are sparse and have not been excluded. Four of these rain gauges belong to KWS Mt. Kenya. Gaps are high, because measurements have temporally been shut down and have been taken up again in 2013.

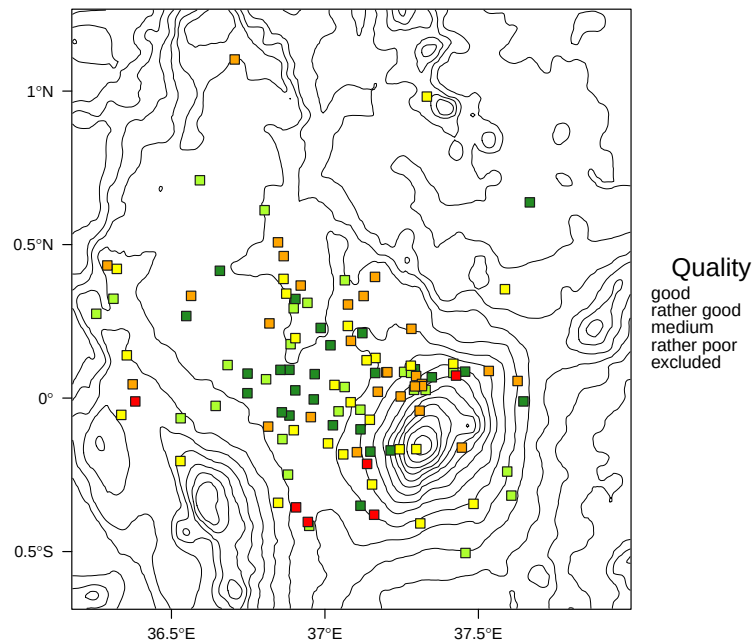


Figure 5.1: Quality classes of all rain gauges including former and newly assessed quality ratings, CETRAD Database, Nanyuki.

5.1.2 Quality Assurance

Next to the quality classes, the visits of the still operating rain gauges identified further issues of the rain gauge network. Summarized, the manual rain gauge network is affected by the following changes:

- **Changes of instrument type:** The formerly widespread used Casella type of rain gauges is slowly being replaced by a different plastic versions of rain gauges and correspondingly often locations changed slightly. Private farmers do replace their rain gauges, however, governmental institutions such as KFS do not continue measurements, when the old rain gauges break.
- **Relocations:** During several field trips, data about relocations has been collected, however no proper storage of this updated information exists. The database contains information of location from the set up of the database, and some remarks about changes, but it is rather incomplete.
- **Practice of data recording:** Two ways of daily data recording exists. Either the record is noted at the date of reading or at the date of event. According to the interviews, for some stations practices seem to have changed since the last field visit, however running correlations (over 30 days) calculated with neighbouring stations indicate that it also might be badly communicated. For Satima Farm, following the reported recording practice of the visits in 2011 would lead to overall lower running correlations, than using the recording practice noted in the database (Fig. 5.2, a). However, for a short period at the end of 2010, running correlations would indicate another practice of record. In order to reduce further adjustment of the data,

it is advisable to check this carefully during the visits. The formerly sent out record sheets that clearly indicate to write down on the date of reading are not provided anymore by KMD and therefore every stations adopts its individual practice of recording. Further, according to the conducted interviews, the time of reading spans from 6 a.m. to 10 a.m. local time. Rainfall is mainly found in afternoons or at night. Thus, different reading times are not of greater concern.

The first and second point are less problematic, if this metadata is stored properly. Such information about relocations and change in instrument types are only insufficiently stored in the database and not fully up to date. The third point does not pose a problem for the calculation of rainfall maps at a seasonal to annual scale. However, at higher resolved time scales a check for shifts in one day for the time series would be required. This issue is known and a function to shift a time series is implemented in the database. However, a careful check of the recording practice during the station visits is recommended.

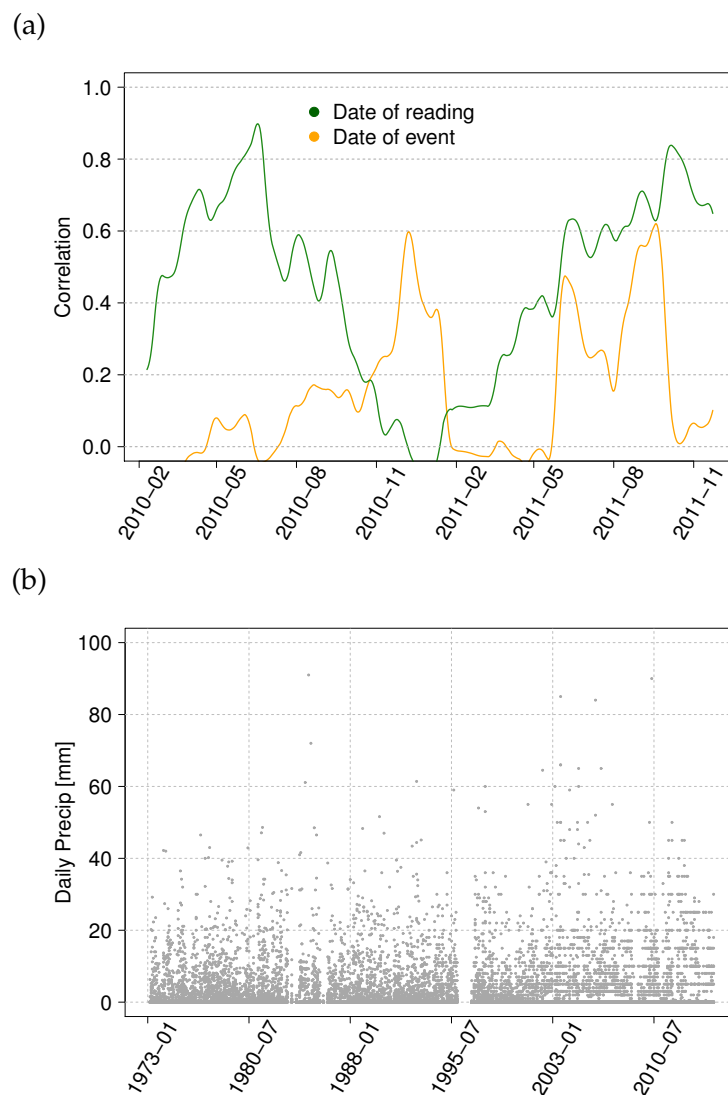


Figure 5.2: (a) Smoothed running correlations calculated for daily records of Satima Farm and Nicholson Farm for 2010 - 2012; as date of reading and as date of event. (b) Rounding of daily data and missing low values starts around 2003, Shamata rain gauge.

5.1.3 Quality Control

The comparison of monthly total precipitation between neighbouring stations identifies very high or low ratios of monthly totals with neighbouring stations that might be suspicious. Comparison of the number of wet days is conducted as well, however, this number is tampered for some time series by missing weekend readings. For very short time series, comparison of ratios is less distinctive. All the undertaken changes in the dataset resulting from comparison of monthly data with neighbouring stations and of daily data are summarized in Tab. 5.1. High amount of changes, such as for South Marmanent KFS or El Karama are due to setting whole years to missing. For South Marmanent KFS this is done, because instead of missing values zero values have been entered to the database for several years. As for most of the data dating back to around 1990 the manual recording sheet is still available. Therefore some suspicious values of the quality control can be attributed to typing errors, such as shifts in commas or high values at the end of month, where instead of the last day, the monthly total is noted. Rows of constant non zero values result either from typing mistakes or are noted, when measurements have not been taken for several days.

Further, the rain data is checked in order to identify how many stations are subject to weekly cycles and double values and whether excessive rounding of data occurs. Double values or rows of constant non zero values and weekly cycles occur, when measurement are not taken every day. Either records are noted on Monday or are distributed equally over two (or more) days. This is further described in Schmocker et al., 2015 for one part of the same dataset used here. Weekly cycles and double values do not pose a problem for the analysis of seasonal and annual data and therefore, this has not been considered further. However, it should be addressed for analysis on a daily basis. Based on a visual check, rounding of daily records is found in 25 % of the stations for at least some part of the measurement period. For the rain gauge at Shamata Gate, rounding even seem to be done up to 5 mm. The records show an increased amount of values at 5, 10, 15, 20, 25 mm (Fig. 5.2, b). This is a rather severe rounding of values and appears together with the missing of low values starting around the year 2003. Roundings may occur because of insufficient instruction of the observers or aged measuring cylinders.

Table 5.1: Changes of daily data after quality control. Value indicates number of adjusted/set to missing days.

Station	Days	Station	Days	Station	Days
Borana	1	Loldaiga Farm	61	Ol Donyo Farm	31
Chuka KFS	1	Loldaiga Northern	30	Ol Pejeta Farm	122
Colcheccio	30	Loldaiga Western	30	Rumuruti (NRM)	31
El Karama	425	Loldoto Farm	31	Rumuruti MOW	1
Embori Farm	1	Lolmarik Farm	31	Satima Farm	2
Enasoit Farm	30	Maralal DC	1	Segera Plant.	30
Gathiuru KFS	2	Mariene CRS	92	Shamata	62
Hombe KFS	5	Mogwoni Ranch	1	Sirimon Gate	61
Irangi KFS	1	Mpala Farm	33	Solio Ranch	122
Isiolo DAO	1	Nanyuki KFS	31	South Marm. KFS	730
Jacobson Farm	1	Nanyuki KAF	1	Teleki (Mt. Kenya)	68
Kabaru KFS	64	Naro Moru Met	2	Teleswani (NRM)	31
Kamwaki Farm	1	Naro Moru Moorl.	1	Tharua Farm	2
Kinamba MOW	60	Ndaragwa KFS	2	Timau Marania	1
Lamuria Met	1	Nicholson Farm	1	William Holden	212
Logilado (NRM)	62	Ol Bolosat KFS	7		

5.2 Model and Scale Selection

Model selection has been performed for the three periods MAM, OND and annual precipitation individually. However, the final model for all three periods is the same and shown in equation 5.1 with β_1 to β_{10} referring to estimated coefficients.

$$v = \beta_1 + \beta_2 \cdot x + \beta_3 \cdot y + \beta_4 \cdot slp_{45} + \beta_5 \cdot d \cdot \sin(a) + \beta_6 \cdot d \cdot \cos(a) + \beta_7 \cdot d \cdot \sin(2a) + \beta_8 \cdot d \cdot \cos(2a) + \beta_9 \cdot slp_{45} \cdot \sin(asp_{45}) + \beta_{10} \cdot slp_{45} \cdot \cos(asp_{45}) \quad (5.1)$$

with:

v = Estimation

x = Longitude

y = Latitude

slp_{45} = Slope at a 45 window size

asp_{45} = Aspect at a 45 window size

d = Distance to Mt. Kenya

a = Angle around Mt. Kenya

The full model 4.6 is compared to different models regarding predictive R^2 , AIC, AICc and absolute MAE (Tab. 5.3). The main criteria is AICc. M1 is the defined full model. M3 is the final model. Slightly changed forms of them have been tested, but not all of them are listed in Tab. 5.3. Superscripts denote a second order polynomial of latitude or longitude. Subscripts denote the window size of a topographic variable. The values are means of all individually estimated years. Thus, they are seen as guidance towards the best model. As we look for the best model representative for all years, it seems legitimate to look at mean values. AICc penalizes for small sample sizes. Yet, this might be too much penalization, as the herein model is used for fitting, rather than inference. Thus, less degrees of freedom can be tolerated. AIC is shown as a comparison. MAE in [mm] is written, in order to give an idea about the size of the cross-validation error.

For all periods, predictive R^2 is smaller than adjusted R^2 , indicating some over fitting (Tab. 5.3, M3). The smallest ratio between predictive R^2 and adjusted R^2 is found for annual models. The models for the OND season show the highest predictive R^2 . Absolute MAE between the periods can not be compared, as total precipitation amounts for the three periods differ. In the next section, the correlations between the different predictors and total precipitation is described, followed by the model selection for the three periods.

5.2.1 Correlation of Total Precipitation with Predictors

Slope values show comparatively high correlations with MAM total precipitation for all years of the rainy season (Tab. 5.2). Correlation values increase with increasing window size. However, correlations differ for the individual years and cover a broad span of values (Fig. 5.3, a). The significance of correlation is not shown, because the degrees of freedom change for every year. The window size 57 was calculated at a later time and shows overall higher correlations with the MAM precipitation totals. Altitude at a 9 gridcell window size is negatively correlated with MAM total

precipitation for around half of the years and its correlation with precipitation is in general low (Fig. 5.3, d). A high negative correlation of -0.512 is found for latitude, as precipitation decreases towards north. For MAM, the correlation with longitude is mostly positive, indicating decreasing precipitation towards the west.

For OND, correlation of slope with precipitation is high and increasing with window size, as it is for MAM (Fig. 5.3, b). Similarly, the correlation with altitude is low, but increasing with smaller window size (Fig. 5.3, e). Mean values for the correlation with altitude is positive, but almost zero. For latitude, the same expected negative correlation exists, however lower than for MAM and annual precipitation (Tab. 5.2). In comparison, OND is more correlated to longitude than is MAM or annual.

For annual precipitation, altitude becomes more positively correlated with decreasing window size, but the values still remain rather low. The mean correlation for slope again increases with window size. Latitude for annual precipitation shows the highest negative correlation for the three time periods MAM, OND and annual. Longitude shows a lower correlation compared to OND, as for annual precipitation the continental rains in the western part of the study area become relevant.

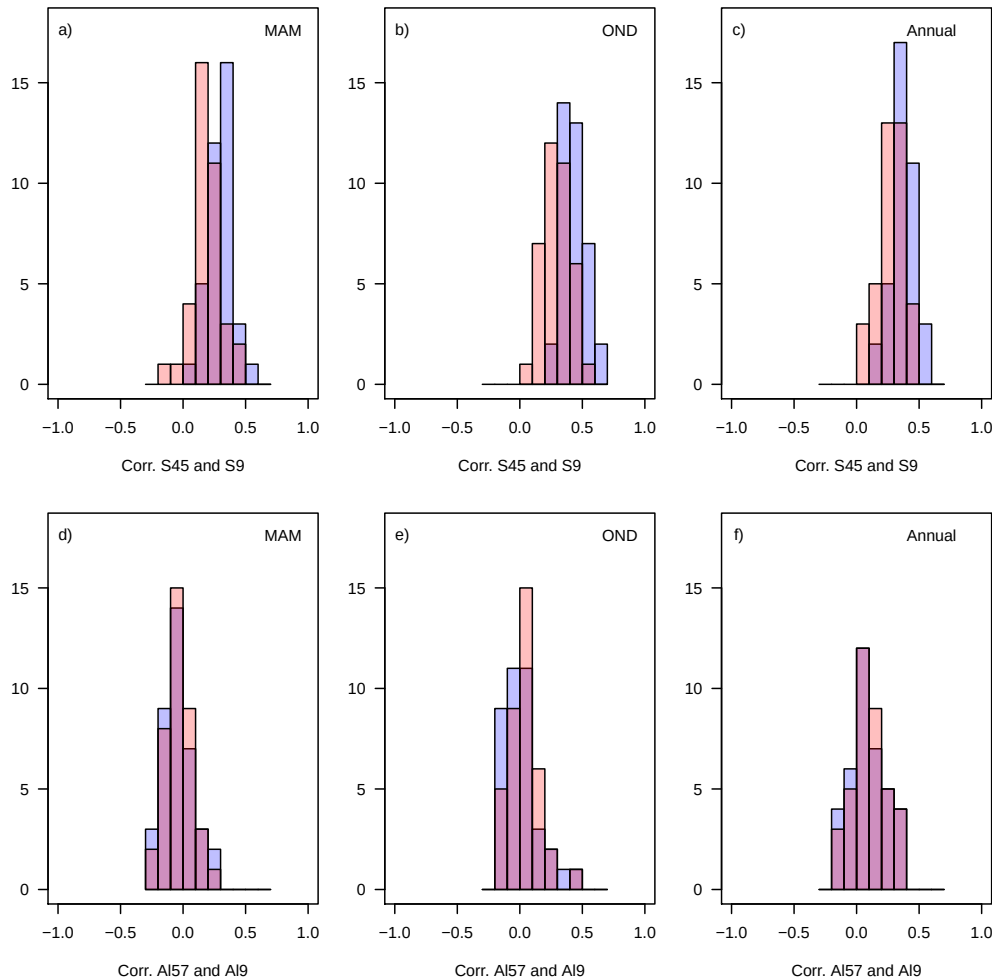


Figure 5.3: Correlation between predictors and total precipitation for MAM, OND and annual precipitation. The upper row shows histograms for correlations with slope at S45 (blue) and at S9 (red) for MAM (a), OND (b) and annual precipitation (c). The lower row shows correlation with altitude at A157 (blue) and A19 (red) for MAM (a), OND (b) and annual precipitation (c).

Table 5.2: Mean correlation for total precipitation and the explanatory variables at different windowsizes for the period of 1977 - 2014. The numbers indicates the window size used for the calculation. Al is the abbreviation for altitude, S for slope, Lat for Latitude and Long for Longitude.

Time	Al ₉	Al ₃₃	Al ₅₇	S ₉	S ₃₃	S ₄₅	S ₅₇	Lat	Long
MAM	-0.032	-0.033	-0.040	0.190	0.270	0.296	0.316	-0.512	0.306
OND	0.032	0.025	0.013	0.291	0.388	0.430	0.458	-0.472	0.488
Annual	0.104	0.101	0.091	0.274	0.358	0.375	0.384	-0.539	0.222

5.2.2 MAM Model Selection

For the MAM season, based on AICc no clear preference between the models can be made. AICc is similar for M1, M3 and M5 (730, 729 and 729). However, the full model M1 has a predictive R^2 considerably smaller than adjusted R^2 . The inclusion of two harmonic oscillations (2MK) in the model, leads to an improvement of the predictive skills. Predictive R^2 of M1 is 0.379 and for M2 it is 0.345. The best predictive skills are found for M5 (predictive R^2 of 0.52). This model has the lowest AICc and the lowest MAE. It is composed of a predictor of slope, latitude and longitude and two harmonic oscillations of the Mt. Kenya variable at different frequencies. The Mt. Kenya variable reaches zero asymptotically at a certain distance away from the mountain (Fig. 4.1). Choosing M5 would lead to a large part of the area only being predicted from slope, latitude and longitude. This might be an oversimplification and hence, M3 incorporating the slope aspect interaction term is preferred with the same AICc value of 729. M6 with only one harmonic (MK) at low frequency has a considerably higher AICc and is therefore not considered. Adding a polynomial of latitude and longitude (M7) leads to a lower predictive R^2 (0.390), higher AICc (733) and high MAE (82.1). Altitude has been added at a 9 grid window size to the model M3. This shows similarly good predictive skills with a predictive R^2 of 0.501 (model not shown in table). Because the coefficients are mostly negative and hence unphysical altitude is not kept and M3 is chosen as a final model.

5.2.3 OND Model Selection

For the OND season, the full model overfits considerably less than for MAM. Comparing the full model M1 to M2, which does not include a Mt. Kenya variable, predictive R^2 decreases by around 0.1. Thus, the 2MK variable is important. Following the AICc, M3 shows the best fit with the lowest AICc of 697. However, the more reduced model M5 and the full model M1 exhibit rather similar values. M6 with only one frequency (MK) shows less predictive skills (predictive R^2 of 0.542) than M3 and a substantially higher AICc value. This legitimises the use of a second harmonic of higher frequency in order to model a second precipitation maximum for the OND season. M7 containing a second order polynomial of latitude and longitude has a lower predictive R^2 of 0.619. AIC is equal and AICc is higher compared to M3 without the second order polynomial. Hence, M3 is chosen. Note however, that for OND, the AIC value of the full model (M1) is considerably lower than the AIC for all other models.

Table 5.3: Model Selection for the different periods MAM, OND and annual and a choice of 7 different models. The highlighted model is the final model for every period. Abbreviations of predictors are: Al = Altitude, S=Slope, SAs = Slope · Aspect, LL = Latitude and Longitude, MK = Mt.Kenya Variable using one harmonic oscillation, 2MK = Mt. Kenya Variable using two harmonic oscillations. MAE is given in [mm].

Period	Model No.	Predictors	adj.R ²	pred.R ²	AIC	AICc	MAE
MAM	M1	Al ₉ S ₄₅ (SAs)L ² L ² 2MK	0.724	0.379	721	730	84.0
MAM	M2	Al ₉ S ₄₅ (SAs)L ² L ²	0.561	0.345	746	751	92.1
MAM	M3	S ₄₅ (SAs)LL2MK	0.699	0.476	724	729	77.8
MAM	M4	S ₄₅ L ² L ² 2MK	0.686	0.464	726	732	80.1
MAM	M5	S ₄₅ LL2MK	0.680	0.520	726	729	76.2
MAM	M6	S ₄₅ (SAs)LLMK	0.590	0.443	741	745	84.6
MAM	M7	S ₄₅ (SAs)L ² L ² 2MK	0.701	0.399	725	733	82.1
OND	M1	Al ₉ S ₄₅ (SAs)L ² L ² 2MK	0.799	0.592	689	699	72.8
OND	M2	Al ₉ S ₄₅ (SAs)L ² L ²	0.703	0.499	712	716	82.6
OND	M3	S ₄₅ (SAs)LL2MK	0.787	0.656	691	697	68.8
OND	M4	S ₄₅ L ² L ² 2MK	0.777	0.635	694	699	70.8
OND	M5	S ₄₅ LL2MK	0.760	0.652	696	700	69.8
OND	M6	S ₄₅ (SAs)LLMK	0.678	0.542	716	720	81.2
OND	M7	S ₄₅ (SAs)L ² L ² 2MK	0.793	0.619	691	700	72
Annual	M1	Al ₉ S ₄₅ (SAs)L ² L ² 2MK	0.745	0.158	715	727	188
Annual	M2	Al ₉ S ₄₅ (SAs)L ² L ²	0.641	0.359	733	738	188
Annual	M3	S ₄₅ (SAs)LL2MK	0.731	0.454	716	722	164
Annual	M4	S ₄₅ L ² L ² 2MK	0.717	0.449	719	726	170
Annual	M5	S ₄₅ LL2MK	0.708	0.518	719	724	159
Annual	M6	S ₄₅ (SAs)LLMK	0.630	0.465	733	737	178
Annual	M7	S ₄₅ (SAs)L ² L ² 2MK	0.730	0.316	718	728	178

5.2.4 Annual Model Selection

For the annual models, a different behaviour can be observed. The full model with two Mt. Kenya variables exhibits a very low predictive R² of 0.158 compared to a reduced model without the 2MK terms (0.359). AIC and AICc denote M1 to be the relatively better model compared to M2. The highest predictive R² (0.518) and lowest MAE is found for M5. But the same argument applies as for MAM. This model is very reduced at far distances from Mt. Kenya. With the idea in mind, to estimate precipitation at further distant and bad sampled mountain chains, such as the Matthew Ranges, this model might be less effective, however it would need to be tested. In terms of AICc, M3, M4 and M5 do not differ essentially. M6 with only one Mt. Kenya variable has in fact a higher R² than a model comprising also the second term of the MK variable. M7 with latitude and longitude as a second order polynomial has lower predictive R² (0.316) than M3. As has been seen for MAM, adding altitude at 9 grid window size (model not shown) reduces predictive R² (0.44). For an annual model, using M3 shows the lowest AICc and is thus chosen as the final annual model.

5.3 Prediction Maps

In the following section, the MAM and OND estimations are shown separately from the annual estimates. Seasonal estimates and annual estimates have a different colorscale, because of their difference in amounts. For both, seasons and annual, some distinct years are selected in order to discuss features of the model and of the precipitation patterns. The final estimation of the remaining years can be found in Appendix B. Every year is composed of one map derived from MLR estimation, the final map and the kriged residuals. The estimated coefficients are not discussed in this thesis since the general goal is prediction. For all described years, adjusted R^2 , predictive R^2 and MAE can be found in Tab. 5.5.

5.3.1 Seasonal Precipitation

In order to have a look at the effects of model selection and of the station network, the two years 1977 and 1989 are shown and described. Compared to 1977, more than 20 additional rain gauges are present in 1989. Also, the distribution of the rain gauges differs substantially. Further, the two rainy seasons for the years 2005 and 2006 are shown. They are interesting, because of their difference in total precipitation amount. The year 2005 was a rather dry year, followed by heavy flooding in OND 2006 (Hastenrath et al., 2011). Because every year is estimated separately, it is possible to depict both these highly variable years.

The correlation of the explanatory variables and precipitation varies considerably from year to year (Tab. 5.4). For example, the correlations for altitude and total precipitation turn their sign from one year to the other. Interestingly, the dry year 2005 shows very low but positive values.

Table 5.4: Pearson-Correlation of OND total precipitation and explanatory variables at different windowsizes for the years 1977, 1989, 2005 and 2006. The numbers indicated the size of the window used for the calculation. Al is the abbreviation for altitude, S for slope. Italic values are significant at a 0.01 level.

Time	Al ₉	Al ₃₃	Al ₅₇	S ₉	S ₃₃	S ₄₅	S ₅₇	Lat	Long
1977	0.004	-0.012	-0.029	0.222	0.286	0.310	0.332	-0.428	0.292
1989	-0.071	-0.085	-0.106	0.343	0.397	<i>0.421</i>	<i>0.438</i>	-0.452	<i>0.544</i>
2005	0.041	0.029	0.010	0.301	<i>0.428</i>	<i>0.461</i>	<i>0.473</i>	-0.596	<i>0.444</i>
2006	-0.030	-0.070	-0.116	0.278	<i>0.381</i>	<i>0.434</i>	<i>0.469</i>	-0.358	<i>0.601</i>

Rainy Seasons 1977

In the year 1977, 42 stations are present for the estimation of rainfall maps (Fig. ??). They are distributed rather equally on the Laikipia Plateau. On the eastern slope of Mt. Kenya, an area which in general receives high precipitation amounts, no stations are operational. Three stations operate on the southern and the south-eastern slopes. The model for both seasons depicts a maximum of precipitation occurrence towards the southeast of the mountain, differently than when more stations at north-eastern slopes are present (Fig. 5.4). A sector of reduced precipitation is found for both seasons on the northern slopes of the mountain. The models for both seasons estimate precipitation on the slopes onto the Laikipia Plateau and in the Matthew Ranges, indicating that incorporating slope or the aspect-slope term is adding precipitation. The altitude of maximum precipitation at Mt. Kenya is found where the distance function has its defined maximum. The MLR is not able to control this with the estimation of the coefficient. Only the slope-aspect term can add further precipitation at steep slopes, depending on its coefficient. Hence, more precipitation is estimated than has been measured by any rain gauge.

Table 5.5: Adjusted R^2 , predictive R^2 and MAE for the years 1977, 1989, 2005 and 2006. For annual estimations only the years 2005 and 2006 are shown.

	MAM 1977	OND 1977	MAM 1989	OND 1989	MAM 2005	OND 2005
adj. R^2	0.826	0.737	0.687	0.860	0.610	0.863
pred. R^2	0.709	0.639	0.486	0.808	0.427	0.794
MAE [mm]	78	47	68	83	88	50
	MAM 2006	OND 2006	Annual 2005	Annual 2006		
adj. R^2	0.792	0.827	0.603	0.756		
pred. R^2	0.671	0.736	0.321	0.605		
MAE [mm]	88	110	167	231		

Precipitation of up to ± 150 mm is added with the residual precipitation (Fig. 5.5). The high differences at the south-eastern edge of the study area show that the regression model for MAM 1977 has difficulties to reproduce the strong gradient between the stations located in the southern most edge (Fig. 5.5, a). OK for MAM produces smooth transitions between residuals and unsampled locations. For OND 1977, the semivariogram for OK is exponential (Fig. 5.12, b). This leads to steep differences in semivariance at short distances of a rain gauge. The overall picture of residual interpolation is therefore slightly dotted.

For both seasons, the predictive R^2 shows rather high values of 0.709 for MAM and 0.639 for OND C. The MAE are very different, but rather because of distinct amounts in total precipitation for the two seasons than because of different predictive skills.

Rainy Seasons 1989

In the year 1989, on almost every exposition of the mountain a station is present. MAM is relatively dry, compared to a seasonal average. The highest amounts of precipitation occur on south-eastern slopes. Stations in a north-western transect away from the peak of the mountain show slightly increased amounts of precipitation. This is the secondary maximum of the MK variable. A dry area on the northern top of the mountain can be observed, as a consequence of the minimum of the MK variable.

In OND, north-eastern stations receive more precipitation. The maximum of precipitation is shifted around the mountain towards the eastern slopes. However, the station with maximum precipitation in MAM is missing for this OND estimation. This might exert some influence on the position of maximum precipitation. The secondary maximum for OND, is not at a north-western, but at a western slope. The two north-western stations mentioned in MAM, still show high precipitation values (filled squares in Fig. 5.6, b). In the final model, this precipitation is added through the kriged residuals. The kriged residuals for OND 1989 are rather punctual. This is evident on the final map. The residuals add some local variability, however, only very local. When spatial relation is short distant, as can be seen in the semivariogram of OND (Fig. 5.12, d) it might be questioned, whether useful information is gained. The range of the semivariogram shows, at what distance, variability between observations does not further increase. In this case, the range is reached at 10 km. Range values for different models (Spherical, Exponential) can not be compared directly.

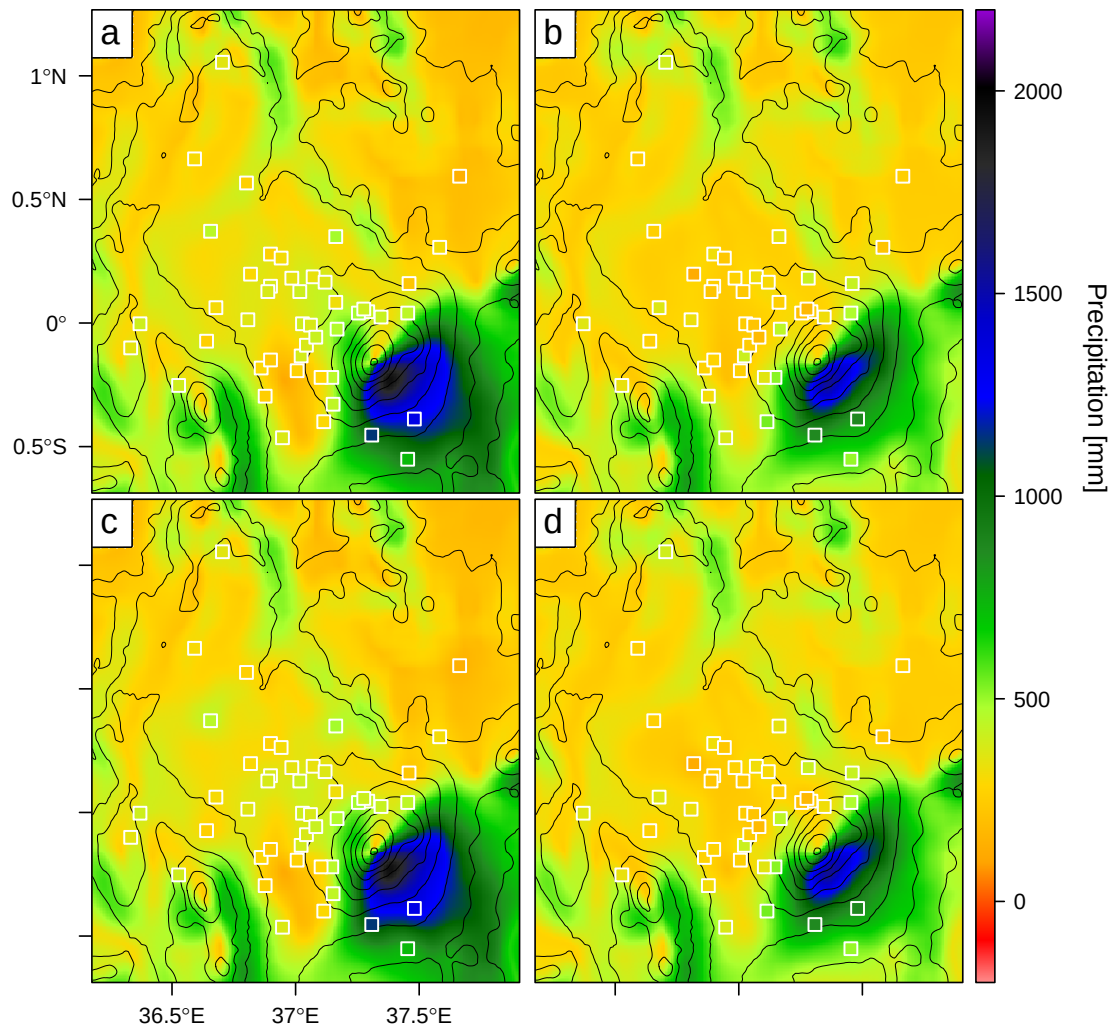


Figure 5.4: Seasonal precipitation 1977: (a) MLR MAM, (b) MLR OND, (c) final estimation MAM, (d) final estimation OND. The filled squares represent the measured value at the station.

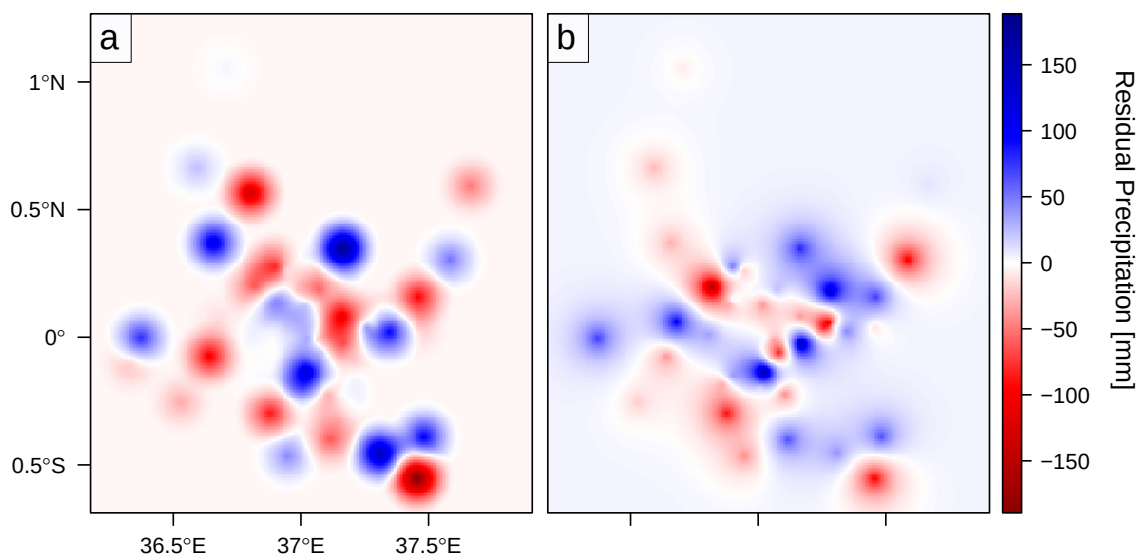


Figure 5.5: Seasonal residual kriging 1977: (a) MAM , (b) OND .

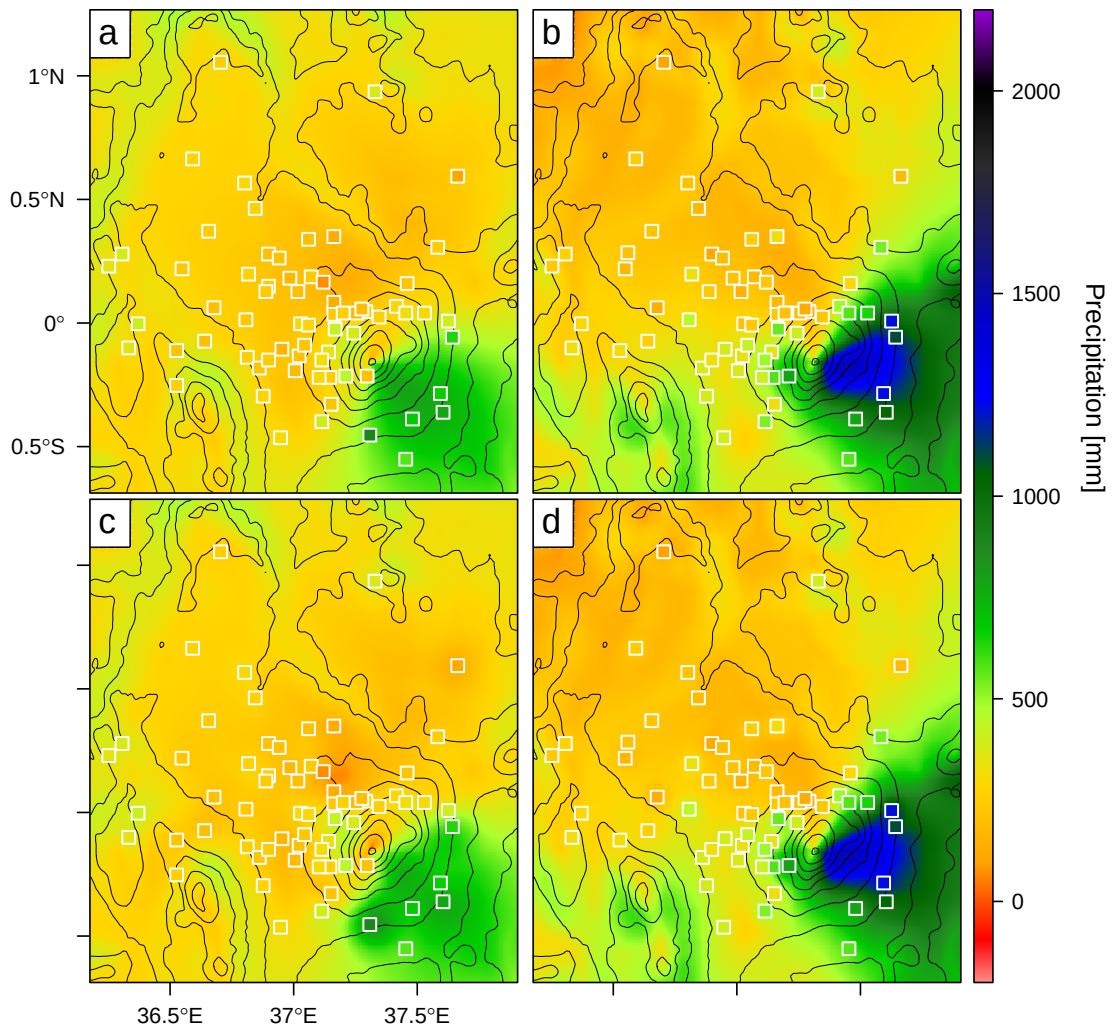


Figure 5.6: Seasonal Precipitation 1989: (a) MLR MAM, (b) MLR OND, (c) final estimation MAM, (d) final estimation OND. The filled squares represent the measured value at the station.

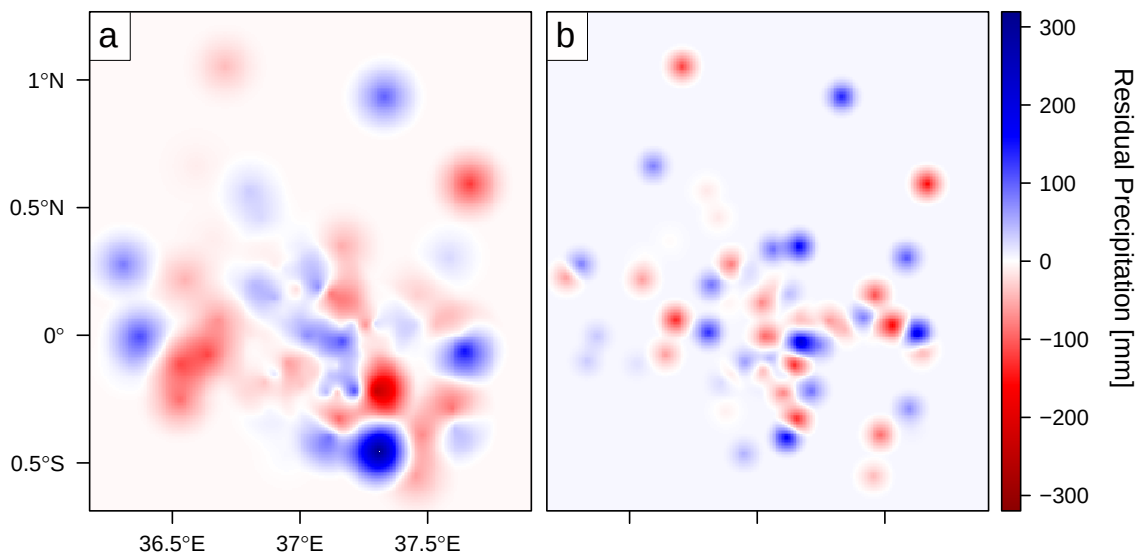


Figure 5.7: Seasonal residual kriging 1989: (a) MAM, (b) OND.

Rainy Seasons 2005

The year 2005 is comparatively dry, especially for OND and has been a drought year. For MAM, a picture similar to previous years can be observed. High precipitation amount are found at southern and south-eastern slopes with difficulties representing the gradient between the southernmost stations and the ones further up the mountain (Fig.5.8, a). The second maximum of precipitation is present, mainly for one station close to Nanuyuki showing increased precipitation values. Along the Matthews and onto slopes of the Laikipia Plateau precipitation is estimated higher than in its surroundings, indicating again a strong slope or slope-aspect term.

For OND 2005, precipitation is much below average. As can be seen from the MLR estimation, a large area in the north shows negative precipitation values (Fig.5.8, b). Adding estimated residuals, the area around rain gauges gets slightly increased values, as the rain gauges have a positive residual. On mountain slopes other than Mt. Kenya, the aspect dependence of precipitation estimation can be seen (Aberdare Range south-west of Mt. Kenya). Northern slopes are drier than south-western or south-eastern slopes. The area of negative precipitation on the Laikipia slopes may result from this same effect as well as the slightly increased values at the Matthew Ranges on south facing slopes. Also the distribution of precipitation around Mt. Kenya results not merely from the MK variable, but from a combination of the two predictors.

Rainy Seasons 2006

In comparison to 2005, 2006 was a wet year for OND (Fig. 5.10, b). The different maximum of precipitation on southern and eastern slopes of Mt. Kenya for MAM and OND are clearly visible. The cluster of stations at north-eastern slopes show higher precipitation values in OND than in MAM, as it is the case throughout all of the described years. Precipitation on the northern sector is again reduced, more so for MAM than OND.

For OND, on a north-west section of the Laikipia Plateau several stations show reduced precipitation. Precipitation in the north-eastern part might be overestimated in OND, as indicated by the highly negative residual value of the north-easternmost station (Fig. 5.11, b). It should also be noted that the latitude and longitude predictor in general show a gradient from south-east to north-west, leading to lower precipitation values in the north-west. This is the case, because they are only used as a first order polynomial.

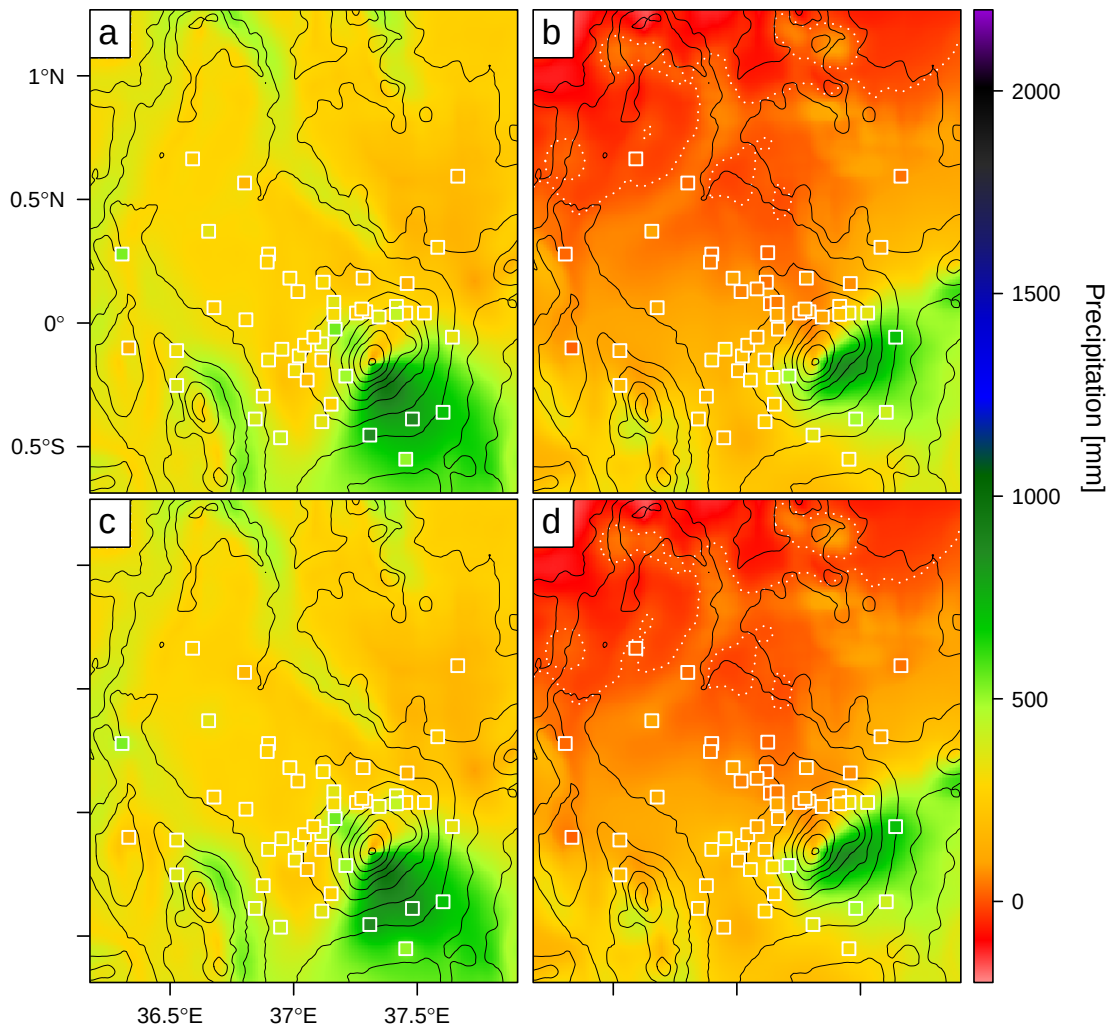


Figure 5.8: Seasonal precipitation 2005: (a) MLR MAM, (b) MLR OND, (c) final estimation MAM, (d) final estimation OND. The filled squares represent the measured value at the station. The dashed white line represents the area where negative precipitation is estimated.

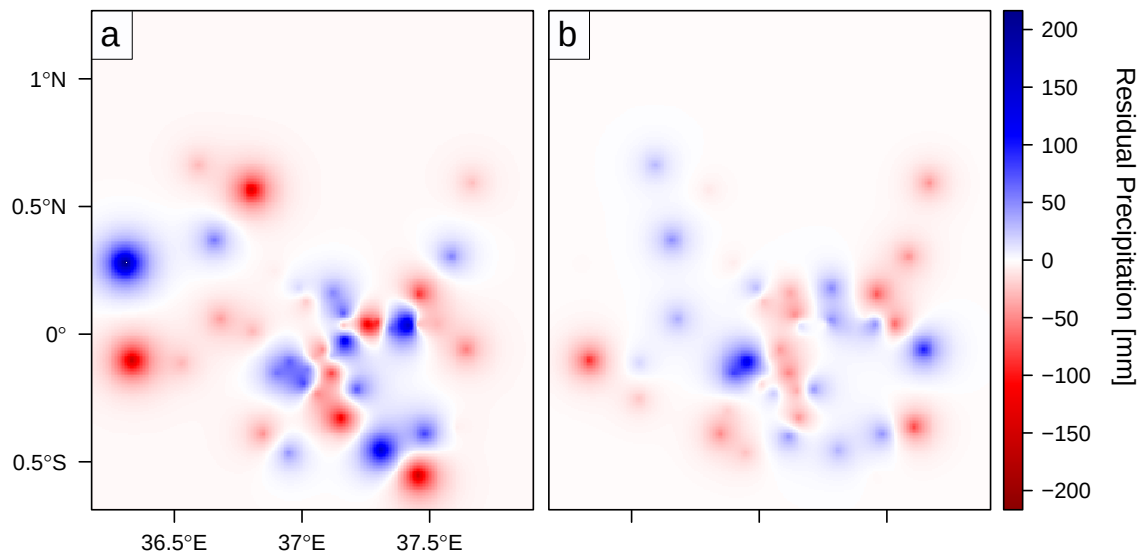


Figure 5.9: Seasonal residual kriging 2005: (a) MAM, (b) OND.

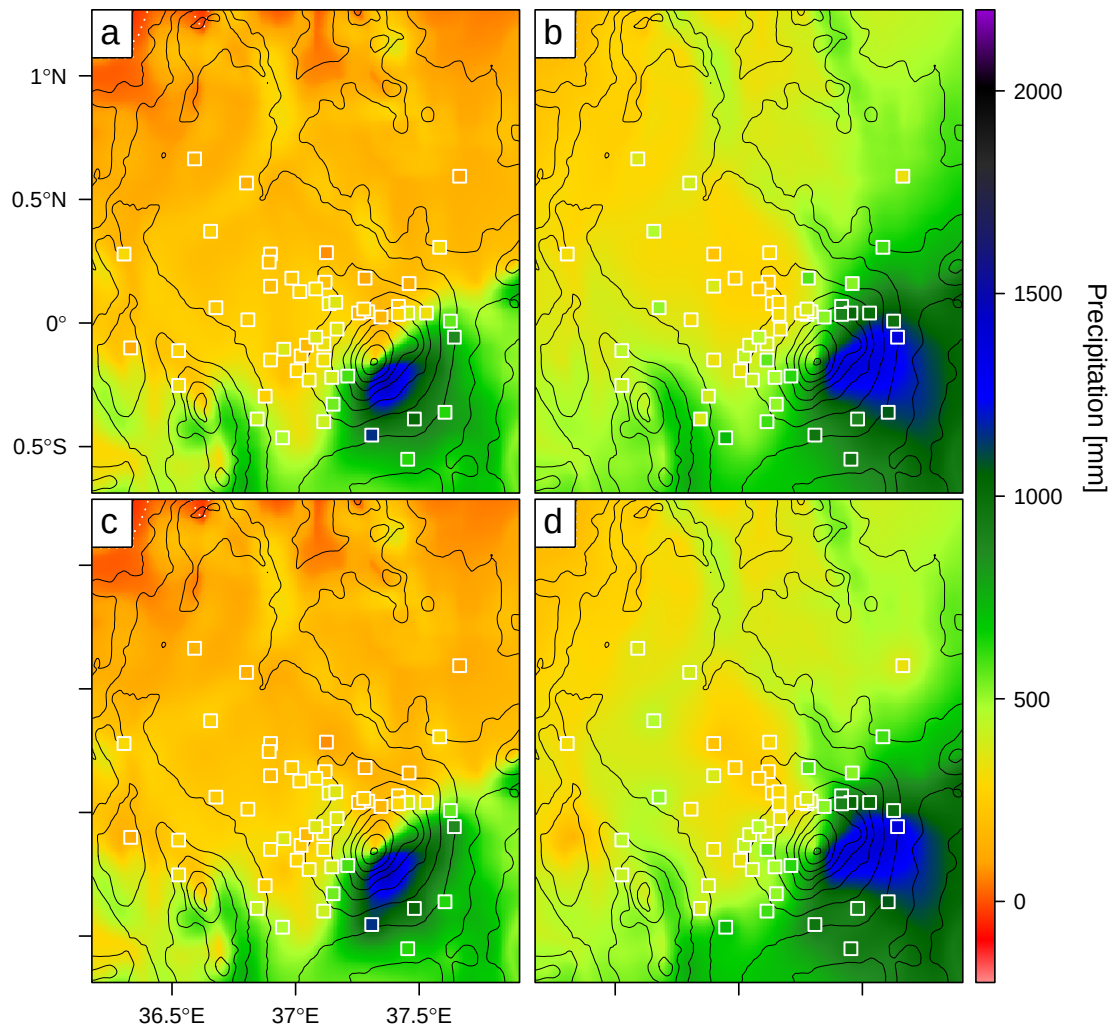


Figure 5.10: Seasonal precipitation 2006: (a) MLR MAM, (b) MLR OND, (c) final estimation MAM, (d) final estimation OND. The filled squares represent the measured value at the station.

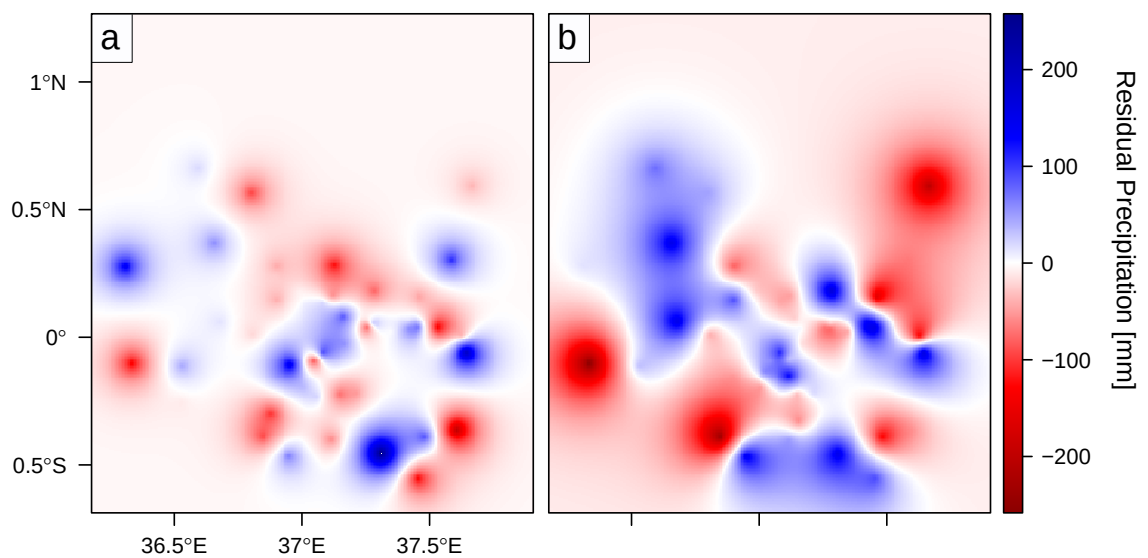


Figure 5.11: Seasonal residual kriging 2006: (a) MAM, (b) OND.

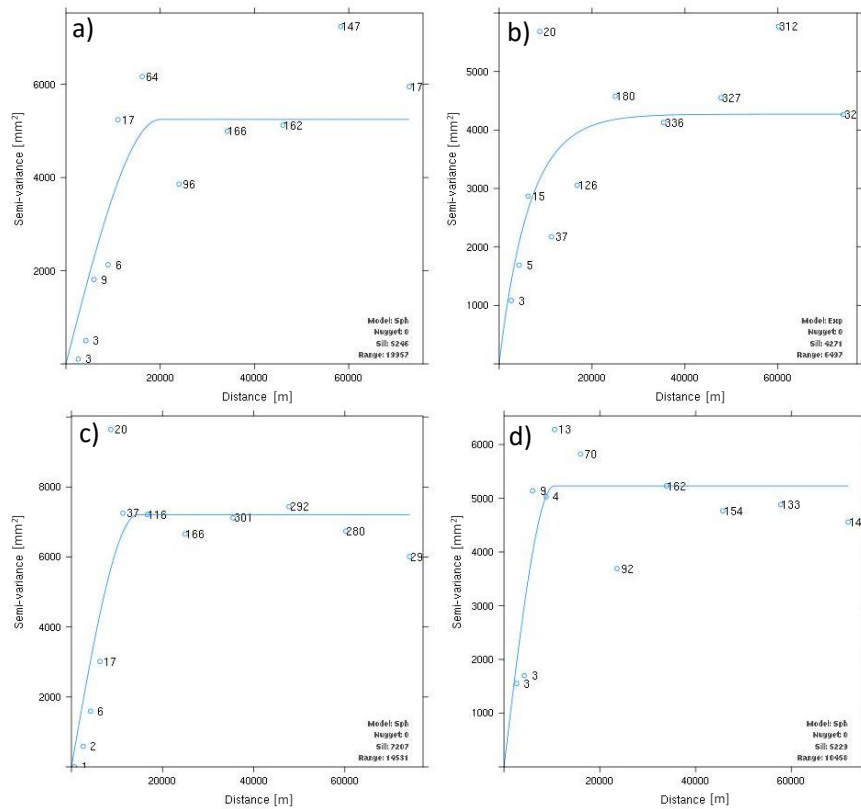


Figure 5.12: Semivariogram for seasonal residuals 1977 and 1989: (a) MAM 1977, (b) OND 1977, (c) MAM 1989, (d) OND 1989.

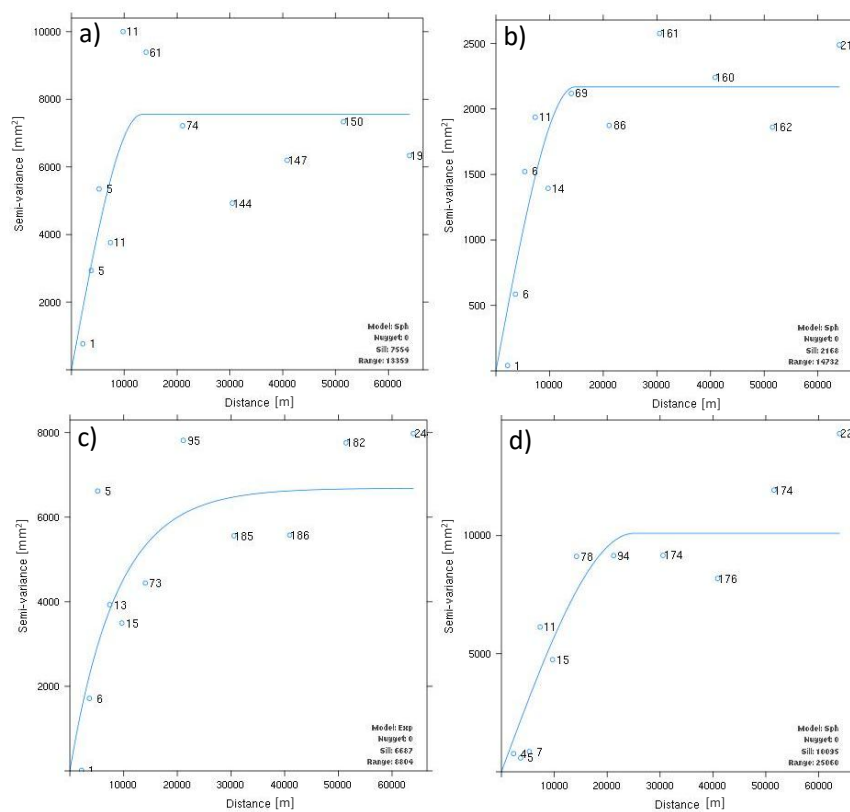


Figure 5.13: Semivariogram for seasonal residuals 2005 and 2006: (a) MAM 2005, (b) OND 2005, (c) MAM 2006, (d) OND 2006.

5.3.2 Annual Precipitation

For annual total precipitation, the years 2005 and 2006 will be described (Fig. 5.15). In the year 2005, the same dry northern sector that has been noted for seasonal precipitation is found. Not a lot of stations are present at higher altitudes, e.g. on western slopes. The north-easternmost rain gauge of the area is missing, however, precipitation estimates decrease towards north-east. Both years estimate increased precipitation for the Matthew Ranges compared to its surrounding. The observational values indicate that for higher altitude stations on south-eastern slopes values are underestimated by the models and residuals are high. Kriging of residuals in both years leads to some rather unsatisfying point pattern (Fig. 5.15, c and Fig. 5.16, a).

In 2006, the amount of rainfall on south-eastern slopes is highly increased. The maximum is found again at a south-eastern position. No secondary maximum can be distinguished visually for annual precipitation. This is in accordance to the aforementioned decrease of explained predicted variability for annual precipitation using a higher frequency oscillation for Mt. Kenya. For both years, the two northern stations are not present anymore.

In comparison to seasonal estimates, the added residual precipitation is on the order of ± 600 mm; which is reasonable, as precipitation values are much higher (Fig. 5.16). For 2005, the semivariogram takes an uncommon form, of an exponential model and a high nugget. The fit is rather poor with a sum of squares error of 54 mm^4 . For 2006, the semivariogram shows fast decreasing spatial relation between the remaining residual precipitation with a sill at 13 km. In the residuals a certain amount of measurement error can be found that can lead to high spatial variability at short distance or a high nugget value. On the Laikipia Plateau and north-western footslopes, some adjacent observational values seem rather different and possibly indicating not sufficient data quality for one of the rain gauges.

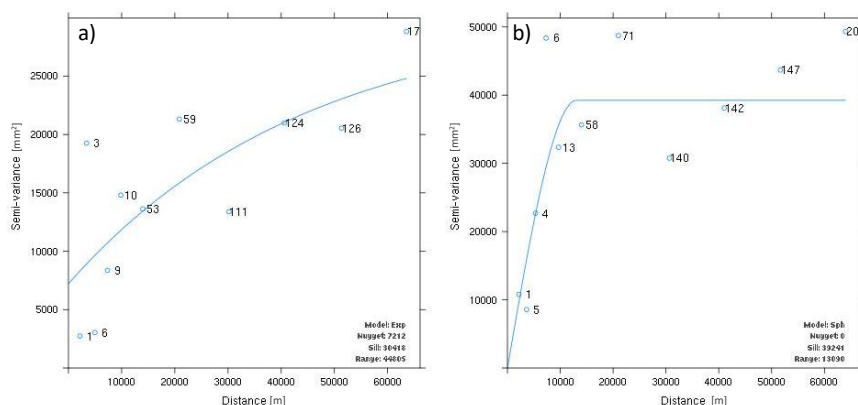


Figure 5.14: Semivariogram for annual residual precipitation 2005 and 2006: (a) 2005, (b) 2006.

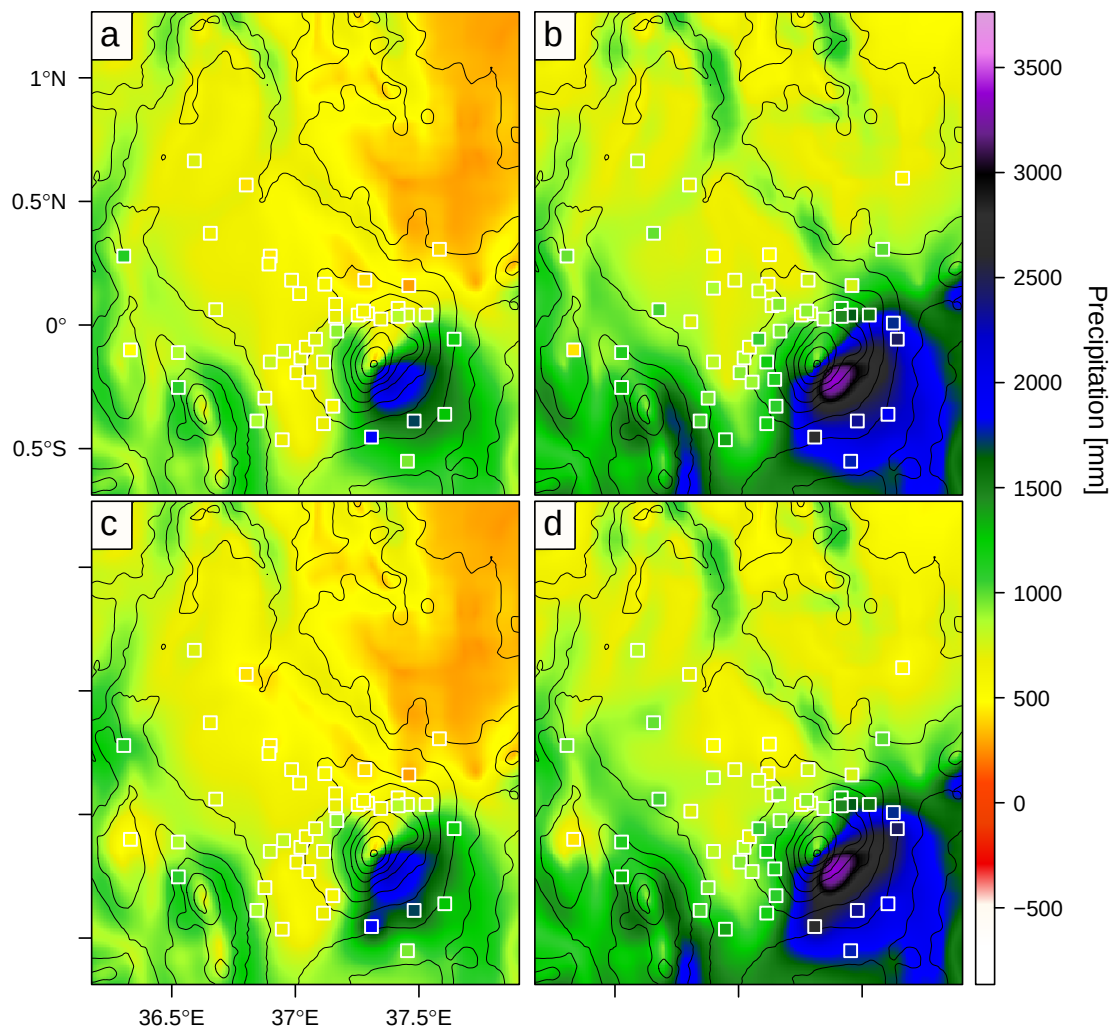


Figure 5.15: Annual precipitation 2005 and 2006: (a) MLR 2005, (b) MLR 2006, (c) final estimation 2005, (d) final estimation 2006. The filled squares represent the measured value at the station.

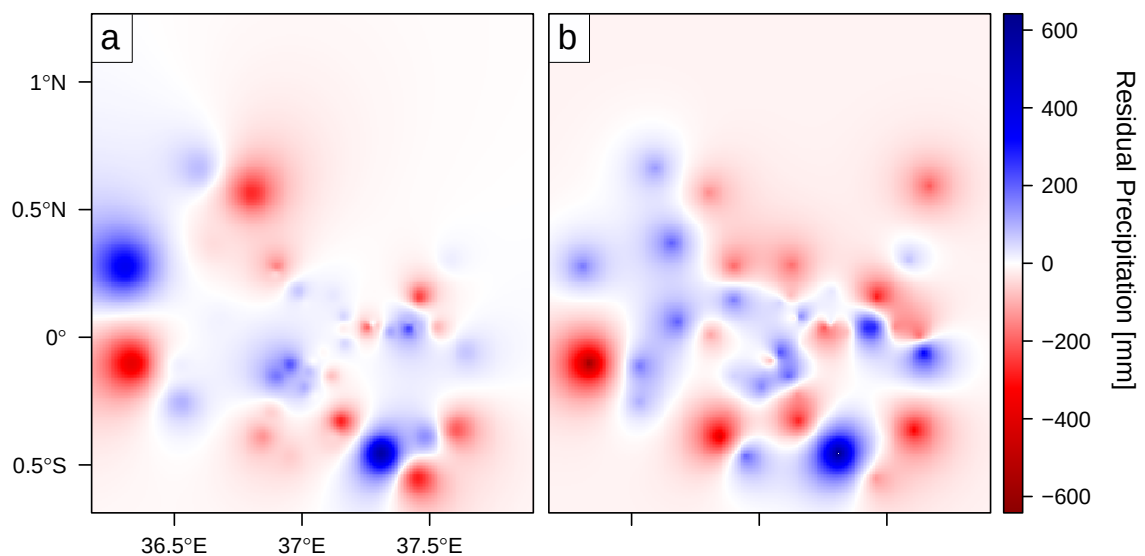


Figure 5.16: Annual residual kriging: (a) 2005, (b) 2006.

5.3.3 Cross-Validation for Individual Stations

In order to have an idea, about how well the MLR models perform in different parts of the study area, the coefficient of determination can be calculated. This is done for every station individually using cross validation predictions and observed values (Fig. 5.17). Only stations with more than 8 observations are used. It should be noted that some R^2 values comprise a total of 38 values whereas others only 9.

For all three periods, the models are able to rather consistently estimate precipitation on the Laikipia Plateau and on the northern slope of Mt. Kenya. For MAM and OND, four respectively five rain gauges show a rather low R^2 on the Laikipia Plateau. Further investigation of the time period covered by these rain gauges also in relation to neighbouring stations would be necessary. Further, two stations exhibit throughout the three periods low values for R^2 ; these are The Ark Gate and Naro Moru Moorland. Both stations have additionally large amount of gaps in their time series (46.19 % and 38.73 %).

In general, estimations for the very southern, the very northern and the high altitude stations show less shared variability with observations (R^2 below 0.4). In data sparse regions R^2 is lower. This is not surprising, as when removing one of these stations, some important information is missing.

For MAM, R^2 are lower around the south-eastern slopes of Mt. Kenya compared to values for OND and for annual estimation. Castle Forest at the southern slope of Mt. Kenya with very high precipitation shows an R^2 of 0.332 (MAM), 0.631 (OND) and 0.288 (annual). MLR estimates of this station suffer of general underestimation compared to the measurements, which has to be compensated by residual interpolation. For Castle Forest, slope values between CV-prediction and observations are 0.316 (MAM) , 0.638 (OND) and 0.368 (annual). The rain gauge in the north-east (called Archers Post) shows for the three periods R^2 of 0.211 (MAM), 0.422 (OND) and 0.291 (Annual). Slope values show that precipitation is overestimated for OND (1.064) and for annual total precipitation (1.145) and surprisingly underestimated in MAM (0.863). All values can be found in Appendix A.

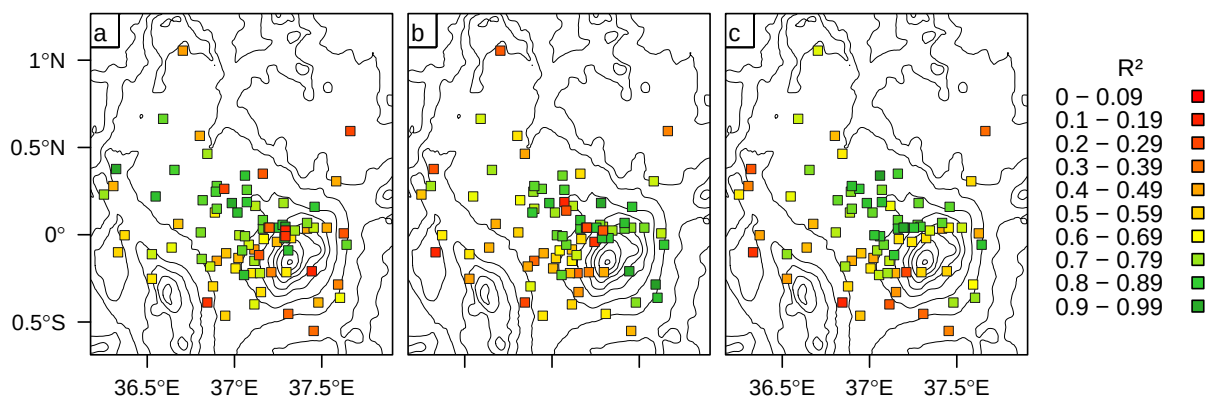


Figure 5.17: R^2 between observed value and predicted values of a Leave-One-Out Cross-Validation for every station between 1977 to 2014: (a) MAM, (b) OND, (c) annual.

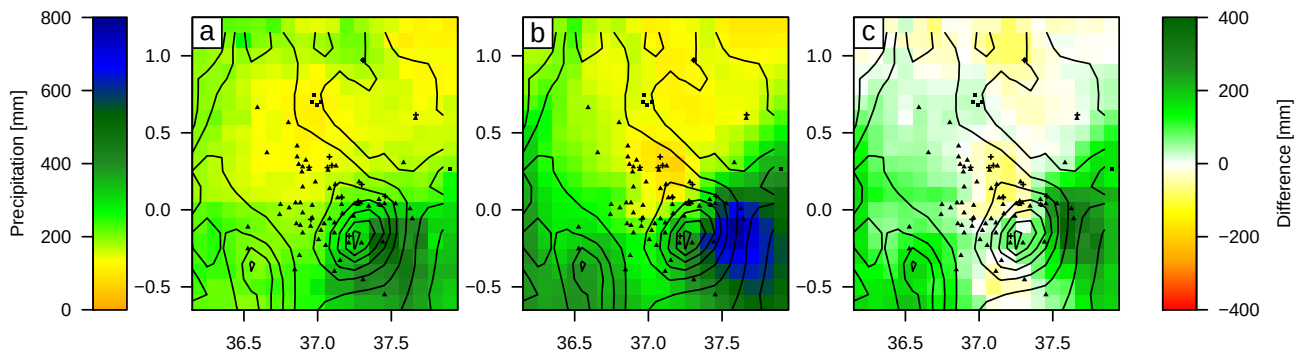


Figure 5.18: Comparison of OND 2014 estimation with the IMERG satellite product, rescaled to 0.1° . (a) IMERG OND, (b) final estimation OND, (c) difference between IMERG and final estimation in mm (right colour scale). Black triangles denote manual rain gauges, crosses denote AWS in the densely sampled area and black squares denote AWS in less dense sample area.

5.4 Comparison with Independent Data

5.4.1 Comparison with IMERG Data

In a first visual comparison, the final model estimation shows higher precipitation values for the mountain slopes to the east of Mt. Kenya and in the southwest of the Aberdare Mountains (Fig. 5.18, c). On the Laikipia plateau differences are of about +100 to -100 mm. For OND 2014, the final model is not able to estimate precipitation at the Matthew Ranges in the north of the study area, where basically no measurements are available so far. Estimation of precipitation in this region by extrapolating the topography-precipitation relation however would be interesting. The IMERG OND shows slightly increased precipitation values at the Matthew Ranges compared to the surrounding lower altitudes.

Applying a simple linear regression, the slope values of 1.522 reveals the general tendency of the model to estimate higher amounts of precipitation compared to the IMERG OND. The satellite is able to explain 0.719 of the variability of the final model (Tab. 5.6). Comparing IMERG OND to the AWS data, the satellite underestimates the AWS measurements with a slope of 0.361, however, the slope value is not significant. R^2 for AWS vs. satellite is extremely poor (0.0123). The AWS are located partly clustered and thus, the capability of explaining the variability might be reduced. For some locations, several AWS are located in the same gridcell of the satellite. Comparison is done using the value of the respective IMERG gridcell the AWS closest to.

Table 5.6: Simple Linear Regression between final model estimation, IMERG and AWS for OND 2014. Model vs. satellite value is calculated for a resampled model to a 0.1° resolution. The comparison with AWS data is based on 14 observations and the resp. extractions from model or satellite map.

	Model vs. Satellite	Model vs. AWS	AWS vs. Satellite
R^2	0.719	0.341	0.0123
Slope	1.522	0.399	0.361
p-value	2e-16	0.0283	0.706

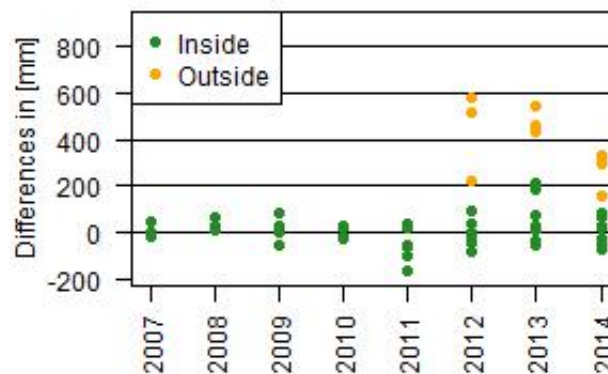


Figure 5.19: Difference in precipitation between AWS data and model estimation for OND 2007 to 2014. Green dots show stations located in data dense region ("inside"), orange dots denote data in extrapolation areas ("outside").

5.4.2 Comparison with AWS Data

RMSE values between model and AWS data for all three periods show large differences between the distinct years. For MAM, a small RMSE of 23.46 can be found in 2011. In 2010, it reaches a value of 150.98 (Tab. 5.7). Values in brackets in Tab. 5.7 denote the RMSE calculated for all stations present in that year, including the newly set up stations at poorly sampled locations of the study area, which are termed "outside" (Fig. 3.2 or 5.18). RMSE values including these AWS increase distinctively, showing rather poor extrapolation ability for the locations of the AWS and for the three years 2012, 2013 and 2014. The same can be seen when looking at differences for the AWS measurements and the estimated precipitation (5.19). As is evident from the RMSE values, the AWS located "outside" the study area show higher differences, than AWS located in the more dense sample area. The year 2013 has also some higher differences values for AWS "inside" the study area.

Table 5.7: RMSE of AWS observations with final estimations for the years 2007 to 2014. Values in brackets denote RMSE calculated for all stations.

Period	2007	2008	2009	2010	2011	2012	2013	2014
MAM	43	72.2	24.6	150.9	23.5	108.8	56.2 (353.9)	59.5(102.7)
OND	27.5	40.4	45.9	23.2	83.7	51.8(257.6)	96.5(335.4)	58.2 (194.0)
Annual	—	108.7	47.8	160.7	134.5	61.5 (228.0)	174.5 (859.2)	119.1 (346.1)

5.5 Trend Calculation

All trends are calculated for the period 1979 to 2011 relative to a reference period from 1981 to 2010. The two rainy season and the annual total precipitation show different trends. The MAM season shows in general a trend of decreasing precipitation for lower south-eastern slopes of Mt. Kenya and northern slopes (Fig. 5.20, a). On the Laikipia plateau this trend is less pronounced and turns to a positive trend for the very western and northern part of the area. Some patterns should be viewed with caution, such as the positive trend on the higher altitude south-eastern slope of Mt. Kenya.

The OND season shows an increasing trend for total precipitation for most of the area, except for a saddle between the Aberdare Mountain Range and Mt. Kenya (Fig. 5.20, b). The trends increases considerably for the north-western part. As it reaches unrealistic values, it is not further depicted. Disappearing stations, rather than an actual trend in precipitation might be responsible. Note the very low station density in areas far north and a disappearance of these stations in the last decades. For annual total precipitation, most stations show an increasing trend, except for the south-eastern area (Fig. 5.20, c). These show a very weak and rather decreasing trend. One station in the east as well exhibits a decreasing trend, restricted to a very localized area around the station. The area represents how far in general residual estimation reaches for this station.

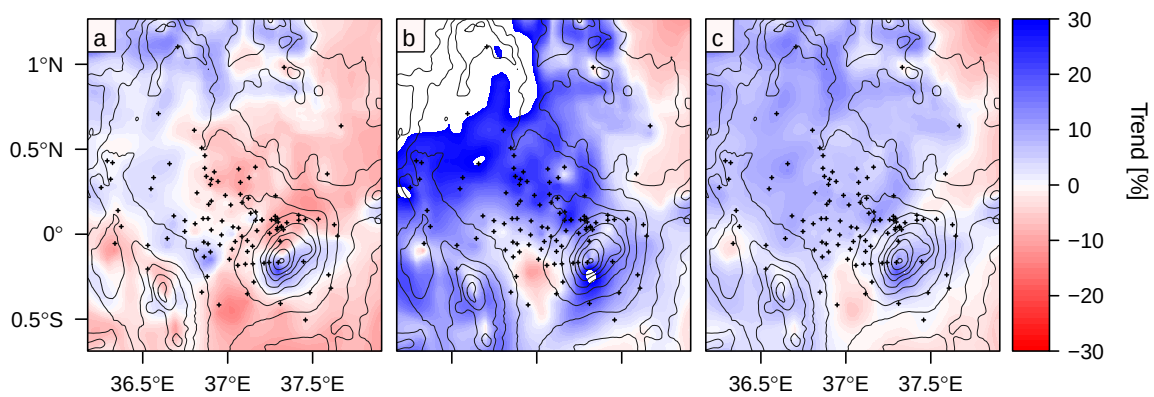


Figure 5.20: Trends for total precipitation in % per decade for the period 1979 to 2011 and relative to a reference period from 1981 to 2010. Crosses denote all stations that have been used for the estimated maps over the entire period. (a) MAM (b) OND, (c) Annual.

Chapter 6

Discussion

Following the structure of Chapter 5, the different topics will be discussed separately: the station network and its influence on precipitation estimation, model selection for multiple linear regression (MLR), the final maps and their ordinary kriging (OK) estimation, the comparison with independent data and, as a last point, a short discussion of the trend maps derived from the final estimations.

6.1 Rain Gauge Network

The rain gauge network around Mt. Kenya has proven its overall reliability in studies before (Schmocker et al., 2015; Sturm, 2002). Also, the extended version of the network is useful in the view of the estimation possible in this study and some further stations with reliable quality exist. The combination of MLR and OK reveals, when a measurement should be viewed with care; either in terms of outlying residuals or, when looking at semivariances of nearby rain gauges. As noted in section 5.3.3, some rain gauges on the Laikipia plateau exhibit a low R^2 from cross-validation for the MAM rainy season, even though they are located in a densely sampled part of the study area. They should be further examined. The two rain gauges exhibiting throughout the three periods a low cross-validation R^2 and high amounts of gaps seem very questionable. A very low R^2 between estimation and observation might indicate unreliable station measurement and an exclusion of the station would be advisable. Further, the visits of the rain gauge sites reveal that it is important to continue these visits on a regular basis and to conduct a quality assurance. Such a visit enables to assess more easily, if a station is not suitable anymore for measurements, even though measurements are still taken (e.g. Ragati KFS, Muringato KFS) and it allows to assess and inform new observers about their recording practice. As KMD does not provide any record sheets any more and does not check any more if measurements are taken regularly, it is important to stay in contact with the stations observers.

The 'interannual' changes in the station network exert an influence on the estimation as can be noted for example in the calculation of trend maps. Precipitation amounts increase or decrease with respect to the missing rain gauge and trends become increasing, respectively decreasing (Fig. 5.20). The comparison of the year 1977 with a lot less stations at the north-eastern slopes and the subsequent years shows a different maximum for the OND season, probably due to missing north-eastern rain gauges. In the last few years (2012 to 2014), the number of rain gauges declined and most of the remaining ones are located on the Laikipia Plateau. This might be a reason for the rather poor estimation with predictive R^2 of 0.436 (OND), 0.155 (MAM) and 0.262 (annual) in the year 2013 (Appendix B and C). In order to estimate precipitation distribution around Mt. Kenya a sufficiently dense network for all aspects of the mountain is necessary. Otherwise, artefacts, like precipitation peaks or minima are estimated because of missing information (Annual 1995, Appendix B). Effects of a changing rain gauge network and a reduction of the network in the recent decades have been

reported for East Africa by Funk et al. (2015) for the GPCC dataset. They mitigate such effects by separating seasonal or annual sums into a background climatology and an anomaly and perform the interpolation on anomalies. Performing interpolation on anomalies, rather than on absolute values would probably moderate leverage of rain gauges with very high or very low values on regression, such as has been noted for the very southern and very northern rain gauges (Castle KFS and Archers Post).

In the view of the station network, the question arises, whether the analysis can be extended prior to 1977. Prior to this date, no satellite data is available and an extension would provide useful, so far not available spatial information of precipitation. It is not merely a question of numbers of observations, but also of what stations are necessary for an extended analysis back in time. The first time series at a south-eastern slope of Mt. Kenya starts in 1957, a second one at a southern slope in 1960, whereas a rain gauge on a north-eastern slope dates back to 1957. In 1965, 35 rain gauges are present, thus an extension to 1965 might surely be worth a try. Also, it is not only a problem of absent operational rain gauges, but of missing observations. A month with one missing daily observation is treated as missing, and correspondingly the whole year. This problem is not restricted to the first years of observations, but arises throughout the measurement period. A bachelor thesis found that a two-stage first-order Markov Chains is able to model the occurrence of precipitation for three of the rain gauges over a period of 1979 to 2014 (Schlatter, 2016). This might be a possibility to fill missing data, at least by knowing the probability of occurrence of rain. Additionally, the high amount of short-term rain gauge measurements might provide a possibility to fill gaps using nearby rain data, rather than use this data in the spatial estimation directly.

6.2 Multiple Linear Regression

6.2.1 Model Selection

Out of MAM, OND and annual precipitation estimation, the OND season shows the highest mean predictive R^2 over all the years (M3 in Tab. 5.3). Thus, the variability of rainfall is best predicted in OND. Several reasons might be found for the different predictive abilities amongst the periods. According to the two studies mentioned in Chapter 2 Camberlin et al. (2009) and Hastenrath et al. (2011), spatial coherence for intensity and number of rain days (first study) and for total amount of precipitation (first and second study) is higher for the OND rains, than for the MAM rains. For MAM, spatial coherence between stations is found to be high at the beginning of the rainy season and then decreases substantially. The high coherence at the beginning is explained with the northward shift of the ITCZ, the resulting low-level moisture flux and south-easterly winds at the beginning of the rainy season. This sparks the simultaneous onset of rains in the entire study area and hence leads to high spatial coherence. In the course of the rainy season, the air mass is more uniform and rain is mainly of convective nature. Sufficient moisture is present and local convection governs the rainfall. Thus, spatial coherence is low for MAM (Camberlin et al., 2009). Similarly, Hastenrath et al. (2011) find boreal spring rain totals (only April-May, AM) to show less shared variance than the boreal autumn rains totals (only October-November, ON). They relate the difference in spatial coherence between stations to the development of the equatorial zonal circulation cell. This cell only builds up in boreal autumn, because of the low upper-tropospheric topography in boreal spring. In boreal autumn, it exerts influence (for example suppressing precipitation due to a high pressure gradient and high subsidence) on the entire study area, leading to high spatial coherence in OND. These two studies are both based on areas comprising all of Kenya. Still, the findings are considered to be valid for the a smaller area of this study.

In that context, the scale on which the explanatory variables are calculated has to be considered. Most variables do not comprise local information and are therefore not suitable to predict local variability. Although they are calculated at a 600 m grid cell, they comprise information of up to 30 km for example for topographic variables. For the Mt. Kenya variable, they rely on a simplification of the shape of the mountain. Thus, the estimated precipitation is representative for a larger grid cell than 600 m. If the MAM rains show more spatial variability because of local convection, this variability might be present at a lower scale, than is represented by the MLR model. Hence, the model might not be able to capture this variability and predictive R^2 is lower. For a comparison, PRISM interpolations are based on a grid of 800 m, that has been smoothed with a Gaussian filter in order to avoid terrain effects on precipitation of extents lower than 7 km (Daly et al., 2008). PRISM estimations are still rather local, as they consider observations from nearby stations only and weight these observations with regard to the topographic similarity between the cell of interest and the rain gauge.

A further reason for the higher predictive ability in OND may lie in model selection. Model selection in this study is not unequivocal and might be further improved. MAM and annual precipitation show good predictive skills also for very reduced models. The choice of model is probably most unequivocal for OND and leads to the better predictive skills for OND compared to MAM and annual totals (Tab. 5.3). For none of the three periods, AICc is able to distinguish a single most informative model, as differences of only a few units in AICc are not relevant (Burnham and Anderson, 2002). Also note that all calculated AIC show the full model to be the one with the least information lost. For annual precipitation, the second term of the Mt. Kenya variable may be less suitable. The distinction of a secondary precipitation maximum may be more difficult, because it can develop at different locations for MAM and OND, regarding prevailing air flow for both seasons. Further, the continental rains contribute to annual precipitation, which can be noted in the reduced correlation with longitude. Identifying the relevant predictors of topography is not a simple task, considering the changing rain gauge network. Also, large-scale mechanisms may cause drier and wetter years with probably dissimilar topographic relations, which would need further consideration. Possibly, the distribution of AICc or predictive R^2 would give a first idea, whether such different relations exists.

6.2.2 Predictors

The first step of this study involved calculating topographic predictors at different scales and their correlation with total MAM, OND and annual precipitation in order to gain a first understanding of how they are related. Very often, altitude is used as a major variable to explain variation in precipitation (Daly et al., 2008; Kyriakidis et al., 2001; Marquinez et al., 2003). For this study, correlation values for altitude are negative and very low, especially for MAM and OND. In the MLR model, altitude similarly shows negative coefficients for MAM and OND, but less so for annual precipitation. This is not expected, as precipitation is in general increases with decreasing temperatures and thus with altitude. Similarly, altitude shows a negative sign in two studies by Oettli and Camberlin (2005) and Hession and Moore (2011), evaluating the influence of topography on mean monthly East African rainfall. In Hession and Moore (2011), using spatial autoregressive and spatial error models in order to account for spatial dependence of residuals, altitude has a negative sign calculated on a scale of 213 km for March, April, October, November and December. Their study area is around 5 times larger than the one considered in this master thesis. Hession and Moore (2011) relate the negative sign to the correlation of altitude with latitude. For our study, latitude and altitude also correlate significantly with values of $r = -0.27$, -0.29 and -0.34 for all scales. For the boreal summer months, Hession and Moore (2011) finds positive signs of altitude coefficients on a scale of 9 km. This scale does not conform with the 9 gridcell scale used in the results presented here,

where, 9 grid cells equal 5,4 km. For this scale, highest and mostly positive correlations are found for annual totals (Fig. 5.3 and Tab. 5.2). The scale of 5-10 km is reported by several studies (Daly et al., 2008; Sharples et al., 2005) as being the most appropriate for effects of altitude on precipitation. They argue that different mechanisms, such as the advective nature of moisture-bearing airflow, viscosity of the atmosphere, delays between initial uplift and subsequent rainout and movement of air around obstacles make this scale the most appropriate. On smaller scales of the area, a positive gradient for altitude and precipitation is found. For example along the western transect of Mt. Kenya, highest precipitation is measured at the rain gauge located at 3771 m a.s.l. (Notter et al., 2007), and precipitation is increasing with altitude. On larger scales, this may be covered by different stronger effects and or collinearity. Slope at a 45 window size and altitude at 9 window size show a correlation of 0.74 significant at 0.01, which may also lead to a changing sign of altitude.

In general, slope proves to be important for predictions, as it shows high correlation values increasing with window size. MLR coefficients are significant at 0.05 for most of the models for all three periods. Slope at larger window size might be more appropriate, because the constantly rising terrain induces a large-scale lifting of air masses. The window size of 45 is still comparatively small, to what is found in Oettli and Camberlin (2005) and Hession and Moore (2011). For these studies slope is an important predictor at various scales up to 213 km. For several month, slope enters the regression equation at 3 different scales. Interestingly, slope is not used as a predictor for the boreal summer month in Hession and Moore (2011), whereas altitude is. According to Yang et al. (2014a) in boreal summer, moisture transport from the Indian Ocean is low and stable air is advected, presumably due to cold SSTs and less moisture uptake. Thus, slope could be relevant in connection with prevailing winds and the high moisture transport during the two rainy season. The effect of altitude for the dry season and for the continental rains would need to be further investigated for the area of this study.

In relation to prevailing wind directions, the Mt. Kenya predictor and the interaction term of slope and aspect should be examined. Both stand in relation to the prevailing air flow. Using a first and a second harmonic of the Mt. Kenya variable seem to be reasonable as the models incorporating them all show low AICc, compared e.g. to a model with only one harmonic (M6 in 5.3). The dry sector in the north of the mountain, already noted by Barry (2008) is found rather consistently throughout the years for OND, MAM and annual precipitation. It is predicted by the negative phase of the Mt. Kenya variable. This sector is slightly shifted counter-clockwise for OND, compared to MAM for most of the years. This is in accordance with the map in Gichuki et al. (1998) and with the isohyets drawn by Berger (1989). North-eastern locations receive more rains during the OND rainy season, compared to the MAM rainy season. Wind directions from ERA-Interim for MAM at 10-m are mostly southern, whereas for OND they exhibit an eastern component (Fig. 2.4).

The increased precipitation on western slopes of the mountain have been noted in the interpolations of Sturm (2002). The second term of the Mt. Kenya variable is able to model this increased amount and can be seen as a further proof of the existence of this secondary maximum. However, influence of available stations on the position of this maximum and differences between OND and MAM in occurrence and position of the maximum should be further investigated.

The distance function of the Mt. Kenya variable possesses some shortages. Its values are determined considering the altitudinal precipitation maximum only at a western transect. As is discussed in Chapter 2, different aspects of the mountain may show different maxima. Yet, data on distinct altitudes and aspects is mainly missing. The model is forced to predict a maximum and a minimum at the defined distance from the mountain. Also, it is predetermined how far the mountain peak influences the precipitation pattern, as it asymptotically reaches zero. Still, using a special variable accounting for flow around a mountain seems to be a good and so far never applied method to determine precipitation distribution around a mountain. It is an easily applicable way, as no numerical modelling and no inclusion of wind is needed.

The interaction term of slope and aspect is similarly able to reflect the amount of precipitation for any mountain chain depending on its exposition. As station density is higher around Mt. Kenya, it shows high collinearity with this variable. Variance inflation factors for several years exceed a value of 10. It still seems useful to predict precipitation, e.g. on the Aberdare Ranges, the Matthews or to some extent on the slopes of the Laikipia plateau as an extrapolation. Following predictive R^2 for MAM and annual precipitation, the term might be dispensable.

6.3 Final Estimation

The final estimation is not merely a product of MLR estimation, but of the combination with the ordinary kriging of residuals, the regression kriging. By adding residual precipitation to the MLR estimates, it is possible to better depict certain gradients, such as the difference between the rain gauge in the south-eastern edge and to rain gauges located at higher altitudes. A study of annual rainfall for Turkey reports RK to show good cross-validation-results, compared to MLR or just OK, because it integrates the two components, (1) a trend induced by topography and (2) spatial autocorrelation (Bostan et al., 2012). Especially, RK shows good results for extrapolations in data sparse areas. However, they find best results for Universal Kriging, a method that does not separate trend and kriging in its estimations. Kyriakidis et al. (2001) finds for the western US that integrating all available information (moisture, windfield) gives best predictions of precipitation, whereas Masson and Frei (2014) claim that a single predictor may be sufficient and considering the spatial autocorrelation may be more effective in precipitation estimation in the Alps. Possibly, the performance of an interpolation scheme depends on the quality of data, the density of the network, the form and scales of topography and just as much on the geographical region.

The question arises what the residuals might be composed of. Considering the range values of the semivariograms, some years show rather short spatial relation reaching their range at 10 km (e.g. OND 1989), resulting in a dotted residual pattern and high spatial variability of the residuals. It might be the high spatial variability of rainfall, coming from non topographic and smaller scale topographic factors triggering convection and subsequent rainout. Most probably it also comprises a measurement error. Very high differences at short distances in the semivariogram might likely also be due to errors in the measurement, rather than local variability. Unfortunately, as several point pairs are averaged to one semivariance value, such outliers may be difficult to detect but can influence the estimation of the theoretical semivariogram. A semivariance cloud, showing all point pairs and its respective semivariances, would reveal such high values. The precipitation pattern is a composition of processes acting at different scales. The MLR allows to extract the signal of higher scales, whereas in the residual pattern smaller scale processes might still be found. For this study, it would be necessary to invest further time in residual kriging and the residual component. For example, the analysis of the semivariogram in terms of the spatial relation of the residual would be an interesting subject to invest in further time.

For some years, negative precipitation values are predicted (5.8, b). A solution might be to determine the occurrence of rainfall prior to the estimation of its amount (Hwang et al., 2012). Thus, the areas with no precipitation can be set to zero, estimation is only conducted on the remaining area and estimation of negative values can be avoided to a certain extent. This procedure has not been applied here, as monthly values during rainy seasons are seldom zero. Because the residual is added to the regression prediction, in general the final estimation in this study does not become zero at the location of a rain gauge. Negative estimations can however occur at a rain gauge, if the semivariogram exhibits a nugget value above zero. A simple approach is to set negative precipitation estimates to zero, as it is for example done in Goyal and Ojha (2010). This is however problematic, because this would indicate an overall higher mean estimation of precipitation. To

avoid negative precipitation at a station, the negative rainfall could be distributed at surrounding stations in order to keep estimates unbiased. A rule of distribution would need to be developed.

6.4 Comparison with Independent Data

The final estimation for OND 2014 and the IMERG OND satellite image show a general accordance with an R^2 of 0.719 between the two datasets. In two areas, the final map shows higher estimates compared to IMERG OND: the south-eastern slopes of Mt. Kenya and the south-western slopes of the Aberdare mountains. For the OND 2014 final map, rain gauges are mainly clustered to the central part of the study area (5.18). Few rain gauges for the western areas and around the mountain are available. With this distribution of stations most of the predictions are extrapolations. This distribution of rain gauges explains the lower differences with the satellite on the Laikipia Plateau and the higher differences with peripheral areas.

The assumption that eastern stations receive more precipitation than south-eastern in OND is not confirmed by the 2014 OND IMERG satellite image. Highest precipitation for the IMERG product occurs further south. IMERG also reports error in estimations [mm] and relative weighting of integrated rain gauges. The IMERG OND errors are highest, where differences between final estimation and IMERG OND are highest. A more thorough study of the error measure would be needed. As the IMERG product is very new, it has not been empirically compared to in-situ observations for the region of Kenya. Thus, the overestimation of the final map in comparison to IMERG can not be compared with other studies.

The comparison between IMERG OND and the 14 AWS yields very poor results. The IMERG OND is not able to explain any variation in the AWS data as a value for R^2 of 0.0123 indicates. However, the AWS are located rather clustered, along the slopes onto the Laikipia plateau and the values are compared to that grid cell of IMERG OND, where the AWS is located. No interpolation or weighting for the extracted satellite values is done. This might improve explained variability.

Simple linear regression between the model and the AWS data (only for 2014), shows that the model underestimates the AWS substantially. The regression is calculated on all data, including the AWS located in poorly sampled regions. In these regions, differences between AWS and model are in general higher (Fig. 5.19, orange dots).

The RMSE values for the different years exhibit a rather broad range. The first few years only comprise stations in a rather dense sampled area and show in general a satisfying interpolation. When comparing the RMSE values of different years, it should be kept in mind that a RMSE calculated for precipitation totals is not independent of the actual amount of rain. In a wet season, a high RMSE value might actually be low, relative to the total amount of rain. As an example, the OND season 2010 shows a rather low RMSE of 23.2 and the season was rather dry. OND 2011 on contrary was a wet season and exhibits a high RMSE of 83.7. This measure is therefore not entirely suitable for a year to year comparison, less so in regions with high interannual variability. The addition of the automatic data in the data sparse area leads to an increase for MAM, OND and annual RMSE values. These additional AWS are located at the slope onto the Laikipia plateau and in the central east of the study area. For the three years 2012, 2013 and 2014, the precipitation maps and the slope or slope-aspect coefficient are not able to predict the increased amount of precipitation on the slope of the Laikipia plateau. But, it strengthens the importance of slope for the precipitation pattern. In addition, the year 2013 exhibits a rather bad predictive R^2 of 0.44 for OND, -0.16 for MAM and 0.26 for annual precipitation. This might again be attributed to changes in the manual rain gauge network.

6.5 Trends

A visual comparison of trends for total precipitation calculated by Schmocker (2013) show a good agreement in densely sampled parts of the study area for MAM, OND and annual totals (Fig. 5.20). This is expected, as both calculations rely on the same data. The MAM rains show in general a decreasing trend, the OND and annual totals both an increasing trend. The decreasing trends for the boreal spring rains have been evaluated in several studies (Funk et al., 2014; Yang et al., 2014b). A reason for the decreasing rainfall in boreal spring might be found in relation to the SST gradient in equatorial central and western Pacific (Funk et al., 2014), whereas a connection to the Pacific Decadal Oscillation has also been made (Yang et al., 2014b). A further discussion of this is, however, out of the scope of this study. For MAM, the maps allow a better visualization of the deviating increasing trend for stations in the western part of the study area. In the MAM rainy season, high altitude areas at the southern slope of Mt. Kenya show a increasing trend. The rain gauge network has been highly variable and the pattern should be compared to a trend calculated only of the station data, not the interpolation. The time series for the higher altitude stations are however quiet short. For annual precipitation, southern rain gauges show no or a rather decreasing trend. This might be related to the decreasing trend during boreal spring, which is the main rainy season in this part of the area.

Chapter 7

Conclusion

The aim of this study was to produce seasonal and annual rainfall maps for a study area around Mt. Kenya using long-term precipitation data and topography. With the incorporation of topographic and geographic predictors into a multiple linear regression model and subsequent residual kriging seasonal and annual rainfall maps have been calculated for a period from 1977 to 2014.

The comparatively dense rain gauge network of this study for a mountainous region proves its general reliability in producing reasonable rainfall maps. The analysis confirms the influence of the particular and unique topography on the precipitation pattern in the tropical study area. A specially designed predictor is incorporated into the regression model in order to account for the influence of Mt. Kenya on the precipitation pattern. This improves the predictive skill of the regression models and reassures the barrier effect of the mountain that can be presumed from station data and processes of orographic lifting. With this predictor, alternating precipitation maxima for the two main rainy seasons can be modeled on the slopes of the mountain. The Mt. Kenya predictor is able to give better confirmation that a secondary maximum exists in the lee of the mountain. However, where exactly it develops for the two rainy seasons, and what atmospheric conditions are especially favourable needs further investigation. The development of this variable has been possible by integrating local expertise of the precipitation pattern and shows the potential of a knowledge-based approach. The manifold reported influence of altitude on precipitation can not be entirely confirmed. For annual precipitation, a weak correlation between altitude and precipitation is found, but for the boreal spring and boreal autumn rains, the correlation is basically non-existent and subject to changes from year to year. The influence of slope on precipitation pattern is more important for the two rainy seasons. This finding is in accordance with other studies for East African Rainfall (Oettli and Camberlin, 2005; Hession and Moore, 2011). Adding kriged residuals to the multiple linear regression estimation allows for improved precipitation estimates, such as steep gradients on short distances.

Several limitations have to be mentioned. Firstly, the high variability of the station network in space and time exerts influence on the predictions. Differences in predictions between distinct years, especially in extrapolation areas and around the mountain, can result from a changing network alone, regardless of climatic factors. Secondly, the residual estimations are only partly successful and varies from year to year. In sparsely sampled locations, only a very local information is added to the precipitation estimation. Further, the comparison with automatic weather station data for the years 2012 to 2014 shows a limited extrapolation ability of the models. These AWS are located on steep slopes and therefore confirm the effects of slope on precipitation.

Based on these findings and limitations, several recommendations can be made for further research:

- The effects of the changing network on estimations could be avoided by a more thorough selection of stations covering a sufficient time span and area. It would be reasonable to decide upon a selection of reliable rain gauges from the present database. The rain gauges used by Schmocker et al., 2015 may be a first basis, but further rain gauges are available that provide reliable data. A thorough review, a quality selection and subsequently disclosure would bring benefits. The dense network is an advantage, as short time series may provide a basis for gap filling of longer but incomplete time series. A good dataset may improve predictions more than a complex prediction method.
- Interpolations based on anomalies, rather than total precipitation should be taken into consideration. This might mitigate persistent problems such as the underestimation on southern slopes and the effects of a changing network.
- The re-estimation of every year independently allows to model both dry and wet year. Still, it might be worth exploring the influence of topography effects for dry and wet years separately. Large-scale circulation may cause topography to act differently on the precipitation pattern.
- No satisfying way to tackle negative estimated precipitation values has been applied so far to the precipitation maps. A method to redistribute negative precipitation would need to be developed.
- For future applications of the rainfall maps, it is necessary to provide spatial error maps. The cross-validation for MLR reveals, where the models perform worse and where they perform better, however no error measure of the residual kriging and of the combined approach has been provided so far.

In the view of expected climatic changes and socio-economic vulnerability, spatial information of rainfall of the past and the present can be helpful. The region around Mt. Kenya is especially exposed to conflicts between different water users due to the uneven distribution of rainfall in space and time. In order to mitigate exacerbating conflicts between different water users, the development of strategies for water resource allocation is necessary. A good database, the maintenance of a reliable network and high resolved and accurate data in space and time are therefore essential.

Appendix A

Data Overview

Automatic Weather Station Data

Table A.1: Automatic weather stations with coordinates (UTM Zone 37 South) and period of measurements.

Database ID	Station Name	Organisation	Start	End	Longitude	Latitude	Altitude [m]
nd	NARO MORU MET STN	CETRAD	2014	2014	9981153	301231	3037
nd	ARCHERS POST	CETRAD	2012	2014	10068190	351632	842
nd	MATANYA	Syngenta	2009	2014	9993143	272285	1831
nd	KALALU	Syngenta	2009	2014	10009037	295644	2026
nd	NYAHURURU	CETRAD	2012	2014	10001867	205042	2362
nd	TIMAU MARNIA	Syngenta	2013	2014	10009709	328483	2502
nd	ETHI POLYTECHNIC	Syngenta	2013	2014	10018188	310733	2199
nd	GACHARAIGU SCHOOL	Syngenta	2012	2014	10077330	279857	1290
nd	MARIMIRA COFFEE FACTORY	Syngenta	2012	2014	10077064	273187	1343
nd	MURIARJAS VOCATIONAL CENTRE	Syngenta	2013	2014	10082394	274485	1119
nd	WATUHA SECONDARY SCHOOL	Syngenta	2012	2014	10075534	276738	1385
164	OL JOGI IL POLEI	Private	2007	2014	10038040	286297	1737
161	OL JOGI MARULA	Private	2012	2014	10030491	270996	1712
163	OL JOGI MAWE NYEUSI	Private	2007	2014	10031789	288187	2079
162	OL JOGI MAWE NYOKA	Private	2007	2014	10030323	284871	1887

Manual Rain Gauge Data

All manual rain gauges are listed in the following two tables. Coordinates are given in UTM Zone 37 South. Gaps are calculated as missing days per total length in days. R^2 is for observation vs. CV-prediction. The rain gauges used in this study are marked with an x .

Abbr.	Station	ID	Org.	Quality	Start	End	Gaps [%]	Longitude	Latitude	Altitude	Analysis	R ² MAM	Slope MAM	R ² OND	Slope OND	R ² Year	Slope Year
ARP	ARCHERS POST	1	KMD	1,2	1957	2014	2,36	351300	10065660	860	x	0,21	0,86	0,42	1,06	0,29	1,15
ARF	ARDENCAPLE FARM	2	Private	1,8	1948	2014	6,09	305700	10004460	2162	x	0,74	0,81	0,86	0,99	0,83	0,92
BOR	BORANA	4	Private	2,8	1975	2014	45,35	308700	10020060	2035	x	0,65	0,69	0,9	0,69	0,76	0,73
CAS	CASTLE FOREST STN	5	KFS	2,3	1959	2014	6,77	311700	9949860	1906	x	0,33	0,32	0,63	0,64	0,29	0,37
CHF	CHOGORIA FOREST STN	7	KFS	1,72	1989	2002	3,01	343500	9968460	1713	x	0,4	0,44	0,91	0,8	0,61	0,61
CHG	CHOGORIA GATE	16	KWS	3,33	1991	2014	51,75	326700	9976860	2967	x	0,02	0,17	0,9	1,55	0,52	0,84
CHU	CHUKA FOREST STN	8	KFS	1,33	1988	2006	5,32	344700	9960060	1474	x	0,61	0,74	0,89	0,86	0,72	0,87
COL	COLCHECCIO	9	Private	1,8	1963	2014	12,22	255300	10062660	1831	x	0,45	0,92	0,6	0,88	0,46	0,93
DOL	DOL DOL DAO	10	DAO	2,67	1957	1993	5,38	295500	10038660	1820	x	0,26	0,44	0,59	0,47	0,79	0,76
EK	EL KARAMA	95	Private	1,9	1950	2014	4,41	266100	10016460	1797	x	0,69	0,58	0,73	0,74	0,76	0,77
EMN	EMBORI (NRM)	13	NRM3	3	1992	2001	10,24	311700	9990060	3436	x	0,77	0,92	-	-	-	-
EMB	EMBORI B	64	NRM3	3	1992	2000	8,45	312300	10000260	2670	x	-	-	-	-	-	-
EMC	EMBORI C	76	NRM3	3	1992	2000	8,36	312300	9999060	2740	x	-	-	-	-	-	-
EMF	EMBORI FARM	12	Private	1,3	1967	2014	0,01	315900	10002660	2691	x	0,65	0,82	0,77	0,99	0,65	0,96
EM	EMBU MET STN	14	KMD	1,8	1976	2014	1,07	327900	9939060	1540	x	0,34	0,81	0,48	0,84	0,35	0,96
ENF	ENASOIT FARM	15	Private	2,33	1971	2002	1,33	285300	10020660	1912	x	0,74	0,98	0,92	1,04	0,85	0,93
GAT	GATHIURU FOREST STN	17	KFS	1,3	1959	2014	2,32	290100	9983460	2330	x	0,65	0,64	0,65	0,67	0,68	0,77
HOM	HOMBE FOREST STN	20	KFS	1,3	1957	2014	2,19	290100	9955860	1991	x	0,63	0,61	0,57	0,47	0,26	0,48
IRA	IRANGI FOREST STN	21	KFS	2,33	1957	2007	8,94	330900	9957060	2009	x	0,49	0,82	0,81	0,96	0,7	1,38
ISI	ISIOLO DAO	22	DAO	2,2	1957	2014	7,79	342300	10033860	1163	x	0,47	0,89	0,71	1	0,53	0,85
JAC	JACOBSON FARM	23	Private	1,7	1934	2014	0,01	282300	9990060	1905	x	0,7	0,75	0,65	0,85	0,71	0,96
JUN	JUNCTION (EWASO NAROK)	24	?	3	1965	2000	8,33	260100	10051260	1541	x	0,66	0,62	0,44	0,45	0,57	0,53
KAB	KABARU FOREST STN	25	KFS	2,3	1959	2014	8,2	294300	9963660	2237	x	0,56	0,7	0,43	0,61	0,39	0,52
KAL	KALALU (NRM)	28	NRM3	1,2	1986	2014	2,41	295500	10003860	2080	x	0,68	0,65	0,86	0,76	0,86	0,73
KAM	KAMWAKI FARM	29	Private	2	1965	2014	0,33	295500	10009260	1980	x	0,71	0,78	0,87	1,05	0,77	0,85
KAR	KARURI (NRM)	31	NRM3	1,33	1992	2004	6,76	313500	9997860	2976	x	0,57	0,71	0,87	1,19	0,54	0,83
KIN	KINAMBA MOW	27	Ministry	2	1990	2003	6,26	201900	10041660	2097	x	0,81	0,87	0,34	0,55	0,25	0,61
KIS	KISIMA FARM	32	Private	2,2	1974	2014	4,92	323700	10007460	2451	x	0,67	0,48	0,89	0,78	0,77	0,65
LAM	LAMURIA MET STN	33	KMD	1,33	1963	2004	6,37	261900	9979860	1876	x	0,64	0,74	0,45	0,72	0,7	0,9
LAR	LARIAK FOREST STN	11	KFS	1,67	1986	2006	0,03	200100	10030860	2020	x	0,47	0,47	0,74	0,66	0,34	0,62
LOG	LOGILADO (NRM)	35	NRM3	3	1992	2005	15,49	309900	9999060	2640	x	0,84	0,97	0,84	0,86	-	-
LOB	LOLDAIGA BARRIER	142	Private	2	2005	2014	7,53	292500	10008660	1889	x	-	-	-	-	-	-
LOF	LOLDAIGA FARM	34	Private	1,3	1976	2014	2,82	290700	10018260	2160	x	0,65	0,63	0,74	0,9	0,63	0,71
LON	LOLDAIGA NORTHERN	144	Private	3	2005	2014	7,58	291300	10031460	1796	x	-	-	-	-	-	-
LOW	LOLDAIGA WESTERN	143	Private	3	2005	2014	7,64	286500	10015260	1868	x	-	-	0,97	1,09	-	-
LOD	LOLDOTO FARM	36	Private	1	1951	2008	1,93	279300	10014060	1790	x	0,79	0,73	0,75	1	0,76	0,94
LOM	LOLMARIK FARM	37	Private	2,1	1965	2014	17,85	308100	10006260	2304	x	0,75	0,84	0,8	0,8	0,78	0,86
LOR	LORUKU FARM	38	Private	2,2	1963	2014	9,27	286500	9993660	2040	x	0,63	0,65	0,82	0,64	0,77	0,78
MAR	MARALAL DC	39	Ministry	2,67	1974	1997	4,88	244500	10116660	2011	x	0,42	0,57	0,37	0,85	0,65	0,8
MAI	MARIENE CRS	42	Research	1,3	1989	2007	2,77	348900	9993660	1670	x	0,7	0,58	0,87	0,86	0,84	0,8
MAT	MATANYA (NRM)	44	NRM3	3	1986	2014	9,02	272100	9988260	1840	x	0,55	0,83	0,45	0,6	0,49	0,83
MER	MERU FOREST STN	45	KFS	2,6	1957	2014	3,01	347100	10000860	1756	x	0,22	0,51	0,73	0,69	0,44	0,6
MOG	MOGWONI RANCH	47	Private	1	1977	2014	0	275700	10020060	1760	x	0,81	0,8	0,89	0,81	0,85	0,82
MPF	MPALA FARM	48	Private	1,3	1972	2014	2,52	266100	10030860	1690	x	0,7	0,67	0,85	0,8	0,86	0,82
MPC	MPALA RANCH CLIFFORDS	150	Private	3	2010	2014	16,66	268500	10035660	1663	x	-	-	-	-	-	-
MPD	MPALA RANCH DARABOI	151	Private	3	2010	2014	16,66	261900	10045860	1706	x	-	-	-	-	-	-
MPM	MPALA RANCH MUKENYA	149	Private	2,33	2010	2014	16,66	261900	10038060	1678	x	-	-	-	-	-	-
MPT	MPALA RANCH TOP SPRAY	148	Private	2,33	2010	2014	16,66	263100	10032660	1772	x	-	-	-	-	-	-
MPR	MPALA RESEARCH CENTRE	147	Research	1,5	1998	2014	4,62	265500	10027260	1698	x	0,71	0,8	0,77	0,81	0,74	0,77
MUC	MUCHEENE FOREST	125	KFS	0	1989	2014	7,79	336300	10004460	2246	x	0,43	0,76	0,81	0,94	0,67	1,01
MUG	MUGIE RANCH	50	Private	1,8	1957	2014	5,18	231900	10073460	1860	x	0,7	0,8	0,69	0,71	0,66	0,68
MUK	MUKENYA FARM	51	Private	3	1958	2004	9,38	257100	10021860	1802	x	0,7	0,71	0,67	0,7	0,76	0,82
MUKG	MUKOGODO (NRM)	52	NRM3	1,67	1989	2002	7,47	284100	10037460	1740	x	0,77	1,09	0,81	0,95	0,95	1,21
MUN	MUNYAKA (NRM)	53	NRM3	2	1991	2014	5,25	283500	9974460	2070	x	0,78	0,86	0,85	0,97	0,67	0,78
MUR	MURINGATO FOREST STN	54	KFS	3	1968	2014	7,64	270900	9950460	1720	-	-	-	-	-	-	-
MUT	MUTARA ADC FARM	55	Private	1,33	1961	2011	2,29	241500	10006860	1941	x	0,54	0,47	0,47	0,49	0,46	0,42

Abbr.	Station	ID	Org.	Quality	Start	End	Gaps [%]	Longitude	Latitude	Altitude	Analysis	R ² MAM	Slope MAM	R ² OND	Slope OND	R ² Year	Slope Year
NAF	NANYUKI FOREST STN	40	KFS	1,9	1989	2004	8,35	293100	9987060	2337	x	0,88	0,94	0,72	0,66	0,82	0,80
NAK	NANYUKI KAF	57	KAF	2,33	1971	2000	1,39	280500	9999660	1860	x	0,64	0,86	0,83	1,08	0,83	0,99
NMP	NARO MORU FG POST	59	#NV	3	1957	1995	17,92	288900	9975660	2195	x	0,59	0,62	0,49	0,50	0,71	0,58
NMF	NARO MORU FOREST STN	60	KFS	2,67	1965	2003	5,57	292500	9971460	2364	-	-	-	-	-	-	-
NMG	NARO MORU GATE STN	61	NRM3	1	1968	2011	1,87	293700	9975660	2420	x	0,64	0,67	0,40	0,49	0,44	0,63
NMM	NARO MORU MET STN	62	NRM3	1	1978	2014	1,19	300900	9976260	3050	x	0,37	0,56	0,40	0,77	0,21	0,53
NMMO	NARO MORU MOORLAND (R2)	58	NRM3	2	1990	1995	46,19	304500	9976260	3771	x	-	-	-	-	-	-
NDA	NDARAGWA FOREST STN	63	KFS	1,6	1981	2014	2,25	224700	9987660	2319	x	0,63	0,95	0,62	0,52	0,73	0,80
NGEL	NGELESHA	43	#NV	3	1993	1996	12,59	198300	10042860	2060	x	-	-	-	-	-	-
NGN	NGENIA (NRM)	65	NRM3	3	1989	2001	17,63	299100	10004460	2110	x	0,88	0,62	0,94	0,74	0,93	0,72
NGB	NGENIA B	104	NRM3	3	1989	2001	14,15	299700	10004460	2225	x	0,90	0,64	0,95	0,77	0,91	0,74
NIC	NICOLSON FARM	66	Private	1,3	1970	2014	3,71	279900	9985260	1950	x	0,54	0,68	0,57	0,78	0,42	0,72
NOM	NORTH MARMANET FOREST STN	3	KFS	2,33	1992	1998	17,83	205500	10010460	2310	x	-	-	-	-	-	-
NYE	NYERI MOW	67	Ministry	1,33	1957	2011	3,24	271500	9948660	1811	x	0,55	0,52	0,63	0,56	0,48	0,51
OA	OL ARABEL FOREST STN	68	KFS	1,33	1983	2003	4,78	194100	10025460	2100	x	0,66	0,72	0,56	0,82	0,57	0,76
OB	OL BOLOSAT FOREST STN	74	KFS	2,32	1957	2013	1,37	203100	9988860	2446	x	0,55	0,62	0,27	0,57	0,16	0,24
OD	OL DONYO FARM	69	Private	1	1955	2014	0,15	309900	10005060	2380	x	0,82	0,95	0,73	0,83	0,79	0,86
JF	OL JOGI FARM	70	Private	1,4	1967	2014	16,68	270900	10029060	1790	x	0,85	0,83	0,84	0,82	0,82	0,79
JO	OL JORO OROK FTC	71	FTC	2,72	1979	2005	4,95	208500	9993660	2506	-	-	-	-	-	-	-
OM	OL MYSOR FARM	72	Private	1	1965	2014	0,02	239100	10041060	1820	x	0,73	0,64	0,76	0,76	0,74	0,68
PF	OL PEJETA FARM	73	Private	1,4	1965	2014	8,93	255900	10001460	1877	x	0,67	0,72	0,57	0,72	0,68	0,68
PG10	OL PEJETA GOLF 10	153	Private	1	2012	2014	0,00	264300	9988860	1844	x	-	-	-	-	-	-
PH	OL PEJETA HEADQUARTERS	155	Private	1	2012	2014	0,00	264300	10005060	1772	x	-	-	-	-	-	-
PLD	OL PEJETA LOIDIEN	156	Private	1	2012	2014	0,09	261300	9990060	1821	x	-	-	-	-	-	-
PLR	OL PEJETA LOIRUGURUGU	160	Private	1	2012	2014	0,09	260700	10005060	1778	x	-	-	-	-	-	-
PRE	OL PEJETA RESEARCH GATE	158	Private	1,17	2012	2014	0,00	266100	9997860	1777	x	-	-	-	-	-	-
PRO	OL PEJETA RONGAI GATE	157	Private	1	2012	2014	0,00	273300	9994260	1829	x	-	-	-	-	-	-
PSE	OL PEJETA SERAT GATE	159	Private	1	2012	2014	0,00	273300	10003260	1794	x	-	-	-	-	-	-
PSI	OL PEJETA SIRIMA	152	Private	1	2012	2014	0,00	248700	9996660	2004	x	-	-	-	-	-	-
PG7	OL PEJETAGOLF 7	154	Private	1	2012	2014	0,00	249300	10003860	1887	x	-	-	-	-	-	-
ONT	ONTULILI FOREST STN	75	KFS	2,72	1957	2014	8,23	296100	9997260	2130	x	0,61	0,68	0,58	0,50	0,55	0,66
PYR	PYRAMID OL JOGI	46	Private	2,67	1993	2014	19,75	285300	10028460	1852	x	0,71	0,73	0,86	0,77	0,74	0,71
RAG	RAGATI FOREST STN	77	KFS	3,12	1957	1992	19,77	294900	9952860	2030	-	-	-	-	-	-	-
RUN	RUMURUTI (NRM)	79	NRM3	3	1989	1996	14,58	228300	10031460	1820	x	-	-	-	-	-	-
RUM	RUMURUTI MOW	80	Ministry	1,3	1982	2003	2,18	227100	10024260	1882	x	0,80	0,67	0,67	0,72	0,64	0,68
SAT	SATIMA FARM	82	Private	2,1	1973	2014	5,96	278100	9978660	1950	x	0,61	1,01	0,53	0,77	0,56	0,79
SEG	SEGERA PLANTATIONS	83	Private	1,33	1965	2004	2,71	264900	10014060	1715	x	0,61	0,52	0,87	0,70	0,68	0,62
SHA	SHAMATA	84	KWS	2	1973	2014	5,82	224700	9972060	2923	x	0,60	0,97	0,43	0,54	0,40	0,76
SIRJ	SIRAJI (NRM)	85	NRM3	2,67	1992	2001	3,61	309900	10002660	2490	x	0,83	0,78	0,98	0,89	-	-
SIRM	SIRIMA (NRM)	86	NRM3	2,67	1989	1998	5,12	256500	9984660	1948	x	0,67	0,76	-	-	-	-
SIRG	SIRIMON GATE	116	KWS	2,8	1983	2014	30,50	304500	9995460	2635	x	0,50	0,56	0,28	0,40	0,56	0,52
SOL	SOLIO RANCH	87	Private	1,6	1967	2014	4,88	263700	9967260	2007	x	0,55	0,81	0,67	0,92	0,52	0,81
SMF	SOUTH MARMANET FOREST STN	88	KFS	2,6	1957	2011	5,29	207300	9999660	2291	x	0,55	0,72	0,63	0,59	0,42	0,59
SE	SUGUROI ESTATE	89	Private	1,33	1962	1995	3,87	237300	9991860	2083	x	0,62	0,80	0,48	0,61	0,46	0,87
TELK	TELEKI (MT KENYA)	90	KWS	2	1978	2014	33,62	310500	9976260	4262	x	0,60	2,28	0,46	1,25	0,53	2,67
TELW	TELESWANI (NRM)	91	NRM3	1,67	1992	2005	11,20	309300	9997860	2724	x	0,76	0,94	0,88	1,02	-	-
THF	THARUA FARM	92	Private	1,9	1943	2014	19,98	266100	9983460	1845	x	0,43	0,59	0,39	0,83	0,43	0,63
TAG	THE ARK GATE	115	KWS	2	1970	2014	38,73	260100	9957060	2170	x	0,20	0,74	0,29	0,71	0,13	0,65
TF	TIMAFLO	140	Private	1,83	2006	2014	5,60	308100	10003860	2373	x	-	-	-	-	-	-
TM	TIMAU MARANIA	41	Private	1	1939	2014	0,12	327900	10004460	2536	x	0,64	0,67	0,79	0,80	0,59	0,62
TM12	TIMAU MARANIA NO. 12	146	Private	1,67	2010	2014	4,44	324900	10003260	2743	-	-	-	-	-	-	-
TM9	TIMAU MARANIA NO.9	145	Private	1,67	2000	2014	2,79	323700	10003860	2684	x	0,43	0,52	0,85	0,69	0,44	0,43
TRT	TREETOPS	49	KWS	2,72	1987	2014	26,30	266700	9955260	1911	-	-	-	-	-	-	-
TRF	TRENCH FARM	93	Private	1,33	1964	2000	2,03	284100	9999060	1903	x	0,66	0,78	0,85	0,81	0,83	0,86
WAM	WAMBA DO	94	Ministry	2,33	1989	1995	0,00	314100	10103460	1500	x	-	-	-	-	-	-
WIH	WILLIAM HOLDEN	141	Research	1,67	2001	2014	6,18	290100	9990660	2110	x	0,42	0,42	0,56	0,45	-	-
LEW	LEWA	nb	Private		1972	2011		328500	10017660	1851	x	0,74	0,84	0,88	0,90	0,77	0,89

Appendix B

Rainfall Maps

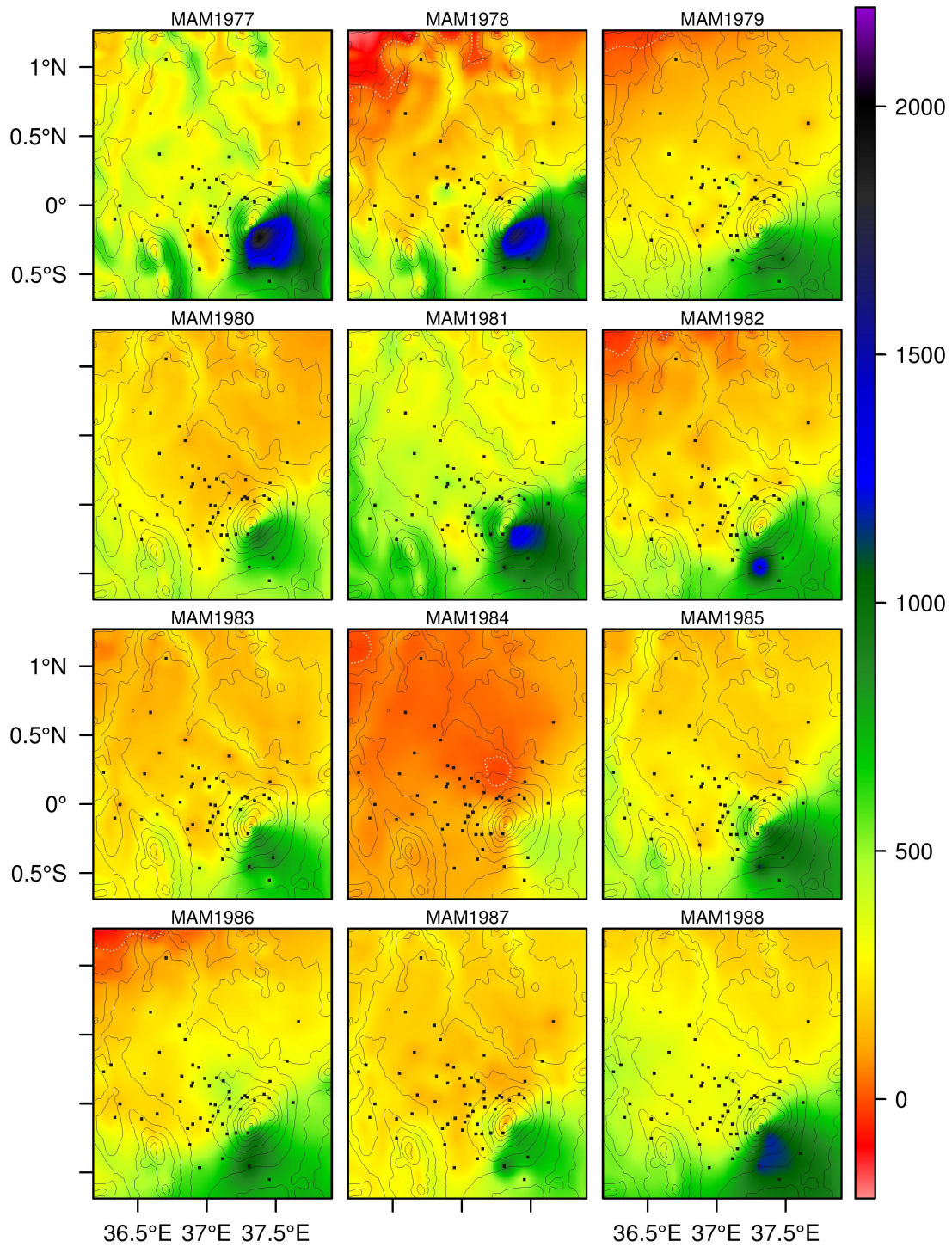


Figure B.1: Final estimations of the MAM rainy season from 1977 to 1988. Dots represent the observations used for the estimation.

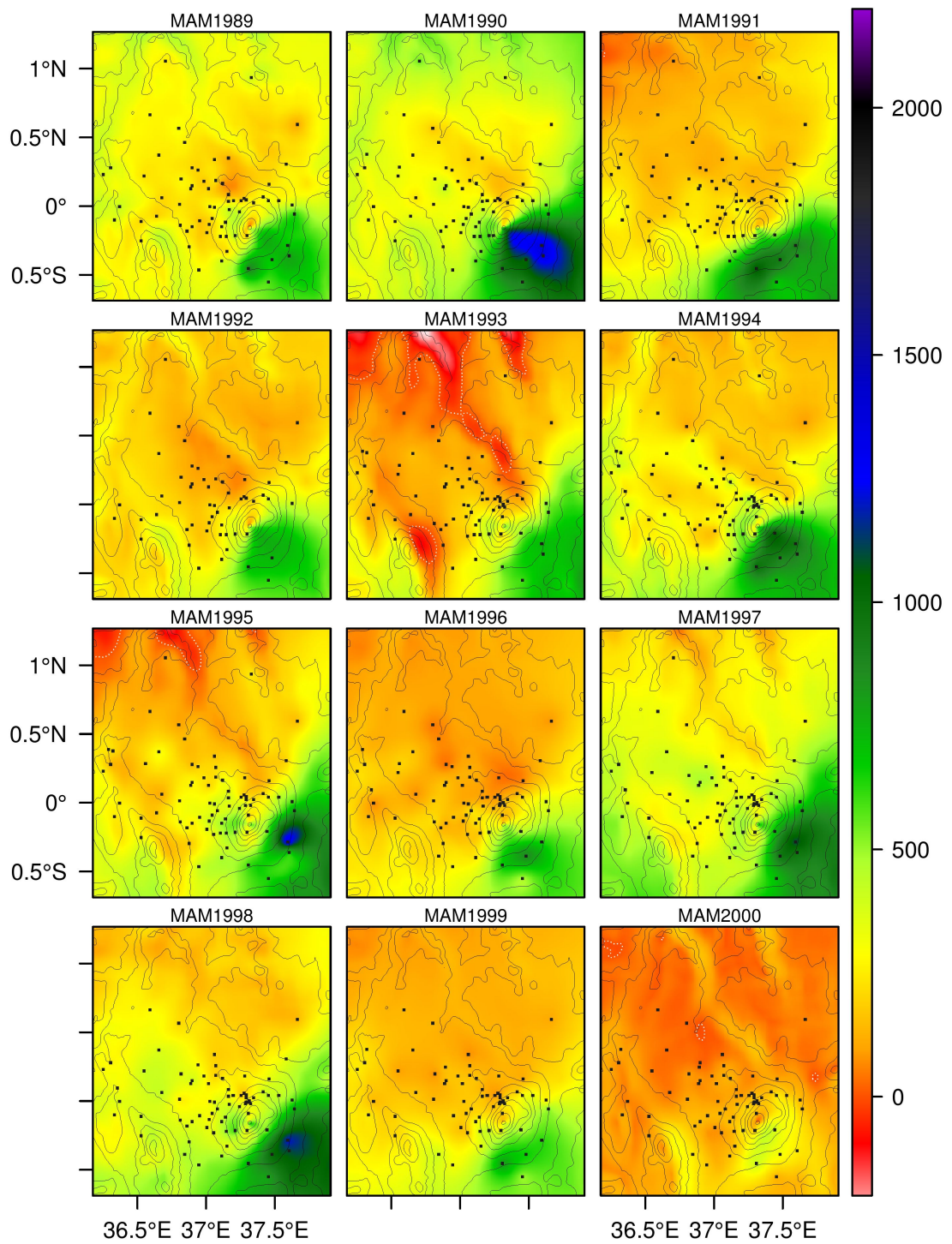


Figure B.2: Final estimations of the MAM rainy season from 1989 to 2000. Dots represent the observations used for the estimation.

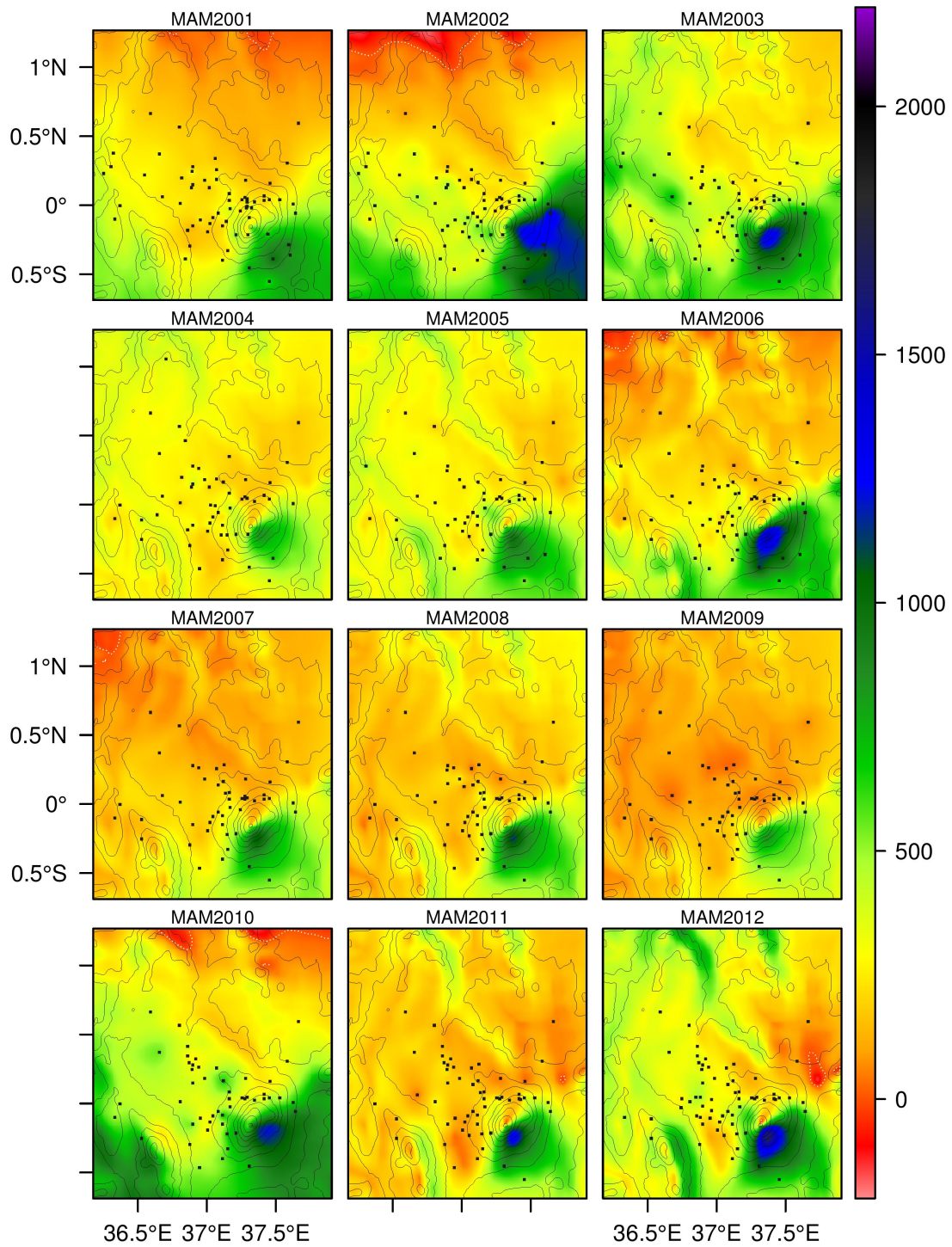


Figure B.3: Final estimations of the MAM rainy season from 2001 to 2012. Dots represent the observations used for the estimation.

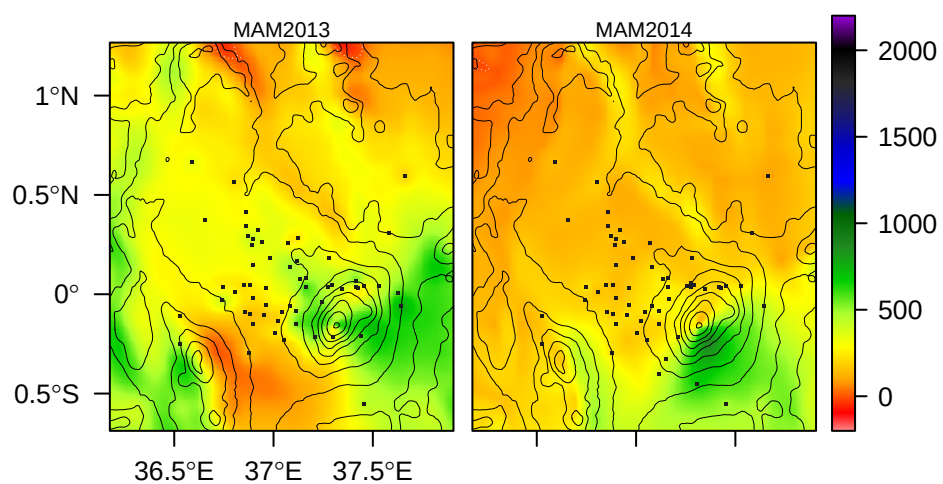


Figure B.4: Final estimations of the MAM rainy season from 2012 to 2014. Dots represent the observations used for the estimation.

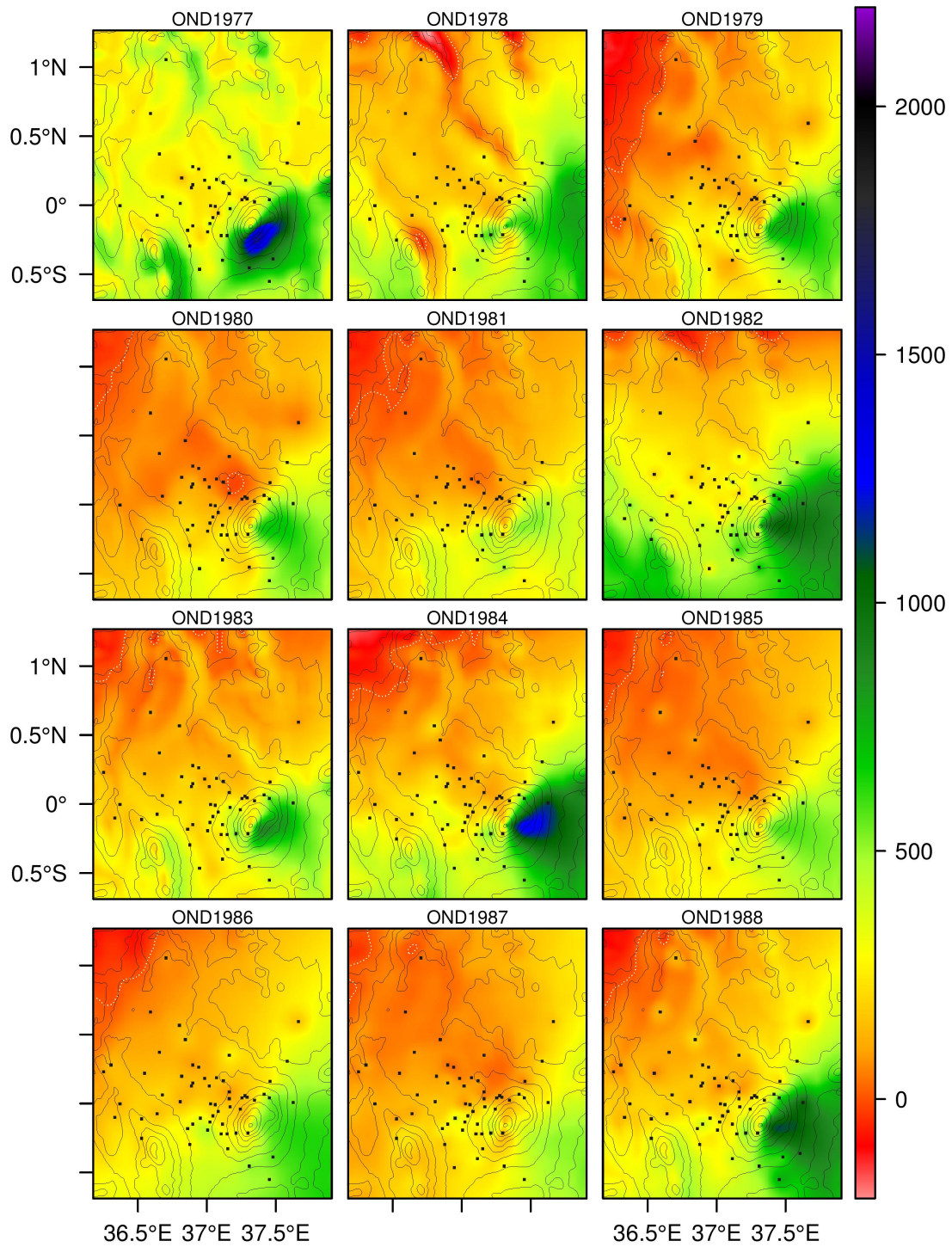


Figure B.5: Final estimations of the OND rainy season from 1977 to 1988. Dots represent the observations used for the estimation.

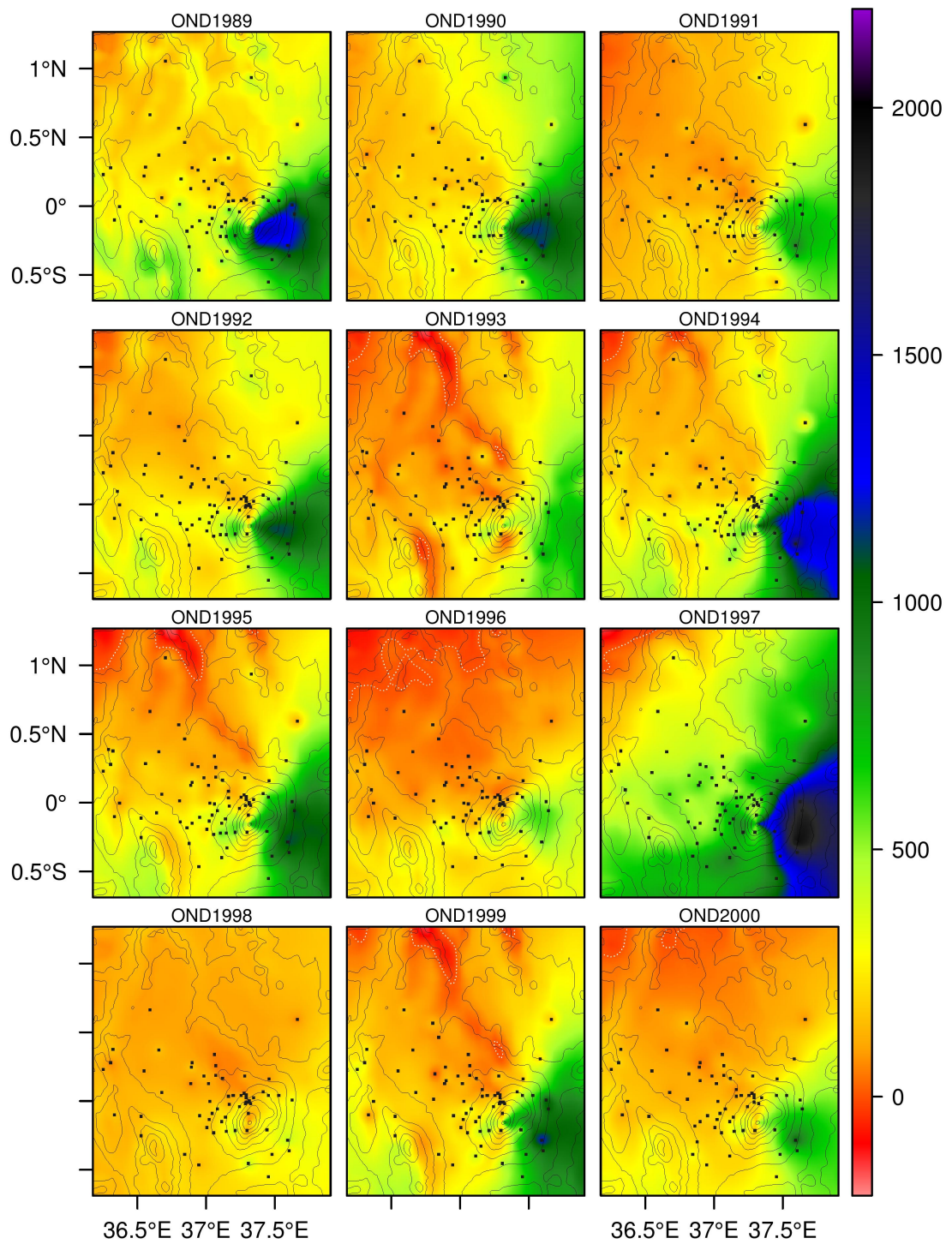


Figure B.6: Final estimations of the OND rainy season from 1989 to 2000. Dots represent the observations used for the estimation.

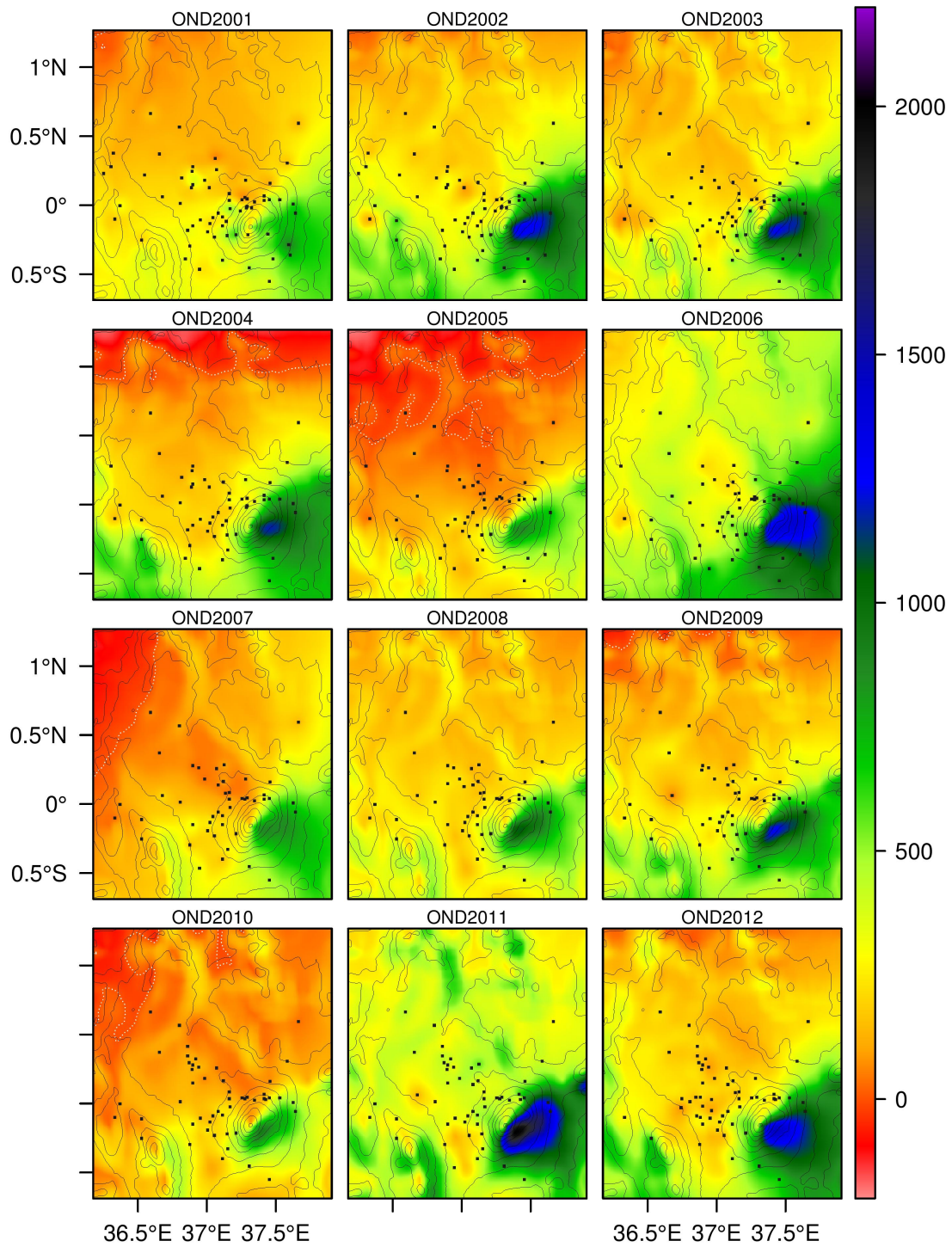


Figure B.7: Final estimations of the OND rainy season from 2001 to 2012. Dots represent the observations used for the estimation.

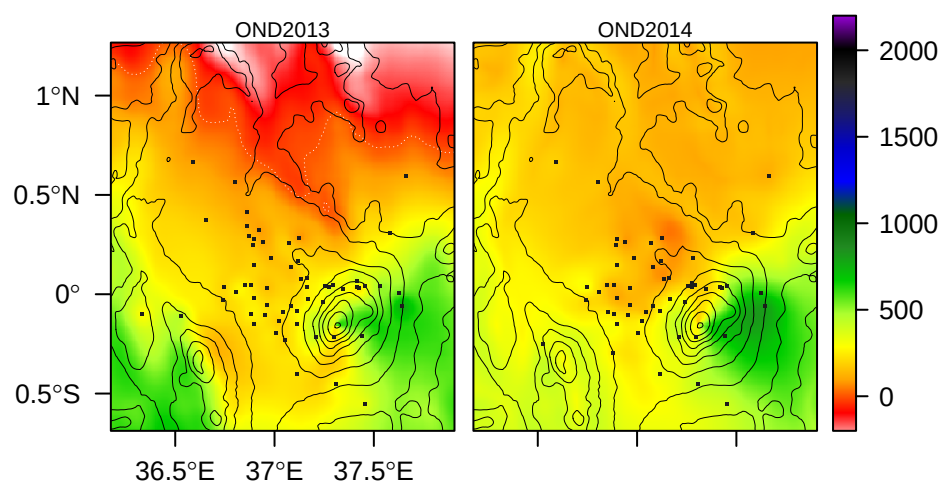


Figure B.8: Final estimations of the OND rainy season from 2013 to 2014. Dots represent the observations used for the estimation.

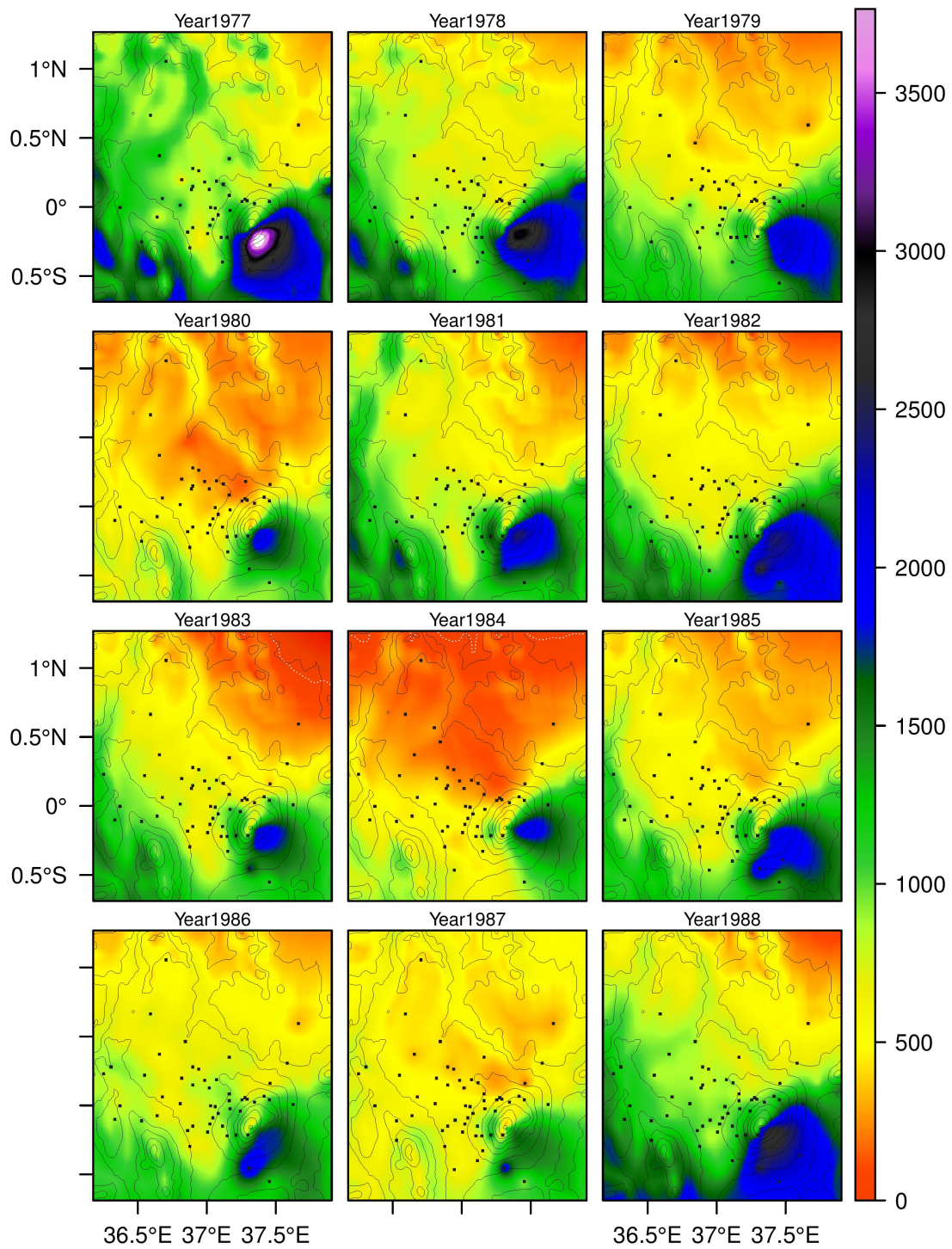


Figure B.9: Final estimations of annual precipitation for 1977 to 1988. Dots represent the observations used for the estimation.

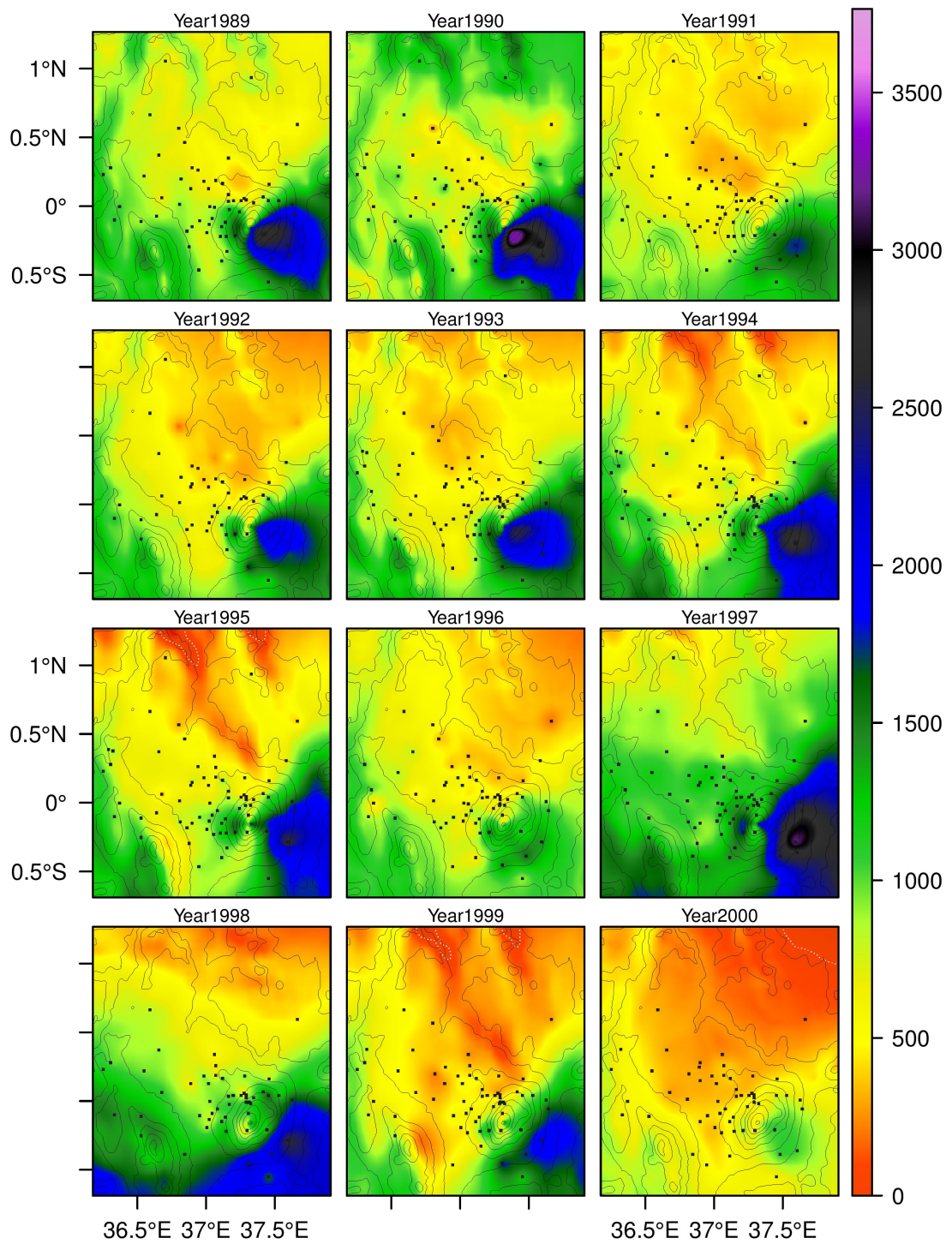


Figure B.10: Final estimations of annual precipitation for 1989 to 2000. Dots represent the observations used for the estimation.

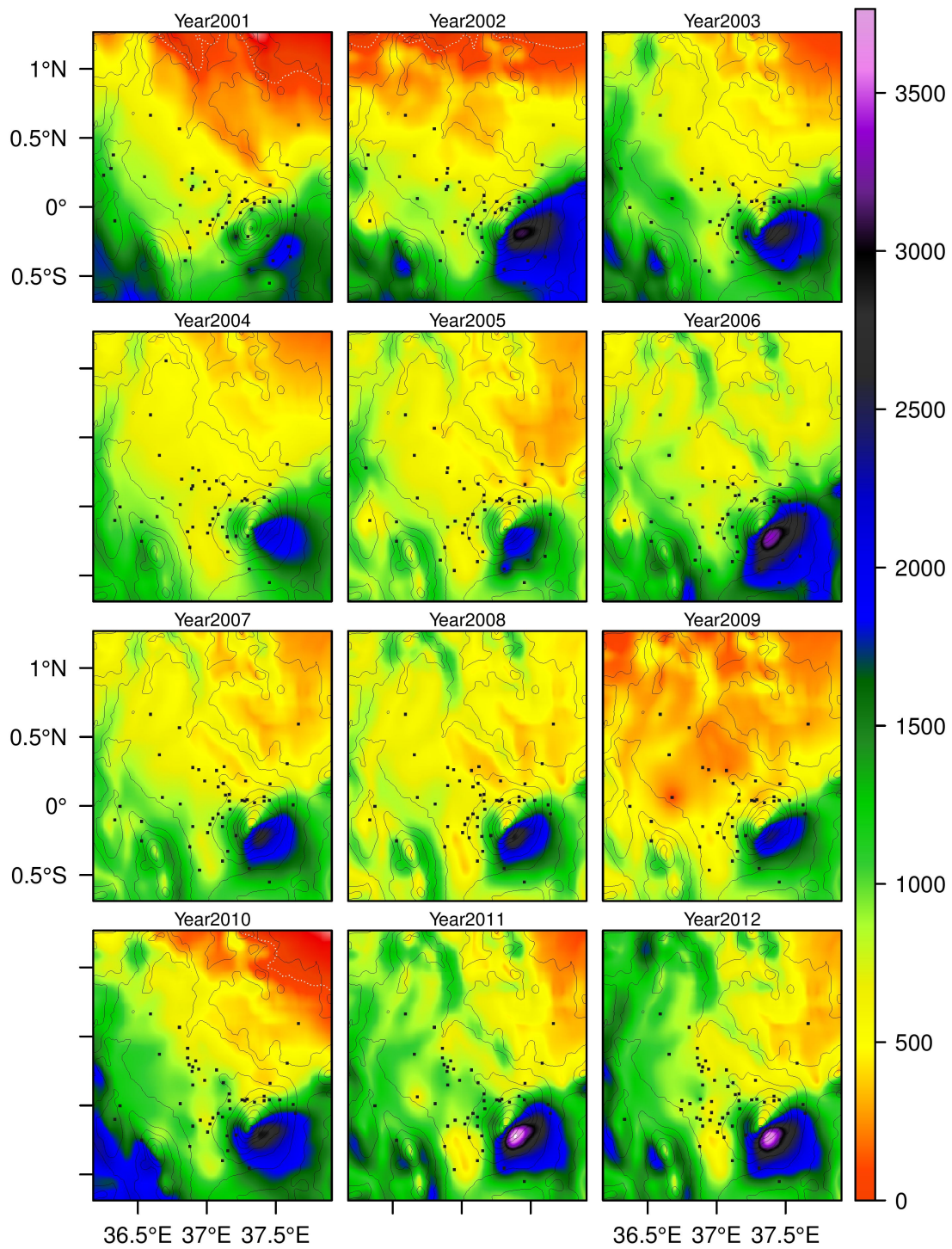


Figure B.11: Final estimations of annual precipitation for 2001 to 2012. Dots represent the observations used for the estimation.

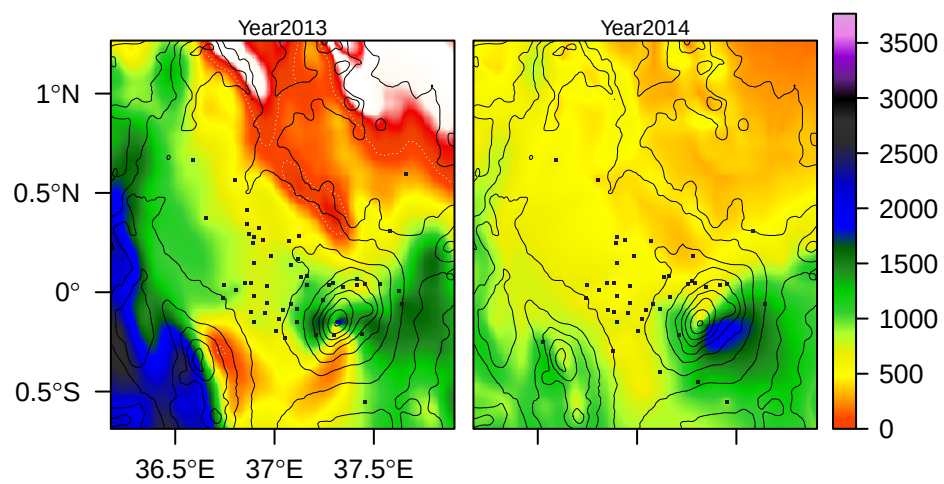


Figure B.12: Final estimations of annual precipitation for 2012 to 2014. Dots represent the observations used for the estimation.

Appendix C

Predictive Skills

Table C.1: Adjusted and predictive R^2 for all years and periods.

Year	adj. R^2 OND	pred. R^2 OND	adj. R^2 MAM	pred. R^2 MAM	adj. R^2 An.	pred. R^2 An.
1977	0,738	0,639	0,826	0,709	0,860	0,723
1978	0,754	0,631	0,822	0,750	0,769	0,657
1979	0,776	0,635	0,676	0,359	0,762	0,455
1980	0,797	0,627	0,758	0,432	0,892	0,772
1981	0,764	0,463	0,617	0,360	0,586	-0,201
1982	0,691	0,552	0,571	0,085	0,612	-0,075
1983	0,709	0,532	0,699	0,420	0,641	0,152
1984	0,823	0,642	0,843	0,713	0,843	0,687
1985	0,713	0,453	0,651	0,349	0,645	0,301
1986	0,702	0,576	0,530	-0,230	0,487	-0,359
1987	0,728	0,662	0,591	0,237	0,503	-0,083
1988	0,822	0,670	0,606	0,242	0,619	0,186
1989	0,860	0,808	0,687	0,486	0,860	0,808
1990	0,825	0,738	0,798	0,710	0,819	0,748
1991	0,847	0,749	0,642	0,381	0,650	0,378
1992	0,845	0,781	0,761	0,630	0,777	0,629
1993	0,802	0,710	0,738	0,581	0,840	0,765
1994	0,884	0,807	0,772	0,668	0,863	0,811
1995	0,846	0,745	0,703	0,508	0,813	0,710
1996	0,778	0,650	0,795	0,620	0,642	0,298
1997	0,838	0,729	0,669	0,518	0,774	0,611
1998	0,533	0,219	0,726	0,523	0,629	0,243
1999	0,812	0,669	0,683	0,489	0,743	0,515
2000	0,774	0,633	0,606	0,440	0,702	0,529
2001	0,652	0,484	0,808	0,664	0,694	0,424
2002	0,763	0,677	0,835	0,761	0,819	0,716
2003	0,845	0,777	0,617	0,482	0,770	0,663
2004	0,875	0,824	0,597	0,482	0,839	0,761
2005	0,863	0,794	0,610	0,427	0,604	0,321
2006	0,828	0,736	0,792	0,671	0,756	0,605
2007	0,855	0,792	0,814	0,640	0,633	0,189
2008	0,847	0,796	0,758	0,566	0,772	0,549
2009	0,804	0,618	0,845	0,764	0,836	0,723
2010	0,855	0,816	0,653	0,560	0,828	0,750
2011	0,745	0,567	0,696	0,498	0,701	0,292
2012	0,853	0,684	0,675	0,380	0,797	0,411
2013	0,681	0,436	0,430	-0,155	0,607	0,262
2014	0,844	0,719	0,688	0,372	0,821	0,328

Bibliography

- Barry, R. G. (2008). *Mountain Weather and Climate*. 3rd ed. Cambridge University Press. Chap. Regional case studies, 363–443.
- Berger, P. (1989). *Rainfall and Agroclimatology of the Laikipia Plateau, Kenya*. Geographica Bernensia, African Studies Series, A7.
- Bostan, P.A. et al. (2012). "Comparison of Regression and Kriging Techniques for Mapping the Average Annual Precipitation of Turkey". In: *International Journal of Applied Earth Observation and Geoinformation* 19, 115–126.
- Burnham, K. P. and D. R. Anderson (2002). *Model Selection and Multimodel Inference*. Vol. 2. Springer. Chap. 1.
- Camberlin, P. et al. (2009). "Components of Rainy Seasons Variability in Equatorial East Africa: Onset, Cessation, Rainfall Frequency and Intensity". In: *Theoretical and Applied Climatology* 98, 237–249. DOI: 10.1007/s00704-009-0113-1.
- Daly, Ch. et al. (2008). "Physicographically Sensitive Mapping of Climatological Temperature and Precipitation across the Conterminous United States". In: *Journal of Climatology*.
- De By, R. A. et al. (2001). *Principles of Geographic Information Systems*. 2nd ed. International Institute for Aerospace Survey and Earth Sciences.
- Dunn, M. and R. Hickey (1998). "The Effect of Slope Algorithms on Slope Estimation within a GIS". In: *Cartography* 27.1, 9–15.
- Funk, C. et al. (2014). "Prediction East African Spring Droughts using Pacific and Indian Ocean Sea Surface Temperature Indices." In: *Hydrology and Earth System Sciences* 18, 4965–4978. DOI: 10.5194/hess-18-4965-2014.
- Funk, C. et al. (2015). "The Centennial Trends Greater Horn of Africa Precipitation Dataset". In: *Nature* 2. DOI: 10.1038/sdata.2015.50.
- Gamoyo, M. et al. (2015). "Rainfall Variability over the East African Coast". In: *Theoretical Applied Climatology* 120, 311–322. DOI: 10.1007/s00704-014-1171-6.
- Gichuki, F.N. et al. (1998). "Knowledge about Highland- Lowland Interactions: The Role of a Natural Resource Information System". In: *Eastern and Southern Africa Geographical Journal* 8.
- Goovaerts, P. (2000). "Geostatistical Approaches for incorporating Elevation into the Spatial Interpolation of Rainfall". In: *Journal of Hydrology* 228, 113–129.
- Goyal, K. M and C. S. P. Ojha (2010). "Evaluation of Various Linear Regression Methods for Downscaling of Mean Monthly Precipitation in Arid Pichola Watershed". In: *Natural Resources* 1, 11–18. DOI: 10.4236/nr.2010.11002.
- Hastenrath, S. et al. (1993). "Atmospheric-hydrospheric Mechanisms of Climate Anomalies in the Western Equatorial Indian Ocean". In: *Journal of Geophysical Research* 98, 20219–20235. DOI: 10.1029/93JC02330.
- Hastenrath, S. et al. (2011). "Circulation Mechanisms of Kenya Rainfall Anomalies". In: *Journal of Climate* 24, 404–412.
- Hengl, T. et al. (2007). "About Regression-Kriging: From Equations to Case Studies". In: *Computers and Geosciences* 33, 1301–1315.
- Hession, S.L. and N. Moore (2011). "A Spatial Regression Analysis of the Influence of Topography on Monthly Rainfall in East Africa". In: *International Journal of Climatology* 31, 1440–1456.
- Hiemstra, P. (2013). *package "automap" for R- Programming*. 1.0-14.

- Huffman, G.J. et al. (2015). *Integrated Multi-satellite Retrievals for GPM (IMERG) Technical Documentation*. Tech. rep. Mesoscale Atmospheric Processes Laboratory, NASA Goddard Space Flight Center, Science Systems and Applications Inc.
- Hwang, Y. et al. (2012). "Spatial Interpolation Schemes of Daily Precipitation for Hydrological Modeling". In: *Stochastic Environmental Resource and Risk Assessments* 26, 295–320. DOI: 10.1007/s00477-011-0509-1.
- Kyriakidis, P.C. et al. (2001). "Geostatistical Mapping of Precipitation from Rain Gauge Data Using Atmospheric and Terrain Characteristics". In: *Journal of Applied Meteorology*. 40th ser., 1855–1877.
- Lau, K. M. and S. Yang (2002). "Walker Circulation". In: *Encyclopedia of Atmospheric Sciences* 4, 53–58.
- Liniger, H. et al. (2005). "Assessing and Managing Scarce Tropical Mountain Water Resources". In: *Mountain Research and Development* 25, 163–173.
- Marquinez, J. et al. (2003). "Estimation Models for Precipitation in Mountainous Regions: the Use of GIS and Multivariate Analysis". In: *Journal of Hydrology* 270, 1–11.
- Masson, D. and C. Frei (2014). "Spatial Analysis of Precipitation in a High-Mountain Region: Exploring Methods with Multi-Scale Topographic Predictors and Circulation Types". In: *Hydrology and Earth System Sciences* 18, 4543–4563.
- Mutai, C.C. and M. N. Ward (2000). "East African Rainfall and the Tropical Circulation/Convection on Intraseasonal to Interannual Timescales". In: *Journal of Climate* 13, 3915–3936.
- Nicholson, S. E. (2015). "Long-term Variability of the East African Short Rains and its Links to the Large-scale Factors". In: *International Journal of Climatology* 35, 3979–3990.
- Notter, B. et al. (2007). "Impacts of Environmental Changes on Water Resources in the Mt. Kenya Region". In: *Journal of Hydrology* 343, 266–278. DOI: 10.1016/j.jhydrol.2007.06.022.
- Odeh, I.O.A et al. (1995). "Further Results on Prediction of Soil Properties from Terrain Attributes: Heterotopic Cokriging and Regression-Kriging." In: *Geoderma*.
- Oettli, P. and P. Camberlin (2005). "Influence of Topography on Monthly Rainfall Distribution over East Africa". In: *Climate Research* 28, 199–212.
- Schabenberg, O. and C. A. Gotway (2005). *Statistical Methods for Spatial Data Analysis*. Boca Raton: Chapman & Hall/CRC.
- Schlatter, L. (2016). "Precipitation patterns analysed with Markov Chains for selected sites around Mount Kenya based on long-term rainfall records". Institute of Geography.
- Schmocker, J. (2013). "Long-term Precipitation Trends and Rainy Season Performance in View of Climate Change in the Mt. Kenya Region". MA thesis. Institute for Geography.
- Schmocker, J. et al. (2015). "Trends in Mean and Extreme Precipitation in the Mount Kenya Region from Observations and Reanalyses". In: *International Journal of Climatology*.
- Sharples, J. J. et al. (2005). "On the Horizontal Scale of Elevation Dependence of Australian Monthly Precipitation". In: *Journal of Applied Meteorology* 44, 1850–1865.
- Sturm, B. (2002). "Development and Use of Long- and Short-term Precipitation Interpolation Models in the Ewaso Ng'iro Basin". MA thesis. Institute of Geography.
- Webster, R. and M. A. Oliver (2007). *Geostatistics for Environmental Scientists*. Ed. by V. Barnett. 2nd ed. Statistics in Practice. John Wiley and Sons, Ltd.
- Winiger, M. (1981). "Zur thermisch-hygrischen Gliederung des Mount Kenya". In: *Erdkunde* 35, 248–263.
- WMO (2008). *WMO Guide to Meteorological Instruments and Methods of Observations*. WMO-no. 8. WMO. Chap. Measurement of precipitation.
- Yang, W. et al. (2014a). "The Annual Cycle of East African Precipitation". In: *Journal of Climate* 28, 2385–2404. DOI: 10.1175/JCLI-D-14-00484.1.
- Yang, W. et al. (2014b). "The East African Long Rains in Observations and Models". In: *Journal of Climate* 27, 7185–7202. DOI: 10.1175/JCLI-D-13-00447.1.

Declaration

under Art. 28 Para. 2 RSL 05

Last, first name: Imfeld, Noemi

Matriculation number: 09-062-621

Programme: Master in Climate Sciences

Bachelor

Master

Dissertation

Thesis title:

Modeling Seasonal and Annual Precipitation using Longterm Climate Records and Topography

Thesis supervisor: Prof. Dr. Stefan Brönnimann

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person, except where due acknowledgement has been made in the text. In accordance with academic rules and ethical conduct, I have fully cited and referenced all material and results that are not original to this work. I am well aware of the fact that, on the basis of Article 36 Paragraph 1 Letter o of the University Law of 5 September 1996, the Senate is entitled to deny the title awarded on the basis of this work if proven otherwise. I grant inspection of my thesis.

Bern, 04.07.2016