Table of Contents

Table of Contents

Tables and Figures

List of Abbreviations

CCN	Cloud Condensation Nuclei
CDE	Centre for Development and Environment
CPU	Central Processing Unit
CRT	Cathode Ray Tube
DTM	Digital Terrain Model
ENSO	El Niño Southern Oscillation
FFT	Fast Fourier Transformation
GCV	Generalised Cross Validation
GIS	Geographical Information System
GNU	GNU's Not Unix
GPL	GNU General Public License
GPS	Global Positioning System
GRASS	Geographical Resources Analysis Support System
HW	Hardware
IDW	Inverse Distance Weighting
ITCZ	Inter Tropical Convergence Zone
KLIMET	Gruppe für Klimatologie und Meteorologie
Linux	Linux is not Unix
LRP	Laikipia Research Programme
MAE	Mean Absolute Error
ME	Mean Error
MRE	Mean Relative Error
MS	Microsoft
MSE	Mean Squared Error
NRM^3	Natural Resource Monitoring, Modelling and Management
RST	Regularised Splines with Tension
SQL	Structured Query Language
SW	Software
TPS	Thin Plate Splines
VBA	Visual Basic for Applications
WMO	World Meteorological Organisation

1 Introduction

1 Abstract

The main focus of this work is on the development and use of a method for the simple and fast production of rainfall maps by means of interpolation from point measure rainfall to any part of an area.
As a starting point, rainfall raw data that was retrieved from a relational database had to be quality controlled, and any missing values were to be re-constructed through the use of a linear regression method. The selection of the used gauges is crucial in terms of covered time and area, as time and space resolution is directly connected to the quality of the resulting spatial rainfall interpolation.
As a consequence, two different datasets had been chosen for this work. Both sets cover different time spans (a longer time span from 1967-1997, and a shorter time span from 1990-1997) and contain different rainfall gauges representing different spatial resolutions.
In a next step the selected data was imported into a GIS for the purpose of the spatial interpolation process. Before the interpolation took place, the best interpolation model had to be selected, which was performed through the use of specially written validation software.
The result of the validation was used for the production of all rainfall maps presented in this study.
From the analysis and application of the above described methods the following conclusions can be drawn from this work:

1. The proposed method for a step-by-step interpolation of rainfall data can be applied to most datasets, and seems to be suitable for a fast and sound analysis in the field.
2. The proposed method for the interpolation of rainfall is based on very low investments in terms of money and time, and offers therefore a great potential for developing countries.
3. The new rainfall maps of the Upper Ewaso Ng'iro basin in Kenya are more detailed than previously produced ones, and can be used for more detailed analysis such as temporal rainfall pattern analysis.
4. A first temporal analysis of the rainfall data revealed a higher variability in the rainfall for the last 10 years. This variability was also mentioned by the farmers interviewed during the fieldtrip in late 1998.

1

A last chapter of this paper deals with the interpretation and the discussion of the produced precipitation maps. This section also includes a short comparison of the awareness of local people in the region towards annual rainfall behavior and the observed and estimated rainfall regime by the proposed method.

1.1 Acknowledgement

I would like to thank to the following institutions and persons who made this work possible: The Centre for Development and Environment (CDE), at the University of Berne, Switzerland. The NRM3 Project Headquarter in Nanyuki, Kenya. The group for climate and meteorology (KLIMET), at the University of Berne, Switzerland. Dr. Hans-Peter Liniger (CDE), Prof. Hans Hurni (CDE), Dr. Dimitrios Gyalistras (KLIMET), Prof. Heniz Wanner (KLIMET), Gudrun Schwilch (CDE).

1.2 Introduction

Water forms the most important resource for any life form on this planet. This is particularly true for human beings, as the abundance of water can guarantee the wealth and health of every culture.

The existence and availability of water sets the limiting natural factor for any prosperous development of living organisms. Where the abundance of water is not given or uncertain, it was and is important to people to control and manage the scarce resource water.

The irrigation of fields and the drainage of swamps together with the control of the floodwaters were often closely connected with the development of complex organised societies such as the cultures that emerged in the lower Tigris-Euphrates valley around 4750B.C. (KEESING, 1981: 50). The more complexly organised a society, the more important becomes the knowledge about water and its availability.

This is especially true for regions where subsistence relies mainly on rainfed agriculture, and where high rainfall is limited to certain periods of the year. These are the typical features of tropical rainfall regimes.

The Upper Ewaso Ng’iro Basin in Kenya forms such a region. The river Ewaso Ng’iro is fed from sources situated at high altitudes at the Mt. Kenya (5199m) that receives, due to orographic rainfall, the highest amount of rainfall in the region. This makes, combined with good volcanic soils, the slopes of the mountain a highly fertile and favoured area for rainfed agriculture.

1

Downstream the Ewaso Ng'iro, in the drier lowlands, traditional livestock subsistence can be found, where upstream high-intensive modern flower or cabbage farms owned by big multi-national companies are located.
The pressure on land, and as a consequence, on water, became in the past years a dramatic importance due to a high-population growth rate, and the subdivision of large farms to small family owned plots.
All these human and environmental factors together mark a highly sensitive highland-lowland ecosystem. And it is obvious that the knowledge about the regional distribution of rainfall (as the main input factor for water) can be vital to the habitants of the Upper Easo Ng'iro Basin.

The herewith-presented study is a master thesis written for the purpose of providing a reliable tool for the production of rainfall maps. Such rainfall distribution maps produced with the methods and models presented in this work, can help to take decisions in a variety of issues. They can be used for future land-use and land-management issues, or they might reveal any past or ongoing changes in the local precipitation distribution pattern.
Therefore, the focus of this thesis is mainly on the development and application of spatial rainfall interpolation models.

It shall also be noted, that the presented methods are not only applicable for the study area in Kenya, but can be transformed without too many restrictions to other tropical or sub-tropical regions of the earth.

Bernhard Sturm
Berne, November 2001

1.3 Aims of the Study

The overall goal of the this study was to develop and apply a rainfall interpolation method which can be generally used, the specific objectives were:

- To develop a simple method for the creation of precipitation maps in general.
- To assess the quality of the measured rainfall and fill data gaps.
- To estimate rainfall within a reasonable amount of time and at the lowest possible costs in terms of money.
- To find out and analyse the spatial and temporal precipitation pattern over the study area based on recent precipitation data and on the developed interpolation method.
- To produce high temporal and spatial precipitation maps.

During the analysis of the raw precipitation data it became apparent that the data was far from being consistent and complete. A way had to be found to correct these inconsistencies without the use of too sophisticated geo-statistical methods. As the main emphasis on the work was set on an 'easy-to-use', and handy method, software had to be written that enabled future users to perform these tasks within reasonable time.
This is also the reason why some chapters were intentionally written in a manual or handbook stylish manner.
If an interested user will be able to follow the herein described steps and methods, then the most important goal of this work was achieved: to provide a useful method for the interpolation of rainfall maps.

1

1.4 Introduction into the Study Area

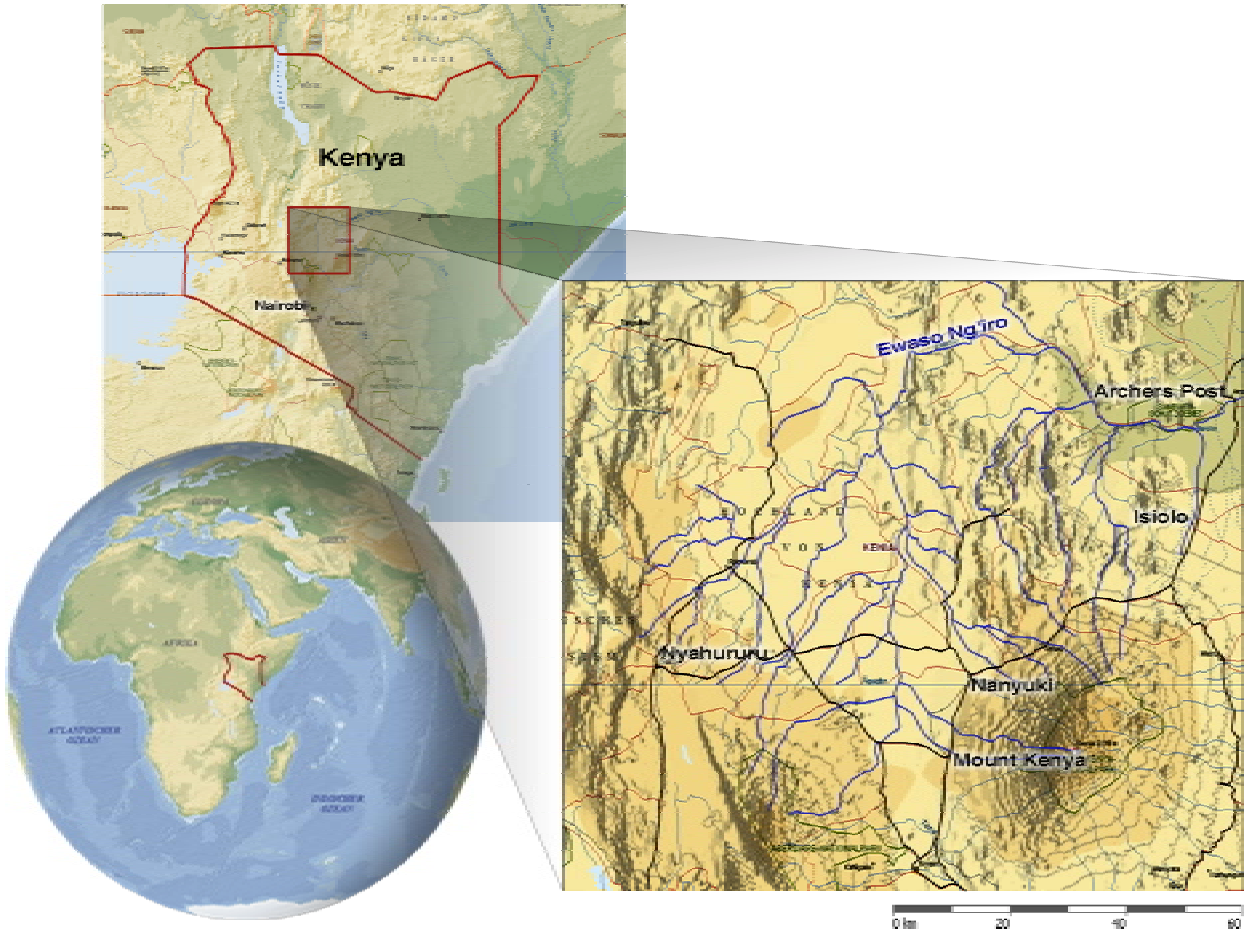


Figure 1.1: Location of the Upper Ewaso Ng'iro Basin (Source: Microsoft Encarta '99, BERGER 1989, B. Sturm)

1.4.1 Location and Zonal Differentiation

Figure 1.2: View from northwest towards Aberdares Hills and Mt. Kenya (in the background). Computer generated image based on a 100m resolution DTM. (Image by B. Sturm, made with the terrain modelling software Terragen, vegetation information was computed depending on slope angle and elevation)

„The Upper Ewaso Ng'iro Basin is located to the North and West of Mt. Kenya (see topographical overview). This area extends between longitudes 36°30' and 37°45' East and latitudes 0°15' South and 1°00' North. The basin covers an area of 15'200km², [...] approximately 2.8% of the total area of Kenya (587'900km²). The area falls under seven administrative districes (Nyandarua, Nyeri, Laikipia, Meru, Samburu, Isiolo and Nyambene)" (GICHUKI et al., 1998: 6)

In order to describe the varying environmental characteristics of the study area, a zonal distinction was introduced (GICHUKI et al., 1998). This differentiation divides the Upper Ewaso Ng'iro basin into six zones following a transect from the mountains to the lowlands. According to the changing altitudes, and the strongly linked precipitation and growing conditions for vegetation, the basin was divided into the major zones described in table Table 1.1 on page 14 (mainly taken from NIEDERER, 2000 and GICHUKI et al., 1998).

Figure 1.3: Alpine zone at approximately 4000mAMSL on the Mt. Kenya (Teleki Valley). Image by B. Sturm.

Figure 1.4: (below) 360° panoramic view from a position close to Mukogodo (UTM coordinates: Zone 37N, Easting: 283828.1877 Northing: 39054.2705). Image by B. Sturm.

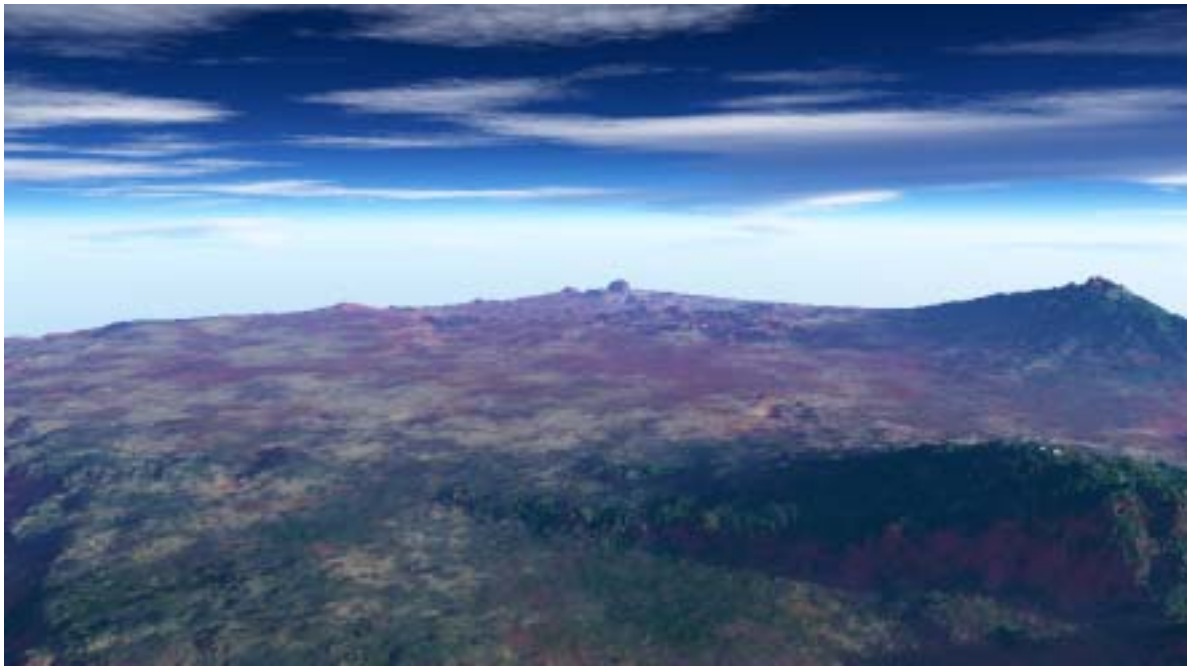


Figure 1.5: Computer generated overview over the study area towards northeast based on a 100m DTM. The observer's position is high above the southwest corner of the area. In the background on the right side Mt. Kenya is visible, the mountaineous feature in the foreground is the Nyandarua range. Elevation is three times exaggerated. (Image by B. Sturm, made with the terrain modelling software Terragen, vegetation information was computed depending on slope angle and elevati

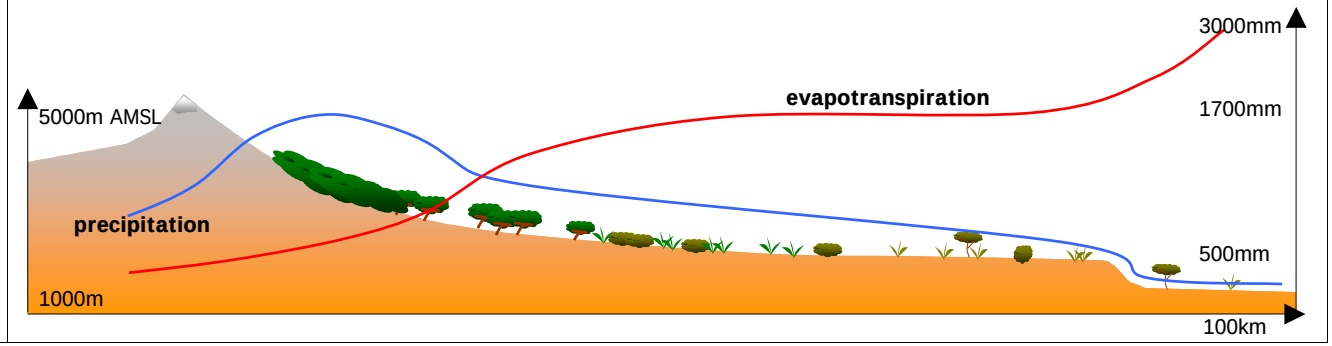
	Alpine Zone	Upper Mountain Slopes	Lower Mountain Slopes	Volcanic High Plateau	Highlands on Basement Complex	Lowlands
Altitude (mAMSL)	5200 - 3500	3500 - 2200	2200 - 1900	1900 - 1700	2500 - 1200	1200 - 800
Annual Rainfall	600 - 1000 mm	1000 - 1500 mm	700 - 1000 mm	500 - 700 mm	500 – 600 mm	300 - 500 mm
Pot. Evaporation	100 - 500 mm	500 - 800 mm	800 - 1200 mm	1200 – 1800 mm	1700 - 2500 mm	2500 - 3300 mm
Rainfall and water availability for plants	Rainfall excess during the wet seasons. Periods of water deficits during dry season.		If conserved, rainfall is sufficient for rainfed agriculture.	Increasing variability and decreasing reliability of rains and thus a lower potential for rainfed agriculture.		Very erratic rainfall with long periods of water stress.
Vegetation	Tussok Grassland	Moorland Ericaceous Montane Forest	Forests / Cropland		Grassland, savannah vegetation	Open savannah vegetation, overgrazed grassland
Soils	Regosols, Leptosols	Humic Andosols	Humic Acrosols, Nitosols, Ferric Luvisols	Heavy Vertic	Volcanic Origin, medium to low fertility	Moderate Deep Soils
Land Use	-	Small Scale Cropland	Rainfed Agriculture (Maize, Potatoes, Wheat, Beans)	Agroforestry, Cropland	Ranching on large scale farms, High grazing pressure	Extensive Grazing
Topography						

Table 1.1: Natural resources and major zones of the Upper Ewaso Ng'iro Basin (after NIEDERER 2000, and GICHUKI et al. 1998).

1

Figure 1.6: Montane forest on the eastern slopes of the Mt. Kenya (Image by B. Sturm)

At an altitude between 3300 - 3500 m, above the timberline, we find the vegetation to be formed typically by senecias and lobelias (see Figure 1.3) (BERGER, 1989: 22).

1.5 Climate

Because of its location and altitude the climatic situation of the higher parts of the Upper Ewaso Ng’iro area can be classified to a tropical highland climate (except for the lower parts of the plateau). The plateau is situated in a zone of transition from a wetter to a drier regime (BERGER 1989: 22). The mean annual rainfall reaches between 1500mm in the forest areas of Mt. Kenya and 350mm in the arid low-lands (GICHUKI et al.: 1998).
The mean annual temperatures lie between 16°C and 20°C. Temperature decreases with increasing altitude and the peaks of Mt.Kenya are glaciated.
In order to understand the specific precipitation pattern of the study area it is of importance to understand the general climatic circulation modes of the tropics.

1.5.1 General Climatic Situation

The general climatic situation for east Africa region is under the influence of the ITCZ (Inter Tropical Convergence Zone). As for all tropical climates the following features are playing an important role (after JACKSON, 1979 and HENDERSON-SELLERS et al., 1996):

- The Hadley Cell regime
- The trade winds
- The equatorial trough
- The south-east Asian monsoon

The Hadley cell regime is part of the - strongly simplified - three-cell circulation model of the earth’s atmosphere. The model fulfils the basic functions of the general circulation to redistribute energy and mois-

ture without changing the angular momentum or mass balance of the planet (HENDERSON-SELLERS et al., 1996). The model is based on the observations that there were zonal belts of low and high-pressure systems distributed on the globe. A converging low-pressure system can be found at the equator and at around 60° latitude, a diverging high-pressure belt is located at 30° latitude and on the poles. From these observations a three-cell model was derived, where each cell transports mass and energy to and from the adjacent pressure zones (Figure 1.7). However, this model was no longer supported as more and detailed empirical global meteorological data was available. Satellite-derived information suggested that the cell model of the atmosphere was, especially for the mid-latitudes, wrong, and a more appropriate wavelike model was introduced (the *planetary scale* or *Rossby* waves). For the tropics, however, we can still use the older cell model, as it describes adequately enough the less complex energy and water redistribution for the tropics (HENDERSON-SELLERS et al., 1996: 178).

The Hadley Cell Regime

The thermally direct tropical, or *Hadley*, cell generally exists within an equivalent barotropic atmosphere. Because of the missing distinctive thermal winter and summer, the influence of the continents are minimal at these low latitudes, therefore we can find a nearly zonal uniform near-surface temperature distribution. There are two heat sources:

For the lower atmosphere: solar radiation absorption through the earth's surface
Within the ascending air of the ITCZ: latent heat released by convective motions.

The heat sinks of the regime can be found on the pole ward limb of the cell and at the top of the atmosphere. The driving force of the Hadley cell is the continuous supply of heat to the equivalent barotropic atmosphere.

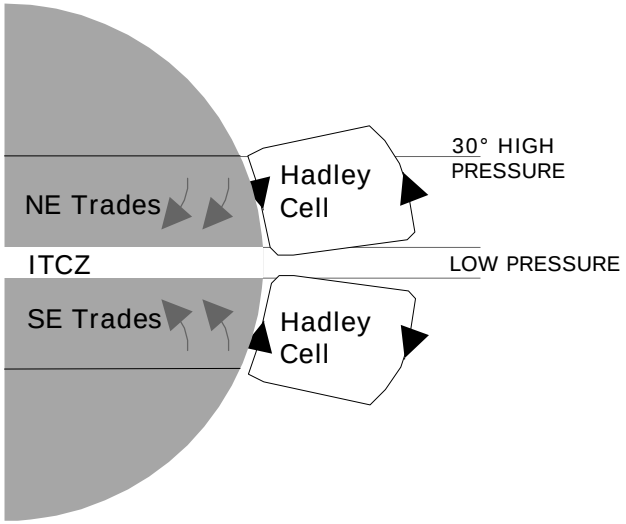


Figure 1.7: Older general circulation cell model for the tropics (after HENDERSON-SELLERS, 1996: 175)

The Trade Winds

The horizontal components of the Hadley cell are only slightly affected by the Coriolis deflection at these low latitudes, so that the resulting winds, the *trade winds*, do not blow parallel to the isobars, but have a marked cross-isobaric component, blowing equator ward at low levels and pole ward in the upper troposphere. This wind pattern allows therefore a direct transfer of energy away from the equator. The low level airflows converge at the ITCZ, which is formed by a belt of low pression known as the *equatorial trough*, and rise to create cloud and precipitation, hence a region of instability.

Virtually all of the rest of the tropics are subject to more or less subsidence inversions in which warm descending air 'traps' somewhat cooler air right at the surface. Because of this feature, known as the *trade wind inversion*, the surface trade winds are able to pick up moisture as they flow over the ocean, but they cannot release it, or the associated latent energy, until they reach the ITCZ. Precipitation from the ITCZ clouds is almost entirely convective. This happens on two different scales: on a small-scale by small groups of convective cells, and on a larger scale by features, which have the potential to become hurricanes.

1.5.2 The South-east Asian Monsoon

The monsoon circulation is mainly fed from surface temperature differences originating from a favourable distribution of land and sea mass, and is represented by an on-shore wind system delivering rain to the Indian subcontinent in the boreal summer. Associated with the monsoon is a low-level jet discovered by John Findlater that influences much of the southeast African climate during the monsoon (The *East African Low Level Jet*) (HASTENRATH, 1991: 134). The jet is characterised by a month-by-month progression, has core speeds of 12-15 ms^{-1}, is centred at

around 1.5km height and must be regarded as an integral part of the Southwest monsoon current. The jet plays an important role in the cross-equatorial water vapour transport[1] and influences directly the climatic situation in the Upper Ewaso Ng'iro basin during the boreal summer months.

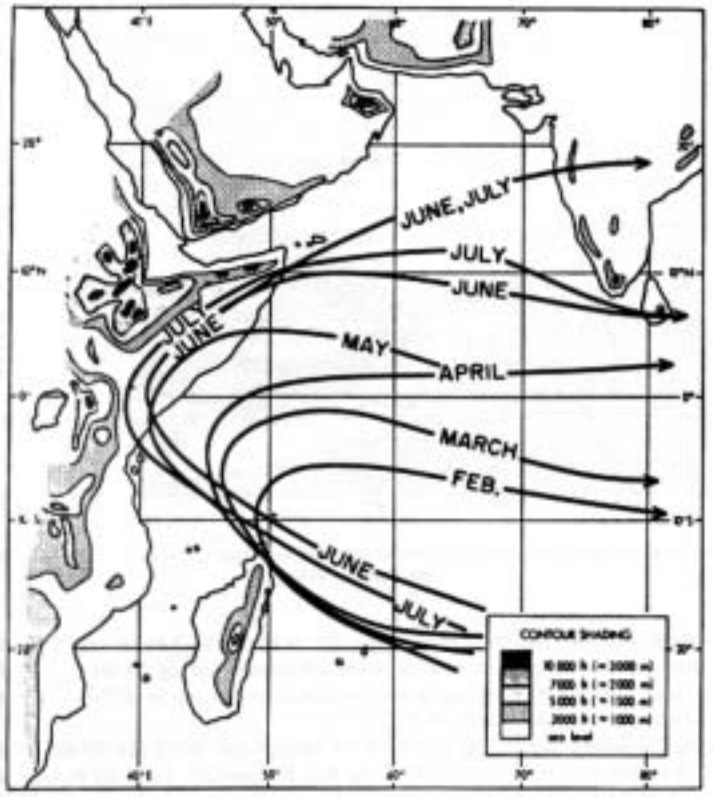


Figure 1.8: Month-by-month progression of the axis of the East African Low-Level Jet. Source: Findlater (1971), taken from Hastenrath (1991: 135)

1.5.3 The Rainfall Pattern

Because of the low latitude location of the study area, the precipitation pattern of the Upper Ewaso Ng'iro basin is determined by

[1] „Findlater (1969b) estimates the cross-equatorial mass transport of the East African Low Level Jet in July at $7x10^{10}$kg s^{-1}, which would amount to about half of the total global lower-tropospheric mass flow across the Equator." (HASTENRATH, 1991: 136) However there seem to be divergent empiric studies about the measured core-speeds of the Jet. Even Findlater himself presents his own measurements in a contradictory way (on some maps Findlater indicates a speed maximum over the Kenyan coast, then in other maps this is labelled as a speed minimum (see Hastenrath, 1991 for a detailed discussion).

1

the seasonal position of the ITCZ and its related climatic processes. Two distinguishable rainy seasons have been identified for the East African region: the 'long-rains' occurring from March to May and the 'short-rains' in October and November. These rainy seasons are in accordance with the seasonal position of the ITCZ. Hastenrath points to the fact, that the two rainy seasons fall into the transition between the two monsoon circulations, and Kiangi et al. (1981) showed that the airstream originating in the Southern hemisphere druing April and May *„[...] being undercut by the Northeast monsoon flow in the region of Equatorial East Africa. Conversely in October-November in East Africa, southeasterlies wedge under the northeasterly current from the Northern hemisphere, along a confluence zone then again extending across Equatorial East Africa"* (HASTENRATH, 1991: 193)

In the Upper Ewaso Ng'iro basin the high mountain areas of the Nyandarua Range and Mt. Kenya break up this two seasonal pattern by forcing the trade winds orographically upwards, which results that these areas receive rainfall in other periods from the trade winds.

The following section discusses main circulation features over the study area for the selected months of January, April, July and October. This overview is taken from BERGER (1989):

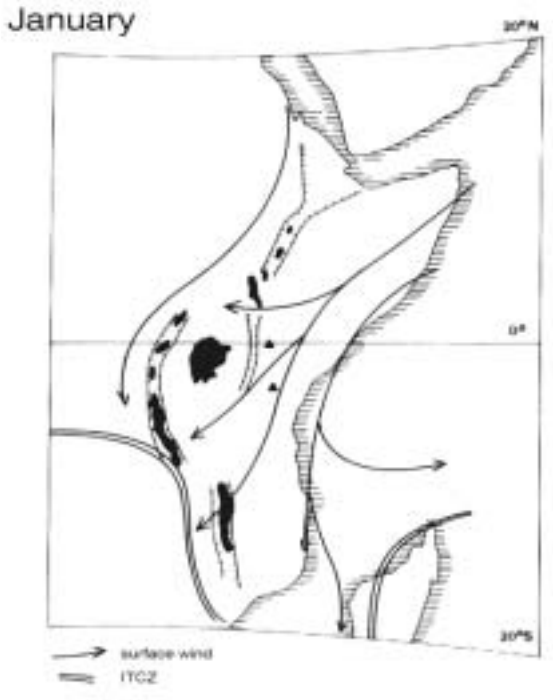


Figure 1.9: Mean atmospheric circulation over East Africa for January (taken from BERGER (1989:26)

With the ITCZ in its southern position, the Eastern Kenyan Highlands are under the influence of a divergent wind field, delivering dry air masses, which originate from Saudi Arabia, and travelled over a long distance over Somalia and Northern Kenya. Therefore almost no rainfall can be observed over the study area.

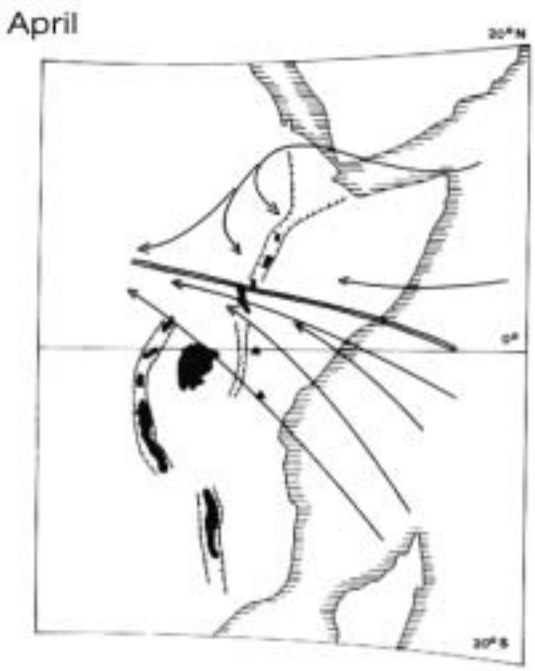


Figure 1.10: Mean atmospheric circulation over East Africa for April (taken from BERGER (1989: 26).

The ITCZ has moved towards the north, and has crossed the equator. This transition is marked by a change in the circulation pattern: the diverging dry north-east air masses converge with air masses originating from the southern hemisphere. The change from divergent to convergent circulation makes high-reaching convection possible and the rainy season sets in. April is therefore a period of extended rainfall in the study area.

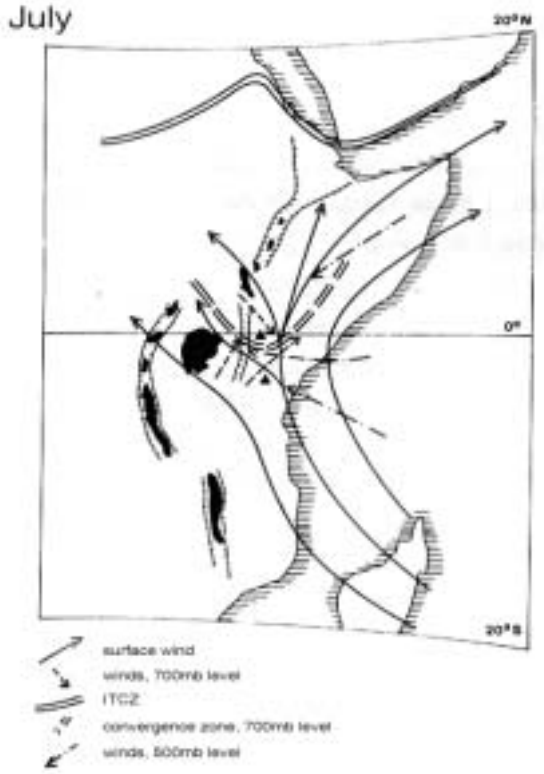


Figure 1.11: Mean atmospheric circulation over East Africa for July (taken from BERGER (1989: 26).

As the ITCZ has reached its highest position on the northern hemisphere the study area

has come under the influence of air mass currents on different pressure levels. On the 700mb level (approx. at 3000 mAMSL) a discontinuity between winds from the north-west and winds from the south-east exists, but these winds have a general western component. The 500mb pressure level is dominated by easterly winds. On the surface layer a divergent current from the south and south-east can be observed. According to Berger this is caused by an area of low pressure over the Lake Victoria, but they could also be seen in the context of the East African Low Level Jet described by Findlater.

However, in July and August, the central and western parts of the study area receive the 'continental rains', but they cannot be explained with the change from trade wind to ITCZ dominated circulation as the equatorial (west) winds are restricted to the lowest layer of the troposphere and cannot rise to the Highlands of East Africa.

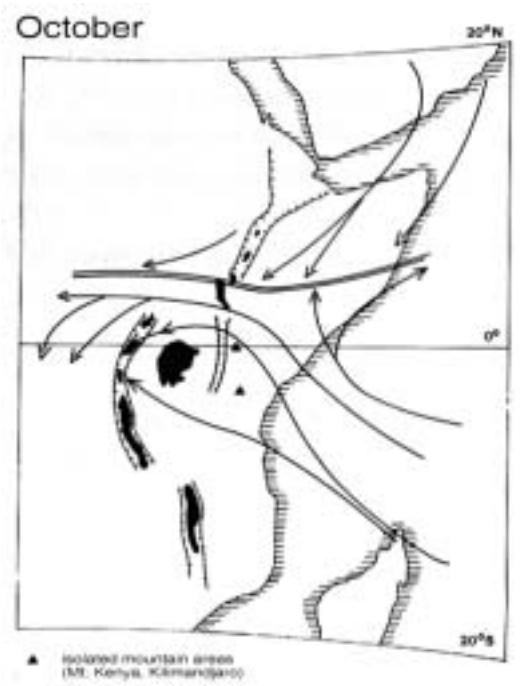


Figure 1.12: Mean atmospheric circulation over East Africa for October (taken from BERGER (1989: 26).

Again the change from the divergent to a convergent circulation marks the month of October. The ITCZ moves southwards and the meteorological equator influences the overall climatic situation in the study area. This is marked by the onset of the second rainy season, the 'short rains'. As the ITCZ

moves further southwards the region comes into the influence of the north-east trade winds, the 'short rains' end.

1.6 Socio Economic Situation

In pre-colonial times, the footzones of the Mt. Kenya, the Laikipia Plateau, and the Nyandarua Mountains were used as grazing land by semi-nomadic Maasai pastoralists. After the colonialisation of the area at the beginning of the 20th Century, European settlers drove away the Maasai pastoralists. During the colonial period, large-scale cattle ranches and mixed farms with dairy and crop production dominated the district. The white settlers occupied the highlands completely: Except for one reserve at Mukogodo, Africans were not allowed to stay in Laikipia[2], unless they found work on one of the colonial farms. Consequently the colonised area was called the „White-Highlands" (FLURY, 1987: 1). After the independence from the British Commonwealth in 1963, the new government put the issue of the Africanisation of land as a main priority on its political agenda. As the white settlers left the land, the population pressure on the new land rose. In this context, intensification in the land-use pattern took place: small holders, mainly from high potential areas, shifted from an extensive to an intensive land-use with crop production. This change took also place in the lower potential areas of the district, where cattle grazing would have been more suitable than crop production.

In the years following the independence, the young Kenyan government introduced so-called „Settlement Schemes" in the high-potential zones on the slopes of the mountains. The intention of those schemes was to subdivide the former big white farms to the incoming small holders on the base of long-term land purchase loans.

But the tremendous demand for the land has changed the situation in the Laikipia District since the 1970s dramatically: speculation and the following illegal subdivision of farms and ranches produced uneconomic small plots (2-5 acres; 0.8-2 hectares) regardless of their production potential. This forced the small holders to intensify the use of land. Compared to the former pastoralists, the small-scale mixed agricultural system was to produce six times more food on the same size of land[3]. This pressure on

[2] Laikipia is one district in the Upper Ewaso Ng'iro Basin.

[3] „Where the pastoralists expected to get the food for one person out of the production from 25 acres (10 hectares, livestock production, meat, milk, blood; figures for Mukogodo, Division in Northern Laikipia), small-scale mixed agriculture with crop production and animal husbandry produces the food for one person, in an average year from about 4 acres (1.6 hectares)." (FLURY, 1987: 1)

scarce land increased in the last years even further for various reasons. One of the most important driving forces is immigration (the population growing rate is as high as 7% for the Upper Ewaso Ng'iro Basin), but also the recent torrential rains of the El Niño (1998) damaged infrastructure such as roads, and forced people to abandon their plots and move to other zones (NIEDERER, 2000).

There are multiple impacts on the socio-economic life of the people and the ecology in the region: by intensifying the land-use system these small holder families are more vulnerable to the risks of short or missing rainy seasons, and need therefore to compensate these risks with off-farm and on-farm strategies. Due to the lack of rainfall water, small holder farmers began to draw water from perennial rivers, thus led to a reduction of the discharge of the Ewaso Ng'iro of more than 50% during the dry season (GICHUKI et al., 1998). Illegal upstream water abstractions are pushing the downstream pastoral communities, their livestock and the wildlife in the surrounding lowland further to the edge of their existence. Off-farm strategies include working for one of the large-scale farms located on the highly fertile northeast slopes of the Mt. Kenya. The social implications of these strategies can lead to disrupted families or the lack of manpower on the home fields (MWITI, 1998).
Because of the fast changing land-use pattern, and the subdivision of land, the Masaai pastoralists were no longer able to follow the rainfall regimes in search of water and pastureland. This disrupted shifting pattern led to overgrazing and there are now serious signs of degradation such as gully erosion and devegetation (KITEME, 1998: 25 and LINIGER & THOMAS, 1998) (see Figure 1.13).

Figure 1.13: Gully erosion at Ol Donyo farm. Image by B. Sturm.

All the described processes are typical for an interacting highland-lowland system such as the one of the Mt. Kenya region. Because of the high dynamics and the conflicting situations between the different users of the natural resources in the Upper Ewaso Ng’iro basin it is of importance to deepen the knowledge of the status of the natural resources and the assessment of the impacts of human activities in such a sensitive eco-system.

Figure 1.14: Pressure on land: deforestation on the Western slopes of the Mt. Kenya. Image by B. Sturm.

2 Precipitation: From the Drops to the Numbers

2 Precipitation: From the Drops to the Numbers

In this chapter we will discuss how and where rainfall is generated and measured. The first paragraph will therefore deal with a general description of the process of condensation and cloud development.
At a later section in this chapter, the quality and standards a precipitation gauging network has to meet will be discussed. As the NRM[3] project in Kenya maintains a natural resource database as a decision making tool, we will elucidate the role of this database in the light of gauging rainfall.

2.1 Condensation and Cloud Formation

Air can only contain a certain amount of vapour at certain temperatures. And vapour will sublimate to ice or condensate to water and thus generating the visible forms of clouds. Condensation and sublimation are the conditions for the process of cloud formation and precipitation.
Condensation can occur for the following reasons (after LAUER, 1993: 75):

- Decreasing air temperatures during ascent, but at a constant volume of a parcel of air.
- Increasing air parcel volume without adiabatic[4] heat input.
- When saturation is reached through decreasing temperatures and reduced air parcel volume, or the constant input of moisture.

There are a couple of ways to force a parcel of air to ascent in order to produce one of the above reasons:

- Through *orographic* uplift (by topographic features such as mountains and ridges)
- Through *frontal* uplift (by two contacting air masses of different temperatures)
- Through the *confluence* of two airstreams (commonly associated with depressions in mid-latitudes)
- Through *convection* (as a result of the hydrostatic instability of the atmosphere.)[5]

[4] *"'Adiabatic' process is one in which the heat transfer vanishes. Processes involving heat transfer are referred to as 'diabatic'. […] Consider an air parcel moving vertically between two levels. If vertical motion occurs on a time scale short compared to that for heat transfer, the process may be regarded as adiabatic.[…] Hence, the potential temperature for an air parcel is conserved for adiabatic motion"* (TRENBERTH, 1995: 73)

[5] Strictly speaking convection has to be treated in connection with the orographic and the frontal uplift processes, as both are able to initiate convective ascent as well. Convection is always associated to tem-

When a cooling air parcel reaches saturation, condensation does not occur immediately. This is true for completely, particle free air where relative humidity could reach several hundred percent before spontaneous condensation would occur. In reality even clean air is not entirely free of particles, and certain particles act as *cloud condensation nuclei* CCN, promoting condensation at relative humidities close to 100% (HENDERSON-SELLERS et al., 1996: 128). The CCN are mainly responsible for the formation of rain drops in clouds as their size and density control the form of visible condensation.
Two effects are forcing a droplet to grow: the *solute effect* and the *curvature effect*. If a CCN is dissolved by the water, the air surrounding it may be saturated at relative humidities <100%, and further condensation is possible. This behaviour is described by the solute effect. The other, simultaneously occurring effect, the curvature effect, requires that the air be supersaturated before further growth can occur. The curvature effect is caused by the non-planar surface of the droplet leading to increased surface tension.
It is of importance to notice that a cloud must contain a variety of different sized CCN to increase the likelihood of producing droplets of sufficient size to grow into raindrops (so called 'activated droplets').
This growth is caused by direct condensation and simultaneously occurring collision and coalescence with other droplets. If the droplets are of considerable size and weight, the Earth's gravity will force them to fall out of the clouds and fall as precipitation to the surface.
At mid-latitudes the observed rain resulting from this process is usually described as drizzle or freezing rain. According to Lauer one can only in the tropics observe heavy rains generated due to highly vapour enriched clouds of a vertical extend of more than 2km (LAUER, 1993: 80).

2.2 Rainfall in the Tropics

In the tropics much of the rainfall is associated with convective activity. Observed rainfall rates of 100mm h^{-1} are not uncommon, and are mainly caused by a fluctuating band near the equator where strong vertical air mass motions occur (HENDERSON-SELLERS, 1996). The reason for this fluctuating band is high temperatures together with abundant

perature differences and wind turbulences. Surface characteristics can therefore trigger the process. Henderson-Sellers uses the example of the afternoon convective showers during the rainy season over the Lake Victoria coast of Uganda originating over the islands in the lake which are faster heated than the colder water in the lake and thus forming a temperature difference. (HENDERSON-SELLERS, 1996: 118)

water vapour, which enable a climate of short-lived storms from cumulus clouds.
Monsoonal circulations are featuring seasonal precipitation regimes that mainly take place over tropical Asia and are responsible for high annual rainfall totals.
Between the latitudes of 15° South and 20° North almost half of the global precipitation is observed (LAUER, 1993: 85)

2.2.1 Seasonal Variations

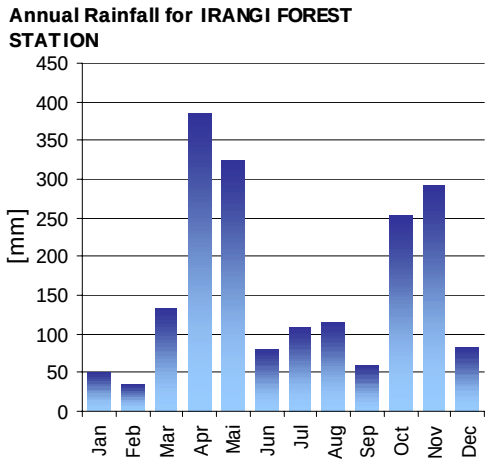


Diagram 2.1: Annual Rainfall from 1960 – 1998 for Irangi Forest Station, Kenya (Source: NRM[3] Database, B. Sturm)

The seasonality of the tropical precipitation climate is mainly a result of the seasonal movement of the ITCZ (see chapter 1.5 Climate on page 15). Near the equator the ITCZ is always fairly close so its influence is felt in most months. Thus there are no dry periods. As the ITCZ follows roughly the position of the maximum solar heating one can distinguish two maximum peaks of observed rainfall during the two solstices over the equator during April and October.
However, the detail seasonal distribution may be influenced by the exact position of the ITCZ as well as by local topography. As described in chapter 1.5.3 (The Rainfall Pattern) on page 18, the Upper Ewaso Ng'iro basin is not only under the influence of the ITCZ, but also by local topographic features such as the Lake Victoria, orographic obstacles and other seasonal variations like the South-east Asian monsoon which affects the seasonal variations of the received rainfall.

2.3 Gauging Rainfall

Rainfall is gauged in units of millimetres. The measured value represents the depth of a water layer covering the ground with the absence of evaporation, vertical infiltration or lateral runoff (GOLUBEV, 1986: 61). A measured rainfall of 1mm equals to 1 litre of water covering an area of 1m² (CAPPEL, 1988: 7).
Rainfall is measured by means of a rain gauge, which is essentially a bucket with a horizontal orifice of known diameter exposed just above the surface of the Earth.

The gauges mainly used in the Upper Ewaso Ng'iro Basin are denoted as 'standard type' within the NRM[3] project. This gauge is produced by Casella London Ltd., and is named as "Splayed-base Rain Gauge, 127mm aperture, British Meteorological Office Pattern MkII". The mechanical specifications of the gauge are described in the Table 2.1.

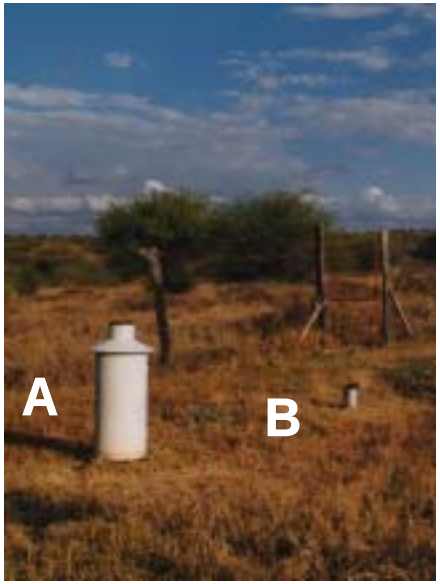


Figure 2.1: Automatic rain gauge and manual rain gauge: A: Hellmann automatic gauge, B: Cassella Standard MkII manual gauge. Picture by B. Sturm, taken at Mukogodo NRM[3] meteorological station.

Casella W5050, British Meteorological Office Pattern MkII	
Diameter of aperture	127 mm
Area of aperture	200 cm^2
Dimensions (height x diameter)	490 mm x 216 mm
Weight	2.9kg

Table 2.1: Mechanical Specifications of the Casella Standard Type Rain Gauge widely used within the study area. (Source: Casella Catalogue, 1983)

Hellmann automatic Rain Gauge (1509)	
Diameter of aperture	127 mm
Area of aperture	200 cm^2
Dimensions (height x diameter)	1200 mm x 420 mm
Weight	Approx. 20 kg
Continuous recording time	31 days

Table 2.2: Mechanical Specifications of the Hellmann (1509) Type Rain Gauge. (Source: Lambrecht, Catalogue 2001, www.lambrecht.net)

There are various different types of rain gauges. The most simple collects the rain that is caught in the bucket of the gauge in a protected measuring cylinder inside of the gauge. The depth of water is collected at the end of a period and then manually noted. For the automatic continuous registration a chart mounted on a revolving drum records the amount of occurred rainfall as a sum curve. This enables not only the exact and continuous registration of the amount of fallen rain but also its intensity and duration. The figure 2.1 shows an automatic and a manual rain gauge typically used in the study area.
Another method of measuring rainfall is through the use of radar. At a certain wavelength raindrops are reflecting the emitted electromagnetic pulse. From the time delay

between the time of emission and received reflection the exact location, speed of motion and direction of the raindrops can be computed.
However, such weather radar techniques need to be calibrated, ideally with a high density surface based rain gauge network (ARKIN et. al, 1985: 11)
Another method of measuring rainfall is by means of space-based systems, such as satellites. This technique can be subdivided into two groups: *active precipitation estimating techniques,* and *passive techniques*.
The active technique uses, similar to the surface based rainfall radar method, emitted radiation to sense reflected energy. A passive system will measure terrestrial and atmospheric radiation in spectral bands in which rain drops and droplets have observable effects (ARKIN et. al, 1985: 11). Despite the fact that there was and is a substantial progress in remote sensing technology, it should be noted that there would still be a need for adequate surface-based ground-truth data.

2.3.1 Rain Days

When analysing a rainfall regime, one could also refer to the number of rain days per year in order to describe the local climate. A rain day is defined as a 24-hour period, usually starting a 0900Z, during which 0.2mm or more of precipitation falls (HENDERSON-SELLERS, 1996: 141). The assigned date of the 24-hour period is defined as the date previous to the actual reading. There are four official times when synoptic data is being observed: 0000, 0600, 1200, and 1800 UTC hours (CAPPEL, 1986: 3). If a rain gauge station records at all four times the amount of rain, the rainfall within 24 hours (rain day) is calculated by the sum of all 4 values.
Because it is desired to get locally comparable observations, measurements should take place at the time of the same height of the sun above the horizon. According to Lauer this is at the climate observation times 0700, 1400, 2100 hours mean local time[6] (LAUER, 1996: 82). However, as these times may vary depending on the geographical longitude most nations have defined their own measurement standards in order to maintain uniformity within their gauging networks (HENDERSON-SELLERS, 1996: 135).

2.3.2 Sources of Error

This is a very crucial topic and there can be found various studies in the literature dealing with the description and compensation of errors involved in rain gauging techniques (SEVRUK et al, 1985).
There are always two types of errors:

[6] For a simple gauging station, Lauer suggests measurements at 07.30 hours mean local time. (LAUER, 1993: 82)

- Random errors
- Systematic errors

A random error would result in additional noise to the measurement (Figure 2.2). Random errors occur unsystematically and are difficult to identify due to the scattered results (HERWEG: 1997, 2). Random errors are therefore responsible for a wider variance of a sample of measured values.
A systematic error on the other hand, will always falsify measured values in the same direction and can therefore be expressed as a bias producing a wrongly derived mean value of a measurement series (Figure 2.3).

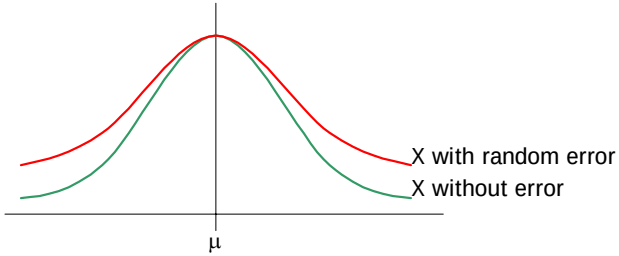


Figure 2.2: Random error forcing a sample distribution (X) to contain a higher noise level (variance). Source: HERWEG 1997

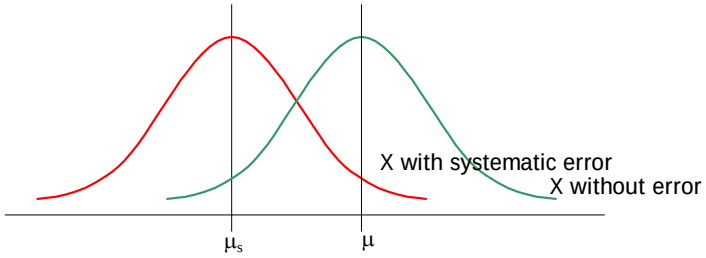


Figure 2.3: Systematic error forcing the mean value of a sample distribution (X) to a deviating direction. Source: HERWEG 1997

Both type of errors lead generally to a constant underestimation of the observed rainfall (RUDOLF, 1995: 21). This is because it is more likely to measure less rainfall due to involved errors than more. The two error types can be found on all levels of an entire measurement system:

- directly at the instrument (due to construction properties of the applied model)
- at the set-up of the gauge (through topographical features, human impacts, and the overall spatial orientation of the gauge)
- at the data collecting process (wrong manually noted values, defective recording devices)
- at the gauging network itself (unfavourable gauge distribution, inhomogeneous observing techniques)

According to Sevruk the systematic error is *"[...] believed to be the most important source of error."* (SEVRUK, 1985: 13), and its impact to rain gauge based measurements can be found in the range between 5-15%.
The main components of the systematic error are listed in Table 2.3:

Component	Magnitude	Instrumental factors
Loss due to wind-field deformation above the gauge orifice	2-10%[7]	The shape, orifice area and depth of both the gauge rim and collector
Losses from wetting on internal walls of the collector and in the container when it is emptied	2-10%	The same as above and, in addition, the material, colour and age of both the gauge collector and container.
Loss due to evaporation from the container	0-4%	The orifice area and the isolation of the container, the colour and in some cases the age of the type of the funnel (rigid or removable)
Splash-out and splash-in	1-2%	The shape and depth of the gauge collector and the kind of gauge installation

Table 2.3: Main components of the systematic error in precipitation measurement (After: SEVRUK, 1985: 14)

It is surprising that systematic errors are not taken into account by most meteorological services, and this is even more astonishing when rainfall observations are compared with the perfection level at which other meteorological observations are carried out (SEVRUK, 1985: 13).
However, simple instructions, such as how to place the gauge with respect to obstacles at the gauge site, help to reduce the loss due to wind effects to some extent, but are not sufficient enough to prevent wetting or evaporation losses.
Richter has therefore published some rules of thumb to correct at least the worst losses (RICHTER, 1995). He suggests to use the mean shielding of the surrounding horizontal features around the position of a rainfall gauge. For a quick assessment and classification of the quality of the gauge shielding in the field, the following rules can be used[8]:

Category	Distance to the Obstacle x Height	Magnitude of error on daily observations
1 (Open space)	>10 – 20x	7.7%
2 (Slightly protected)	5-10x	6.2%
3 (Better protected)	5x	4.5%
4 (Strong protection)	2-5x	2.4%

Table 2.4: Mean error induced by wind (After RICHTER, 1995: 40)

However it is difficult to use the correction values listed in Table 2.4 as the terrain itself can produce local turbulences, which render a windshield of a gauge useless. This is crucial for extremely wind-

[7] Other studies suggest a much bigger influence of the loss due to wind-field deformation. Rudolf found a maximum error of 20% for a precipitation measurement with a Hellmann standard gauge (RUDOLF, 1995: 23)

[8] Richter suggests that a simple visual assessment of the surrounding obstacles is sufficient enough in order

exposed locations such as windward or leeward orientations of the gauge.
Casella Ltd., the manufacturer of most of the standard rain gauges used in the study area recommends that the site around the gauge should be "[...] *sheltered from the full force of the wind, but no object should be nearer to the gauge than a distance four times its height. Where there is no natural shelter, the gauge should be placed inside a ring of turf about 3m (10ft) inside diameter, 300mm (1ft) high, vertical inside and gently sloping outside.*" (Casella Catalogue, 1983: 1).
None of the visited gauges was specially protected against wind nor were the operators of these gauges aware of the fact, that wind could produce considerable errors onto the observations.
However, it is noteworthy that wind induced systematic errors are only of concern when the gauge is located too high above the ground (1-2m), and when the raindrops are of small diameters (<3mm). Otherwise the loss due to turbulence is according to Kreuels and Breuer "*neglectibly small.*" (KREUELS and BREUER, 1985: 106). Therefore, it can be concluded that due to the typical bigger drop sizes of tropical rainfall the induced error is of not such big influence as for rain gauges located in mid-latitudes.

Systematic errors caused by the relocation of the gauge

Another source for systematic errors is buried in the history of a rain gauge. Most studies only seem to take the present situation of gauges into account, as this may be appropriate for modern western rain gauge networks where one is concerned to provide a constant and homogenous measurement environment as possible, this is not the case for networks located in less developed countries where governmental maintained networks are also relying on privately operated gauges originating from colonial times.
Hanssen-Bauer and Førland state that even for a set of Norwegian precipitation series 50% of the inhomogeneities are caused by a relocation of the precipitation gauge in the past, and most of the inhomogeneities led to increased gauge catch (HANSSEN-BAUER, FØRLAND, 1994: 1001)[9]

to determine the induced systematic error (RICHTER, 1995: 31).

[9] According to Hanssen-Bauer and Førland the better shielded rain gauges could also lead to the observed increased gauge catch, but that it is difficult to exactly identify the cause for the measured trends: *"The tendency for artificially increasing trends in unadjusted precipitation series may be global. Improvements of instrumentation, which in most cases lead to increased gauge catch, have taken place all over the world."* (HANSSEN-BAUER, FØRLAND, 1994: 1012)

For the presented study no corrections were applied, although a quality assessment based on the following points was made for each rainfall gauging station:

- Wind protection of the gauge
- General orientation of the terrain (slope, windward, leeward)
- Material and condition of the gauge itself[10]
- Homogeneity of the recording process[11]

As stated above no adjustments to the gauge, or the observed data were made for the presented study. However, the data originating from the unadjusted gauges was tested according to the Standard Normal Homogeneity Test (SNHT) developed by Alexandersson (ALEXANDERSSON, 1986) in order to identify inhomogeneous datasets. The results of the SNHT test were compared with eventually known metadata from the gauges history such as the relocation or the replacement of the gauge (see also chapter 2.6 Quality Control of Raw Data).

Potential Evaporation Loss Test in Nanyuki 28.10.1998 – 5.11.1998

In order to quantify the magnitude of the induced error originating from evaporation and the loss due to the manual readings[12] of the gauge cylinder evaporation tests in Nanyuki, Kenya were performed. These tests took place in Nanyuki in the backyard of the NRM[3] office building. The make of the used gauge was a Casella W5050 standard gauge MkII. To produce worst-case results the gauge was placed directly at the brick / concrete wall of the building where high temperatures were measured, and high evaporation losses could be expected. The gauge was sheltered against rainfall, and initially filled with 10mm water. Every 24 hours the gauge was emptied and the contents of the container were measured. The table (Table 2.5) lists the results.

[10] Condition of gauge includes any leaks, condition of funnel, and container.
[11] This includes readings at the official meteorological observation times and proper recordings. Furthermore metadata such as any known relocation of the gauge in the past was used.

[12] During reading of the gauge the water collected by the gauge cylinder has to be filled into a measurement glass cylinder. This process involves always wetting losses.

Start	Stop	Difference	Remarks
0900 28.10.98	0830 29.10.98	0 mm (0%)	Rain at 1515 hrs
0830 29.10.98	0900 30.10.98	+1.3mm	Rain!
0900 30.10.98	-	-	Cancelled due to rain
0900 03.11.98	0830 04.11.98	-0.2mm (2%)	Cloudy, few rain
0830 04.11.98	1100 05.11.98	-0.2mm (2%)	Cloudy, few rain

Table 2.5: Measured evaporation losses from the evaporation loss experiment from 28.10.98 – 05.11.98 (Source: B. Sturm)

Although the conducted tests are far from being statistical solid, with some care we could conclude that the systematic error due to evaporation loss is considerable small and can be neglected (see also Table 2.3, page 31).

We have seen, that a lot of environmental parameters influence any sort of ground-truth based rainfall measurement, and it is difficult to measure the real amount of rainfall that actually arrives at the Earth's surface. But we have also seen, that we may reduce the errors involved in rainfall observation techniques by the application of considerable simple counter measures, such as shielding the gauge properly against wind turbulences.

2.4 Network Density

On the level of an entire rain gauge network the distribution and density of the rain gauging stations can have a strong influence on the quality of an areal rainfall analysis.

Because rainfall measurements are point measures that assume to represent the rainfall regime around the rain gauge, it is obvious that the farer away from a gauge the more unreliable this assumption becomes.

The correlation of rainfall amounts for two rain gauges separated by various distances gives insight into what the sampling problems will be to determine area mean precipitation with a given accuracy. For a station in Valdai (former USSR) the correlation dropped to 0.6 at a site 50km away, and a rain gauge at one site can explain only 36% of the rainfall variance at a site 50km away. It is evident to note that the sampling intervals and the spatial resolution of the network plays an important role for the quality and reliability of the measured rainfall values. For the above example a sampling interval of 24 hours was chosen, when reducing the sampling interval to monthly precipitation, the correlation drops to 0.6 at a station 170km away, and for seasonal precipitation, more than 300km away (WMO, 1985: 4).

Therefore, on daily precipitation estimates errors are substantially higher than decadal or monthly values (see Figure 2.4).

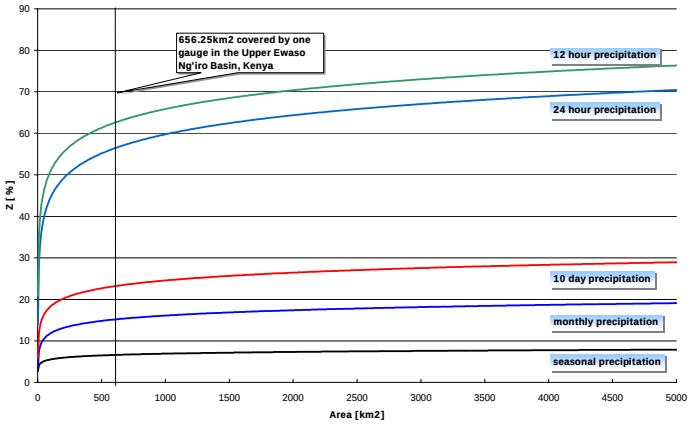


Figure 2.4: Standard Deviation of the error of determining mean precipitation amount over an area on the basis of data from stations situated at its centre. The data was measured at Valdai (USSR), summer period. (Taken from WMO, 1985: 16)

2.4.1 WMO-Recommendations

The World Meteorological Organisation (WMO) has developed some recommendations for the time and scale requirements of precipitation estimations. The idealised requirements are as follows (after WMO, 1985: 17):

- Synoptic Scale: 6-hourly data; 160km spacing
- Meso Scale: 3-hourly data; 80km spacing
- Small Scale: 1-hourly data; 40km spacing
- Cumulus Scale: 15-minute data; 10km spacing

However, these recommendations depend greatly upon the application. The WMO has also recommended the minimum station density for hydrological observations depending on topography features:

Region Type	Range of Norms (km^2 per stn)	Range of Norms tolerated
Flat	600-900	900-3000
Mountains / island	100-250	250-1000
Arid and polar	1500 – 10000	

Table 2.6: Recommended minimum station densities for hydrology (from WMO, 1985: 17)

The gauge density for the study area for the minimum stations used for this work is:

Area Upper Ewaso Ng'iro: 175km x 150km = 26'250km^2
The minimum density for 40 stations[13] is:

$$d_{40}=656.25km^2/Station$$

The density of one gauge representing an area of 656.25km^2 is low, but still in the range of norms tolerated for mountainous topography, and in the range of norms for arid climatic regions (1500km^2-10000km^2).
These recommendations are rather crude and another approach by Stephenson (SUMMER,

[13] Number of stations used for the interpolated time period in the presented paper from 1.1.67-30.9.98.

1988: 305) uses the mean and standard deviation of monthly rainfall to compute the coefficient of variation for a range of catchments area sizes, together with gauge density, in order to determine the minimum number of gauges that represent a certain area. For an area of 50'000km² Stephenson suggests the minimum number of 25 gauges (SUMMER, 1988). Compared with the 26'250km² area of the Upper Ewaso Ng'iro Basin and the 40 stations, used for the 30-year analysis, we can conclude that the used gauges form a sufficient solid statistical ground for spatial rainfall analysis. However, one has to consider that the rain gauges operating within the network are not evenly spaced. The network has its highest density in the higher altitudes around the mountains, and a lower density in the semi-arid lowlands.

2.5 Precipitation Database

Because of the three main pillars of the NRM³ project (Monitoring, modelling and management), a natural resource database was established.

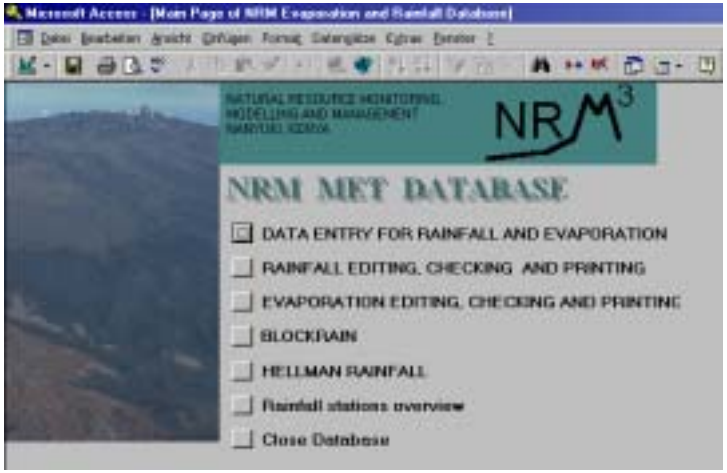


Figure 2.5: Main menu of the NRM³ database

This database stores soil, water and vegetation data sampled at different measurements sites in the study area. Because data management forms a major issue the main database is kept and maintained at NRM³ headquarters in Nanyuki, Kenya. The relational database (MS Access) serves for the following purposes (after GICHUKI et al., 1988: 12):

- Provide access to the data for researchers, government, development agencies and users
- Ensure long-term data quality and consistency of data collection and storage
- Serve as a monitoring tool and an example for other projects in the same context

The database enables users to query and analyse the major environmental parameters. For this purpose a set of reports and tools was developed and integrated (STURM, 1997).
The rainfall database keeps the observed daily rainfall of 116 rain gauges, and the first measured daily rainfall event was recorded at the 1.1.1934[14]. At present the rainfall database contains almost 1'000'000 samples. Because the stored rainfall values are collected from different sources (governmental, NGO or privately operated gauges), it is clear that the database is not updated immediately with the most recent data. Usually every two years a data collection campaign is conducted during which each rain gauge is visited and data even from the far remotest gauge is being collected (usually by a team of 1 research assistant and 1 driver). The author participated at the last campaign in late 1998. During this campaign additional information for each gauge was collected[15], which helped to assess the quality of the gauges, and to identify unreliable gauges.

[14] First monthly observations start as early as 1925!

[15] This included the precise geographical position (GPS), condition of the gauge, sketches and images from the gauge's environment, interviews with observers, reading times and meta data of the gauge's history.

Q	Station	Start	Longitude	Latitude	mAMSL	UTM Zone	Spheroid	Manufacturer	Material	Unit
1	ARCHERS POST	1957	351624.283	70807.0123	860	37N	ARC1960	Casella	Glass	mm
1	CHUKA FOREST STN	1988	344923.385	9965172.17	1474	37M	ARC1960	Casella	Glass	mm
1	COLCHECCIO	1963	255409.16	68804.3074	1831	37N	ARC1960	Unknown	Plastic	mm
1	EMBORI FARM	1967	316071.142	7766.9033	2691	37N	ARC1960	Casella	Glass	mm
1	GATHIURU FOREST STN	1959	290496.978	9989256.09	2330	37M	ARC1960	Casella	Glass	mm
1	HOMBE FOREST STN	1957	290267.591	9961541.78	1991	37M	ARC1960	Casella	Glass	mm
1	KALALU (NRM)	1986	295548.955	9332.62183	2080	37N	ARC1960	Casella	Glass	mm
1	LAMURIA MET STN	1963	263186.069	9986005.94	1876	37M	ARC1960	Casella	Glass	mm
1	MARIENE CRS	1989	349303.747	9999065.13	1670	37M	ARC1960	Casella	Glass	mm
1	MOGWONI RANCH	1977	275763.207	25543.0605	1760	37N	ARC1960	Casella	Glass	inch
1	NARO MORU MOORLAND (R2)		304420.031	9981854.2	3771	37M	ARC1960	Belfort	Glas	mm

Table 2.7: Output from the rainfall database: the first top quality rain gauges with their geographical position

2.6 Quality Control of Raw Data

Quality control of the collected precipitation data is involved at different stages: at a first stage during the data collection campaign. Because the gauge operators note the measured rainfall manually on record sheets, each value also has to be copied manually by the NRM[3] researchers. This is a very time consuming process, and can also form another source of errors, but any doubtful values may be already identified during the manual copy process[16].

Later the collected rainfall recording sheets are entered in the rainfall database. The software automatically performs various checks during this process, such as inch/mm conversion or gap identification.

After the 'punch in', quality control reports can be generated which check for missing months, gaps, and unusual high values.

Before the data is being released as 'raw' a sound data quality control takes place: the entered data is cross-checked against the paper recording sheets by another person who was not involved in the data collection process. This ensures a high degree of reliably stored data in the database.

[16] As the observers at governmental operated stations are not always present, it is possible that during bank holidays the noted rainfall represents the mean value for each day during the holiday period or only the sum is noted at the end of the holiday period.

3 Precipitation: Analysing Raw Data

3 Analysing Raw Data

In this chapter the raw precipitation data stored in the NRM³ database will be a major issue. Based on this raw data all further analysis and interpolations were conducted. Due to the incomplete nature of the raw data set a reliable set of rain gauges has to be selected for the later spatial interpolation process. This set must not only provide a long as possible time span, but also represent a statistical robust spatial distribution. Another part of this chapter will then deal with the topic of gaps and missing data in those time-series.

3.1 Presentation of Available Data

As already discussed in chapter 2.5 on page 36, the NRM³ database keeps all the observed rainfall values of the NRM³ monitoring network. For this study the following summary of the database rainfall values formed the base for the later analysis and interpolation:

- 116 rainfall gauges[17]
- First recordings start at 1.1.1934
- Geographical position of all gauges
- 923'526 daily observed rainfall values
- The median of the time covered by the measured precipitation values is 20 years.

3.1.1 Metadata and Fieldtrip

From September to November 1998 a data collection campaign in the Upper Ewaso Ng'iro Basin took place. The goal of this campaign was to gather missing rainfall recordings from remote gauges in the study area. Because some of the gauges are very difficult to access it is not possible to enter the data on a daily base in the NRM³ database, so every two years data is being collected by researchers from the NRM³ project in Nanyuki.
The fieldtrip for this work was part of this data collection campaign. During this fieldtrip the following information was collected in order to gain a better understanding of the conditions and implications of the rainfall observation in the Upper Ewaso Ng'iro Basin:

- Exact geographical position and altitude was measured using a standard GPS receiver[18]

[17] See Appendix A for a full list of all rain gauges stored in the database

[18] Type used for the positioning: GARMIN GPS12 receiver. A position was determined by means of RMS average based on at least 60 sampled positions and with a minimum of 5 satellites available.

- Drawings of the set-up and the vicinity of the gauges were made
- Assessments of the grade of protection of the gauges were made
- Where it was possible interviews covering several issues (Reading times, inconsistencies, abnormal wet/dry years) with the responsible observers were conducted.
- For later reference pictures from at least two different locations of the gauge and its surroundings were taken.
- If possible, the history of the gauge, such as any known dislocation or damage of the gauge in the past, was recorded.

When visiting gauges that were maintained by farmers, it was possible to conduct longer and deeper interviews with the responsible farmers. During these interviews the farmers were asked whether they could remember especially dry years (droughts) or wet years (floods) and if they believe that the rainfall pattern, such as the onset or the end of the rainy seasons, has changed. For later reference most of the answers stemming from these interviews and the environmental details of the gauges were also stored in the NRM[3] database, and form additional data which can be used as metadata when exploring the raw data of the rainfall observations.

Refer to chapter 7.3 (Farmers Perception and Interviews) on page 139 for the results of the above-mentioned interviews.

3.1.2 Time Series

After entering the daily observed rainfall values a first analysis will always be a time series analysis.
For a detailed overview of all the stored precipitation values in the NRM3 database, refer to appendix B-Tables (chapter 12, page 169) where one can easily identify the starting point of the recordings and the eventually occurrence of any missing values.

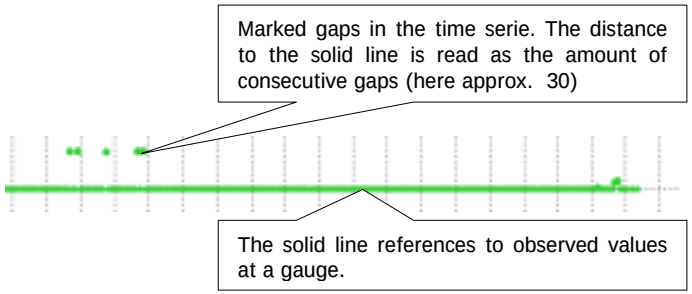


Diagram 3.1: Detail from the overview of all available data.

Table 3.1 lists the distribution of the length of the observed time spans. Half of the gauging stations are covering 20 years or more of continuous rainfall recordings. Only 12.7% of all gauges in the study area cover more than 42 years.

Coverage in Years	Percentage of gauges
64	0.0%
42	12.7%
40	16.3%
31	37.9%
20	50.0%
11	60.4%
9	76.6%
2	100.0%

Table 3.1: Distribution of the length of the recorded time span in percent.

If we analyse stations towards the occurrence of gaps we see that out of 112 stations 84 stations do show missing values in their recordings.
Another important issue is the time when regular rainfall observations started. Generally most of the farmers gauges started recording rainfall long before Kenya's independence in 1963, these gauges show the longest and usually thoroughly documented rainfall recordings in the entire area. However, as these gauges tend to be clustered in a small area located at the (fertile) northern slopes of the Mt. Kenya and towards the big cattle ranches at the southern borders of the drylands, they form some sort of an isolated group of long rainfall time series. Another well-documented group of gauges are the gauges maintained by the Kenyan Meteorological Department and the ones at the Forest Stations. During the history of the NRM³ project many more observation stations were introduced in order to study

the rainfall pattern at other areas. As these gauges were only installed at the end of the eighties or at the beginning of the nineties in the last century, they do not offer a long enough time period for a long-term trend analysis (see Table 3.2).

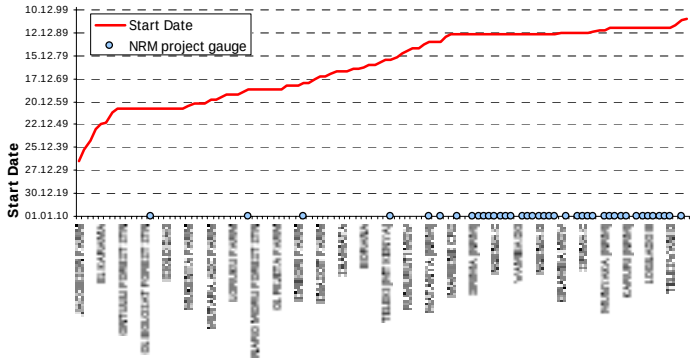


Table 3.2: Start date of recording from all gauges (red line) compared to NRM3 project maintained stations (blue circles).

Therefore, it is obvious that the rainfall recordings kept in the NRM[3] database are far removed from representing high quality precipitation data. Nevertheless, they are the best ground based precipitation data we can get from the Upper Ewaso Ng'iro Basin[19].

[19] It is even doubtful whether there are better and more complete time series available for the rest of Kenya.

3.1.3 Spatial Distribution

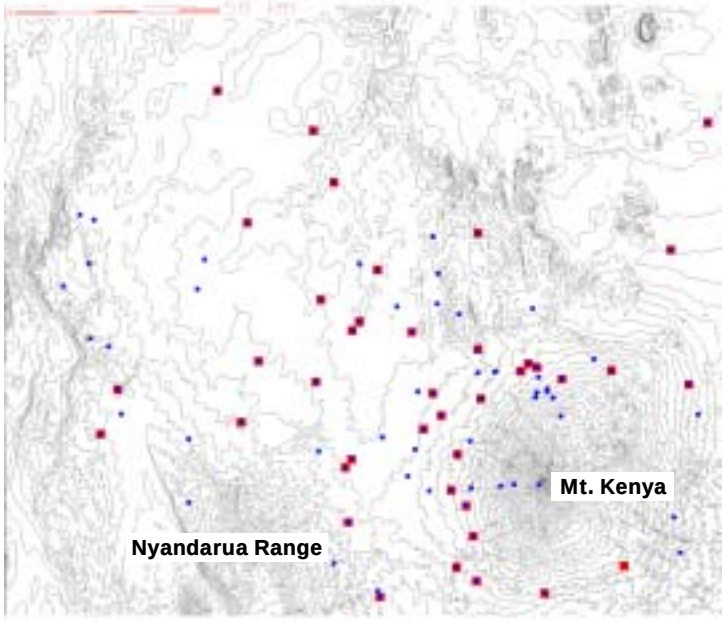


Figure 3.1: Spatial distribution of all available rainfall gauges in the NRM3 database. Blue squares indicate all gauges; red squares indicate the gauges used for the spatial interpolation from 1.1.1967 – 31.12.1997 in this work. (Source: B. Sturm and NRM3 database)

Because precipitation shows a considerable spatial variation (SUMMER, 1988) and any precipitation map based on gauge observations is produced by referencing in one or the other way to point data, we have to be aware of the difficulties arising from the spatial distribution of the gauges.

As already concluded in chapter 2.4 (Network Density) on page 34, each gauge in the Upper Ewaso Ng'iro Basin represents enough area to get a reliable statement about the spatial rainfall pattern.
However, it is not only the ideal spatial representation of one gauge which is important for a spatial interpolation, but also the overall distribution pattern of the network itself.
Figure 3.1 shows the geographical location of the gauges referenced in the NRM[3] database, as mentioned in chapter 3.1.1 a clustering of gauges can be found on the northern slopes of the Mt. Kenya, and at the southern edges of the drylands in the center of the study area. On the other hand, mountaineous features, which are part of a different rainfall regime than the lower parts of the basin, are hardly populated by gauges, and their rainfall pattern might therefore erroneously estimated by the use of the other gauges.
In order to overcome such a problem additional gauges, positioned on various altitude levels, were introduced by the NRM[3] project on the Western slopes of the Mt. Kenya in order to form a linear transect ranging from altitudes of 2195mAMSL to 4262mAMSL. As these gauges[20] were only installed recently they cannot be used for any long-term analysis but for short-term analysis, such as 10-year analysis, they already offer enough sufficient data.

3.2 Selecting Sampling Locations and Time Span

When analysing precipitation data it is desired to cover a time period as long as possible in order to identify any eventually underlying trend to get reliable mean values from the data. Furthermore, the used data should origin from a network with sufficient gauge density to avoid too large spatial estimation errors. As shown in the previous chapters 3.1.2 and 3.1.3 (Spatial Distribution) a trade off between these two demands has to be found as the network in the study area offers not ideal conditions.
As all previous studies on the subject of areal rainfall in the study area were conducted using a maximum length of time period of 23 years (LINIGER, 1998), it was desired to base the estimations on a longer time period for a sounder analysis.
In chapter 3.1.2, we saw that only 12.7% of all gauges cover 42 years of complete rainfall observations. These 12.7% represent approximately 14 gauges, which are insufficient for spatial estimations. But for a

[20] The gauges are: TELEKI VALLEY (MT KENYA) [4262mAMSL], NARO MORU MOORLAND (R2) [3771mAMSL], NARO MORU MET STATION [3050mAMSL], NARO MORU GATE STATION [2420mAMSL], NARO MORU FOREST GATE POST [2195mAMSL]

coverage of 31 years approximately 43 gauges could be used.

Therefore 40 stations were chosen which covered 31 years (from 1967-1998) of rainfall recordings for a long-term estimate. In order to estimate precipitation on another set of gauges (for quality control reasons) a second set was chosen which covered only 8 years (1990-1997). But, this second set was based on data provided by 63[21] rain gauges, which enabled a more robust spatial estimation.

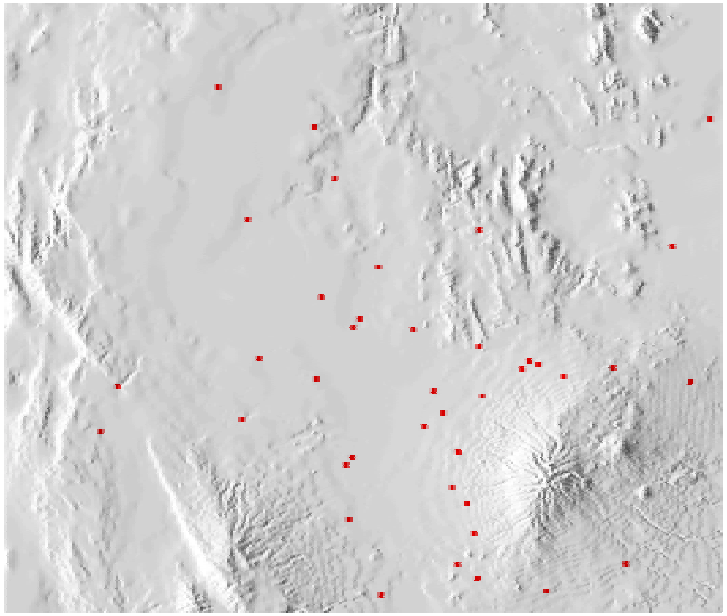


Figure 3.2: Geographical distribution of all gauges used for the 31 year analysis from 1967-1997 (Source: B. Sturm)

[21] From these 63 gauges only 61 were actually used for the later interpolation, as 2 gauges were located outside the study area.

3

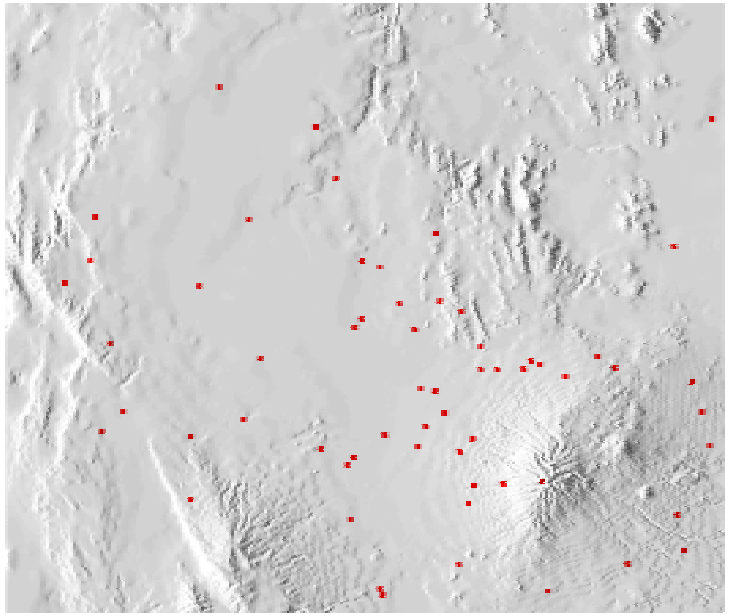


Figure 3.3: Geographical distribution of all gauges used for the 8 year analysis from 1990-1997 (Source: B. Sturm)

3.2.1 Sampling Resolution

In chapter 2.4 (Network Density) on page 34, the problem of the sampling resolution was discussed by comparing the correlation coefficients between two gauges at different sampling resolutions. Although the raw data stored in the NRM³ database is based on daily observations any spatial estimation based on these daily measurements would lead to intolerable estimation errors (see also Table 2.6 on page 35).

One goal of this work was to provide rainfall maps offering higher time resolutions than previously drawn maps for the Upper Ewaso Ng'iro Basin[22], which provided annual and monthly resolutions. Because the rainfall regime over the study area is mainly of convective origin, and thus forced by the seasonal variability of the ITCZ (see for a detailed discussion chapter 1.5 (Climate) on page 15), which is responsible for the different rainy seasons in the area, it is of importance to recognise the onsets and ends of those seasonal changing patterns in the rainfall domain. The onset and end of the rainy periods is of considerable importance, since it determines the extent of the growing seasons. Estimates based on monthly averages are not able to distinguish the different rainy periods or rainy seasons as a monthly resolution is too coarse as Berger puts it: *"[...] the monthly rainfall figures may give an initial idea of the general pattern of rainfall, but the are often not precise enough."* (BERGER: 1989: 43). In his work "Rainfall and Agroclimatology of the Laikipia Plateau, Kenya", Berger uses a set of rules to identify onsets and ends of

[22] Peter Berger has produced maps based on mean annual calculations (BERGER, 1989). The maps composed with ClimShell by Gudrun Schwilch at the Centre for Development, University of Berne, covered mean monthly estimations (LINIGER, 1998).

rainy periods. According to these rules the onset of a rainy phase is marked as:

- The first day of a period of 5-days with consecutive rainfall of at least 20mm of rainfall followed by a period of 10-days with at least 20mm of recorded rainfall. (BERGER, 1989: 43)

According to this rule the onset of a rainy phase is defined by a window of 15 days.
The rule for the end of a rainy season is:

- The last day of a 10-day period without rainfall.

If we apply these rules to any set of consecutive rainfall observations a rainy phase would theoretically fit into a 25-day period, which could not be recognised within a monthly resolution. However, with a smaller resolution, such as a 10-day (decadal) time window, it is possible to detect the onset and the end of a rainy period with considerable more accuracy.
According to the WMO recommendations a decadal (10-day) analysis would lead to a <25% standard deviation of the error for determining the mean amount of precipitation for an area of 656.25km^2. This is the area that is covered by one gauge out of 40 stations in the Upper Ewaso Ng'iro Basin.
A daily resolution would increase the error to >55%, and on a monthly scale the error drops to approximately 15% (WMO, 1985: 16).
The decision to use a time resolution of 10-days was also supported by a study covering the mean area rainfall for the Ivory Coast, Africa. Depending on the used number of gauges for the calculation of the weekly rain the estimated error was ranging from 12% (at 39 used gauges) - 54% (at 5 used gauges) (WMO, 1985: 23).

3.3 Gaps Infilling

So far we have chosen two different sets of raw data on which further geostatistical analysis will be conducted:

- **Long-term set:** 40 rain gauges, covering a time period from 1.1.1967-31.12.1997 (31 years) with a decadal resolution.
- **Short-term set:** 63 rain gauges, covering a time period from 1.1.1990-31.12.1997 (8 years) with a decadal resolution.

As we have already stated in earlier chapters the time series contain, depending on the quality of the rainfall station, more or less 'holes' with missing values. These missing values are referred to as 'gaps'. Because any spatial estimation can only be

computed based on complete time series, these 'holes' or gaps are to be filled up with estimated rainfall values.
There are two methods to re-construct these missing data. The first may be used when there is typically little spatial variation in precipitation, and involves the calculation of the arithmetic mean of values from three or four adjacent gauges. Where we cannot assume such uniformity in precipitation measurements, a second method can be used. The ratio for a longer-time period is being calculated, between the means of adjacent gauges and the gauge for which an estimate is required. These ratios may be used as weighting factors to produce the missing values at the desired location. (SUMMER, 1988, 403). Departing from this second method a gap infilling method was applied which was already used at the Centre for Development and Environment, University of Berne, by Lindsay McMillan and Gudrun Schwilch. Chapter 6 (The Interpolation: Step-by-Step) on page 89, describes the application of this method to any series of rainfall recordings, but first we will take a brief look at the methodology of the gap elimination process.

3.3.1 Methodology

The method, which eliminates the missing values, is based on the assumption that gauges that are geographically close to each other tend to show similar rainfall characteristics. The steps that are involved are the following:

1. Identification of neighbouring stations
2. Correlation and linear regression analysis
3. Identification of the best fit station
4. Calculation of missing data based on regression

1. Identification of neighbouring stations

For each station where data is missing, surrounding stations likely to yield similar decadal characteristics were selected manually. Stations that were located in similar topographical features were therefore preferred instead of close neighbours, which were situated in entirely different topographical regions, and hence climatic regimes. As well it is important to choose stations, which show observations of a higher quality, such as a consistent and complete time series, than the station that needs the missing values to be estimated.

2. Correlation and linear regression analysis

By the means of general statistics (correlation coefficient, mean values, linear regression curve) the gauges which are assumed to yield similar characteristics are com-

pared with the gauge where the missing values were found.
General statistics are always calculated for each neighbouring gauge (denoted as X) and the gauge for which the values need to be estimated (denoted as Y). The following statistics between the two variables X and Y were computed:

Pearson Linear Correlation Coefficient
The Pearson correlation coefficient determines the linear independence between two set of variables, according to the following formula:

$$r = \frac{n\left(\sum XY\right) - \left(\sum X\right)\left(\sum Y\right)}{\sqrt{\left[n\sum X^2 - \left(\sum X\right)^2\right]\cdot\left[n\sum Y^2 - \left(\sum Y\right)^2\right]}}$$

where r is the correlation coefficient ranging from –1.0 <= r <= 1.0, a value of r=1.0 indicates a complete positive dependency between the two variables X and Y.

When using the Pearson Correlation Coefficient and the equation of the linear regression we will assume a 'Zero Rainfall Coincidence Assumption' which is applied to the curve described by the line of linear regression:

$$Y = a + bX$$

where a describes the y value when x is zero. In order to maintain the sequence of wet and dry 10-day periods the constant a is not used, and hence forcing y equal to zero when x is zero. This removes what would be an illogical assumption of a constant in the regression of rainfall data, but yields statistics that are not based on the usual assumptions (MCMILLAN, 1997).

The Cases
The number of values found for the two variables X and Y were counted. The more cases were found for X, then this gauge was assessed of having a more complete set of observation, and hence was assumed of being more reliable.

Median, Mean, Minimum and Maximum Statistics
These general statistics are used to assess the overall distribution of precipitation of the considered gauge. The values are not used for the assessment of the best fit because it is difficult and hence impossible to assess the dependency of two sites only by the use of these general statistics (SUMMER, 1988), but help to identify any found anomalies in the recordings. At this point it is important to note that the World Meteorological Organisation recommends the adop-

tion of a minimum 30-year data sequence for 'reliable' mean series. However, such a 30-year sequence could only be achieved with the long-term data in this work, and it shall be noted that the quality of the estimated gaps for any shorter period will considerably degrade.

The Skewness

The skewness of a distribution indicates whether the frequency of the distribution tends to have more values above the mean value (positively skewed) or below the mean value (negatively skewed). Precipitation data are rarely normally distributed. Two rules of thumb concerning the statistical spread exist for precipitation data. The first is that the shorter the duration under consideration, the less likely it is that the resulting distribution will be normal. The second rule is that the drier the site being considered, the more variable is the precipitation from year to year or month to month (SUMMER, 1988). For the assessment of the best fit a low skewness was therefore preferred.

The Coefficient of Determination (R^2)

The R^2 was used to assess the explained variance between X and Y. At an R^2 of 1.0 all points described through X and Y are located exactly on the line of regression defined by

$$Y = a + bX$$

Therefore a complete dependence between X and Y can be assumed.
The square root of the R^2 equals to the Pearson Correlation Coefficient.

T-test

In order to test whether the two variables are dependent, a two-sided t-test was used. The hypotheses to test the independence of the two variables are as follows:

$$H_0 : \rho = 0$$
$$H_A : \rho \neq 0$$

Where ρ is the Correlation coefficient. H_0 assumes that the two variables are completely independent, where the alternative hypothesis H_A assumes that the variables are dependent ($\rho \neq 0$).
Under the condition of H_0 the test value t is t-distributed with a degree of freedom of $df=n-2$ where n is the number of cases found in the two variables (BAHRENBERG, 1990: 155).
The two-sided critical t-test level is therefore defined by $t_{n-2,\alpha/2}$ and checks for

$$\hat{t} = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}} > t_{n-2,\alpha/2}$$

Where r is the correlation coefficient, and n is the number of cases.
A high t-test value above the mentioned critical t-level was used to assume, with a 95% probability, that the two variables are dependent.

3. Identification of the best fit station

With the above described statistical indicators the station which was used for the regression was determined through a weighting factor.
This weighting factor was introduced in order to weight each statistical indicator independently. The site, which got the highest sum of all weights, was considered being statistically closest and therefore representing the most dependent gauge with the Y-gauge (the gauge with gaps). The weights were distributed according to these rules:

Statistical Indicator	Weight
Pearson Correlation Coefficient (high, close to 1, values >0.5 were highlighted)	+2
t-test (high, above limit)	+2
Coefficient of Determination (high, close to 1)	+2
Cases (the most)	+1
Skewness (low)	+1
Maximum of points	**8**

Table 3.3: Weightings for each statistical indicator to find the best fit

Because the best-fit regression station was determined automatically by the use of the above described weighting algorithm, it shall be noted that this method can lead to erroneously selected sites. For instance could a high correlation coefficient coincide with a low number of cases, and hence show highly scattered pairs of (X, Y) points (for a discussion see Bahrenberg page 154 (BAHRENBERG, 1990)). To minimise the risk of a wrong selection scatterplots for each pair of stations was used.

4. Calculation of missing data

The best fit was used to compute the line of regression for the missing decade value. The final calculation of the missing data was completed within the EXCEL spreadsheet (see chapter 6.3.4 (Step 4: The Regression Analysis) on page 102 for a detailed discussion).
The regression line was calculated only for decades that are missing from a station. Thus for example if a whole year is missing, then the regression exercise has to be carried out for each of the 36 decades.

The method described in this chapter proofed to provide reliable results for the estimated gaps, however it shall be noted that this method cannot be used to predict rainfall for certain sites. As McMillan puts it "*Rather, the estimated data is an estimate*

of the likely rainfall for that particular month[23]*. An example of this is the zero occurrence of rainfall. For the majority of occasions, zero monthly rainfall may be coincident at a selected pair of stations, say 16 Februarys out of 22, however for the remaining 6 Februarys rain occurred at one station, but not at the other. The regression always assumes the coincidence of zero rainfall (the most frequent scenario), but in reality in any one month rain may have occurred at one station and not the other."* (MCMILLAN, 1997: 3)

In the next chapter we will have a look at the quality of the interpolated values.

3.3.2 Quality of Interpolated Gaps

For the assessment of the grade of quality of the described method a worst-case scenario was applied to a complete set of decadal rainfall recordings. For the period of 3 years from 1.1.1985-31.12.1987 all observed rainfall values were deleted and then re-constructed using the 3 neighbouring stations. The linear regression was calculated without manual consideration of scatterplots and the interpolation software was run in the full automatic mode. After estimating the values of the previously eliminated years, theses values were compared with the original values by means of a linear regression analysis.

The station for which the values were eliminated was OL DONYO FARM, located at the northern slopes of the Mt. Kenya (see Figure 3.4 for the location of the gauges).

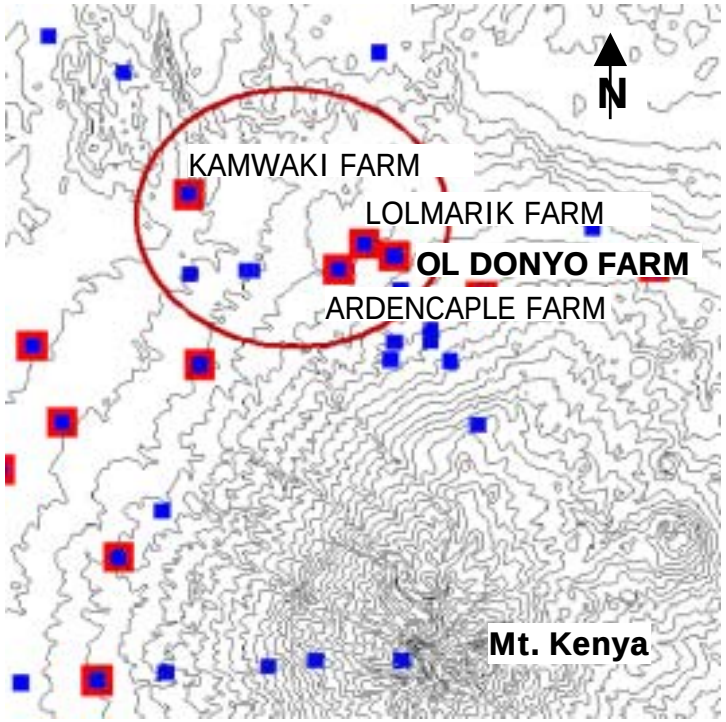


Figure 3.4: Geographical location of the 4 gauges used for the gap infill test

The linear regression and the scatterplots between the computed and observed values is shown in Diagram 3.2 and the resulting lin-

[23] McMillan was estimating monthly values with this method.

ear regression resulted in a line described through:

$$y = 0.9608x + 0.3141$$

$$R^2 = 0.8128$$

An explained variance of 81.2% allows us to conclude that the estimated values represent the original values with sufficient quality. The median of the difference between the estimated values and the observed values is 0, which supports the robustness of the applied interpolation method. The scatterplot diagram shows no remarkable outliers, and we can therefore assume that the interpolation method which uses best fit linear regressions to estimate missing values can be used as a general method to interpolate even longer series of missing values.

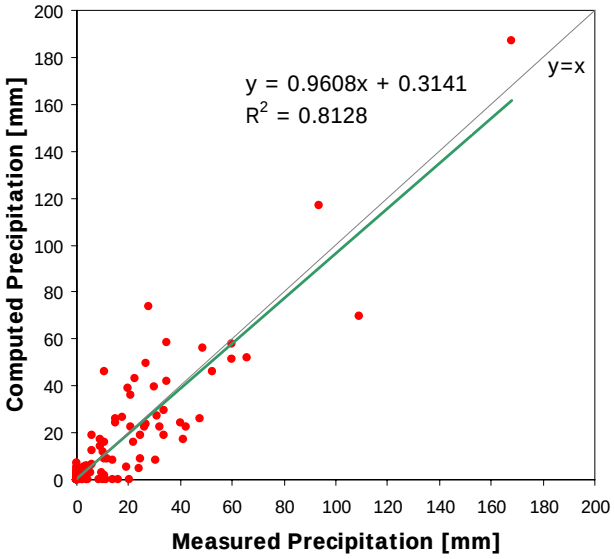


Diagram 3.2: Scatterplots and linear regression for the estimated and observed values for a 3-year period of the OL DONYO FARM gauge.

4 Spatial Interpolation: A Model Overview: Theory and Methods

4 Spatial Interpolation of Rainfall: Methods and Theory

This chapter deals with the theories that are involved within the realm of spatial rainfall estimation. It also forms the basement and starting point for the applied interpolation method used for this work introduced in the chapters 5 and 6.
At the beginning we will introduce the different estimation techniques, starting from the most basic ones and then moving on to more sophisticated geostatistical models developed solely for the purpose of spatial point estimation.

4.1 Theoretical Overview

4.1.1 Deterministic Models vs. Probabilistic Models

As rainfall samples (or any other point based sampling method such as temperature, chemical deposits or soil samples) reflect only point measurements we can not directly use the sampled data to draw conclusions about the rainfall at any other location than at the rain gauge. To overcome this problem we need to find a method, which will enable us to estimate rainfall at any desired point in a given area. There have been various interpolation techniques suggested in the literature, and all those methods can be grouped into two different main categories: deterministic models and probabilistic models (ISAAKS, SRIVASTAVA, 1989). In a deterministic model there is sufficient knowledge about the phenomenon to allow a deterministic description of it. But such a sound knowledge is only available for a limited number of physical processes. And as Isaaks and Srivastava puts it: *"For the vast majority of earth science data sets, we are forced to admit that there is some uncertainty about how the phenomenon behaves between the sample locations."* (ISAAKS and SRIVASTAVA, 1989: 197). With the use of probabilistic models we try to minimise or at least to quantify these uncertainties.
As rainfall and the involved atmospheric process are not fully understood, we have to rely on a probabilistic model if we want to predict unknown values.

The starting point of a probabilistic model is that the observed or measured data are viewed as the outcome of a random process. It should be obvious that such an assumption collides with reality as no physical processes, be it precipitation, the runoff of a river or the growth of vegetation, and is actually based on random coincidences. Our poor understanding of those extremely complicated and interconnected mechanisms let us assume that we have to deal with randomly

generated processes. In other words: our ignorance leads us to the probabilistic model path. However, it would be wrong to conclude that if we have to deal with a set of randomly generated data it is therefore impossible to predict all the other, also randomly, generated data. In practice probabilistic models have proved to be very good in estimating unknown values and a useful set of tools, based on probabilistic models, had been developed in order to interpolate irregularly spatial organised data.

In a random model one decides how to weight the nearby samples so that the outcome (the estimate) is unbiased. The unknown value $\hat{v}$ we want to estimate is expressed through a linear combination of the known and available samples:

1) $$\hat{v} = \sum_{j=1}^{n} \left(w_j \cdot v \right)$$

where v is the known sample and w_j represents the weight we apply to this point v. n is the number of available samples in our data set.

4.2 Different Models, different Worlds

The following chapters provide a brief overview of the major models and methods used for the spatial interpolation of rainfall.

4.2.1 Thiessen Polygons

Analysing the spatial variability of hydrological processes in general and precipitation in particular had always been of great concern to scientists.
As early as 1911 A. H. Thiessen suggested a method for the estimation of the precipitation distribution by assigning polygonal areas to each rainfall data point (Thiessen-Polygons). The weighted mean precipitaion for the entire area was then calculated by:

$$P_{mean} = \frac{\sum \left(A_i \cdot P_i \right)}{\sum A_i}$$

where
P_{mean} = Weighted Mean Precipitation
A_i = Area of Polygon i
P_i =Precipitation measured at point i
(After: WILHELM, 1993: 47)

4.2.2 Delaunay Triangulation

Precipitation maps drawn by the Polygon Method will always suffer from an undesired discontinuity[24].

[24] The limits of the polygons form the edges of areas containing the same estimated precipitation. By moving from one area to the other the estimated precipitation suddenly jumps from one value to the other.

4

The Triangulation Method tries to eliminate the discontinuity effect by fitting a plane through three samples that surround the point being estimated. This plane is described as:

2) $$z = ax + by + c$$

The three coefficients of 2) (a,b,c) can be computed by solving a set of equations which contain the three points at the edges of the polygon:

$$ax_1 + by_1 + c = z_1$$
$$ax_2 + by_2 + c = z_2$$
$$ax_3 + by_3 + c = z_3$$

where:
(xi,yi,zi) is a corner point at the coordinates xi (easting), yi (northing) and with the observed precipitation value of zi.
After: ISAAKS and SRIVASTAVA, 1989: 252)

By inserting coordinates (x,y) in the solved equation 2) we are now able to compute the estimated precipitation z at any desired point in the plane. By calculating all equations for all triangles in a given area we will end up with a reduced discontinuity at the borders of each polygon. The Delaunay Triangulation is fairly easy to calculate and it produces triangles that are as close to equilateral as possible. Three samples form a Delaunay triangle if their polygons of influence share a common vertex. However, it is very unwise to estimate a point not located within the borders of a Delaunay triangle as this leads us to terra incognita (The estimated point is not part of a calculated triangle).
The Delaunay method cannot be used to extrapolate points that are located close to the borders of an area of interest (see Figure 4.1 on page 56).

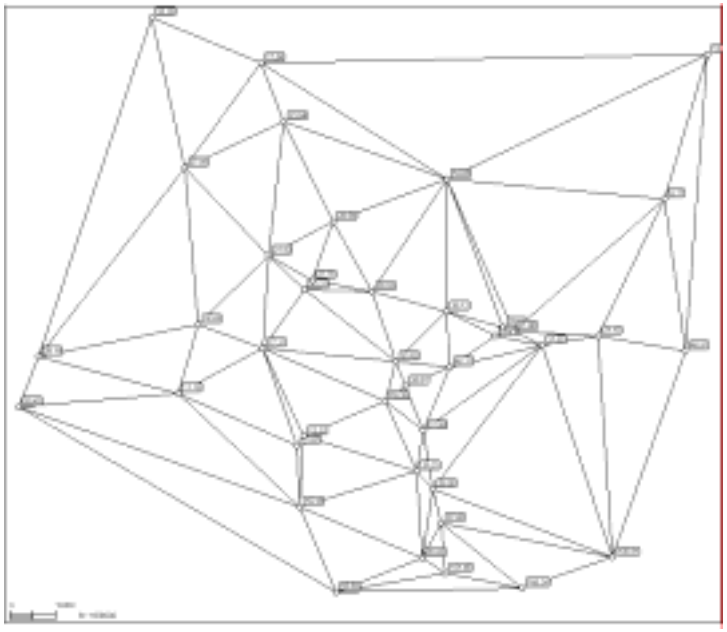


Figure 4.1: Delaunay Triangulation of the monthly mean precipitation values for May 1967-1997 of the Upper Ewaso Ng'iro Basin. Note that there are no estimates possible for points that lay outside the mesh. (Source: B. Sturm, Computed with 3DField 1.71)

4.2.3 Inverse Distance Method

By using only limited information about the nearby samples in the polygon and triangulation method we may not properly consider extreme variations in a group of very close sample points. If one weights the samples differently one should get a better estimate as local extremes are treated differently and not equal as in the two previously described methods. Inverse distance methods belong to a family of distance weighting techniques that use the distance between sampling points and the points of interest for the interpolation function. (TABIOS and SALAS, 1985: 367):

$$w_j = \frac{f(d_{oj})}{\sum_{i=1}^{n} f(d_{oi})}$$

where $f(d_{oj})$ represents a given function of the distance d_{oj} between an observed point (xo,yo) and an point to be estimated (xj,yj) and w_j is the calculated weight. For $f(d_{oj})$ the following function is commonly used:

$$f(d_{oj}) = \frac{1}{d_{oj}^{b}}$$

where b is an appropriate constant. It is of importance to note that the weight approaches zero as the distance and/or the parameter b increase. If we set b to the value of 1 the technique is referred to as the Inverse Distance Method, and with a b of 2 one refers to the Inverse Square Distance Method.

An important aspect of the Inverse Distance Method is the consideration of what we actually mean with 'nearby' samples.
The search for the neighbourhood around a sample that controls which samples are included in the estimation procedure is an important consideration in statistical approaches. There are different varieties of the Inverse Distance Method available which use either a circular neighbourhood range or search for a minimum number of gauges around the sample point to be estimated (see Figure 4.2 on page 57).

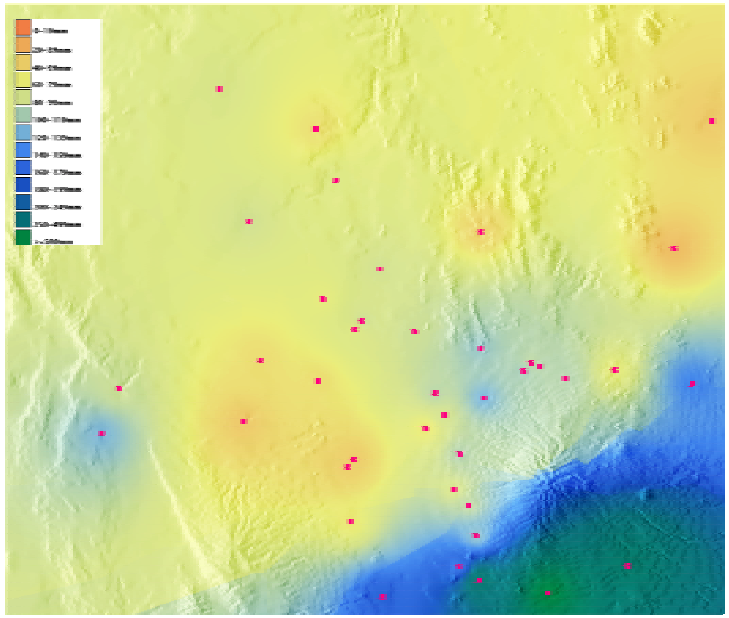


Figure 4.2: Inverse Distance estimation for the monthly mean precipitation for May 1967-1997 of the Upper Ewaso Ng'iro Basin. For the interpolation the 12 nearest neighbouring gauges for each estimated point was used. (Source: B. Sturm, Model: r.surf.idw, GRASS5b10)

4.2.4 Kriging

Kriging is a widespread estimation technique and was applied for many hydrological studies. The Kriging methodology is derived from the so-called Optimal Interpolation method. Departing from the equation 1) the Optimal Interpolation technique determines the weights given in 1) by minimising the variance of the error of interpolation σ^2_ε. Where

the Optimal Interpolation method uses spatial correlation functions to minimise the variance of the error of interpolation, the Kriging technique replaces the spatial correlation function with a so-called variogram. Kriging interpolation requires that the observed process is second-order stationary. Essentially, this assumes homogeneity in the means, variances and covariances. An isotropic spatial covariance structure is also assumed (TABIOS and SALAS, 1985). Other Kriging techniques such as the universal Kriging proposed by Delfiner and Delhomme in 1975 try to incorporate nonhomogeneity in the mean of the process by representing the mean m_o at a given point (x_o, y_o) as a linear combination of the observed station means. As these means are unknown they may be represented by a polynomial trend (TABIOS and SALAS, 1985).
Kriging is mainly preferred because it offers unbiased estimates and estimates of the interpolation error (DINGMAN, et. al, 1988). However Dingman et. al mention that according to a study from Bras and Rodriguez-Iturbe from 1985, Kriging tends to underestimate the interpolation error. Also does Kriging not provide better predictive results than much simpler techniques like the inverse square distance method. This is mainly true for high-resolution networks (for instance 13 raingauges for an area of 35km^2) (GOOVAERTS, 2000: 114). Beside the estimated errors, another advantage of Kriging is that it can be used together with associated secondary variables that lead to an extension of the Kriging technique, called co-kriging. Co-kriging incorporates a multivariate approach where sparsely sampled rainfall observations are combined with densely sampled information like weather-radar observations.

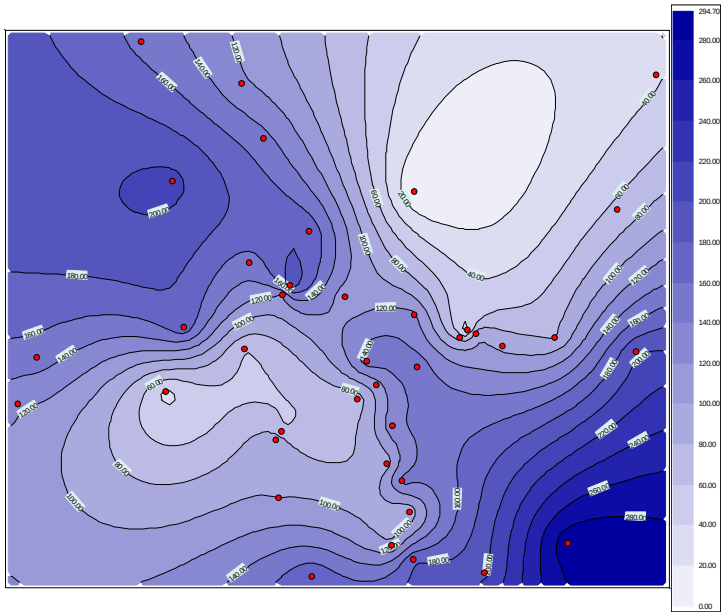


Figure 4.3: Kriging Interpolation for mean decadal rainfall from 01.05-10.05 1967-1997 of the Upper Ewaso Ng'iro Basin. (Source: B. Sturm, Computed with 3DField 1.71)

Another application of co-kriging is demonstrated by P. Goovaerts who has used the

multivariate approach to combine topographical information, provided by a DEM, with observed rainfall to estimate the areal topography dependent rainfall for the Algarve in Portugal (GOOVAERTS, 2000). Topography is cheap and abundant additional information that is related to rainfall because precipitation tends to increase with increasing elevation[25].
Despite the various advantages of Kriging, there is one big drawback: "*Kriging requires the spatial covariance function or variogram to be estimated first, and is critical to process.*" (KESTEVEN and HUTCHINSON, 1997)

There is one technique left, which offers a slightly less critical handling than Kriging, and is also based on the strive for minimum error estimation: the thin plate splines model.

4.2.5 Thin Plate Splines and Relatives

Like Kriging, the thin plate spline method models spatial distribution as a function of observational data across a region without prior knowledge of the distribution or its underlying causes, and both methods essentially have the same underlying computational structure (KESTEVEN and HUTCHINSON, 1997).

[25] This is mainly due to the orographic effect, which causes air to be lifted vertically, and hence causing condensation due to adiabatic cooling.

Kriging and thin plate splines base on the same computational formulation in relation to observed data:

$$y_i = z(x_i) + e_i (i = 1,..., n)$$

where:

- y_i are observed values
- n number of observed values
- x_i position of the observed values
- $z(x_i)$ function to be estimated from the y_i observed values
- e_i mean error term (usually spatially discontinous)

In the case of kriging $z(x_i)$ are assumed to be values of a spatially auto correlated random field, whereas in the case of splines the $z(x_i)$ are assumed to be values of a smooth unknown function (taken from HUTCHINSON, 1999: 4).
However there are more significant differences between these two multivariate geostatistical interpolation approaches. As already stated Kriging requires a prior calibration of the spatial covariance function (variogram), where on the other hand, thin plate splines uses a smoothing parameter that determines the optimal balance between fidelity to the data and smoothness of the fitted spline function. Most thin plate splines algorithms compute this parameter automatically by minimising the GCV (Gener-

alised Cross Validation), and hence are much easier to use than Kriging (HUTCHINSON, 1998).
But before we discuss the thin plate splines and its derived methods exhaustively we will have to take a look at the basic pillars of this method.
The general idea behind splines is that the resulting interpolation function is a close approximation to the surface that represents the topographical plane described by the variables to be estimated.
This approximated surface can be imagined as a thin steel plate that is forced to pass through the constraints representing the data points of the observed variable (MITASOVA, 1993) (for an illustration see Figure 4.4 on page 60).
The thin plate spline method has its origin from a growing need for not only spatial estimation approaches but also for surface analysis where the interpolation function minimises an appropriate function which represents some measure of smoothness of this function.

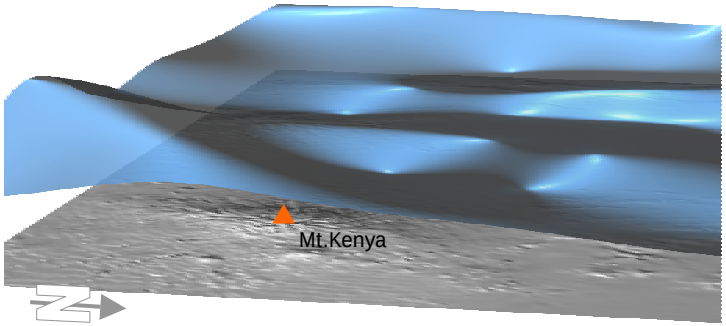


Figure 4.4: 3 dimensional view of the interpolated mean precipitation over the study area from 01.05.-10.05. for the years 1990-1997. The light blue plane is interpolated by means of regularised splines with tension. Note the typical 'hills' and 'valleys' representing areas of high and low rainfall. The scale is 70 times exaggerated. The grey plane represents the underlying topographic structure (Mt. Kenya is in the lower left corner, the grey arrow points towards geographic north). Source: B. Sturm, visualisation by NVIZ 2.2.

The thin plate splines (TPS) function is directly derived from the minimum curvature (MINQ) interpolation introduced by Briggs, 1974 (MITASOVA, 1993). The MINQ approach was computationally costly and had a tendency to produce overshoots due to plate stiffness in regions with rapid change of gradients. Attempts to overcome these problems were made (e.g. by Hutchinson in 1989), but all those methods needed a numerical solution of the variational condition, which was time consuming.

The TPS method reduces calculation time by using an explicit solution for the variational condition (MITASOVA, 1993). However the TPS model was still very difficult to handle and was not widely used in geosciences because of its *"[...] global character and consequent difficulties in application to large datasets."* (MITASOVA, 1993: 4).

Regularised Splines with Tension

Departing from the TPS method, Helena Mitášová and Mitáš Luboš have developed a new, more robust and reliable spline interpolation method: the regularised splines with tension. With the regularised splines with tension it is possible to control the behaviour of the resulting surface from membrane to thin plate. The value of the generalised tension parameter has to be determined empirically. But according to the authors *"[...] experience suggests that suitable values may be found within a few trials."* (MITASOVA, 1993: 8).
The following properties characterise the regularised splines with tension (RST) model:

- Accuracy: With the proper choice of the tension parameter the RST gave the most accurate estimates.
- Flexibility: Because the tension parameter can be used for interpolations from membrane to thin plate, overshoots due to gradient changes can be avoided.
- Local behaviour: This enables the application of segmented processing of large datasets (tens of thousands of data points)[26]

Mitášová and Luboš have incorporated their proposed method in the open source GIS GRASS as a surface interpolation module. The command was first introduced as r.surf.tps (GRASS 4.1) and later as the regularised splines with tension (RST) command r.surf.rst (GRASS 5.0beta).

Because Kriging and RST do share common basic structures, a direct comparison of the two methods in this work was not preferred[27], however a cross validation between the inverse distance weighting and the RST estimation method was used to assess the accuracy and reliability of these two models. The ap-

[26] The greater the tension at a data point the more local is the behaviour of the RST function. The proposed algorithm incorporates an automatic determination of optimal size of segment and flexible size of overlapping neighbourhood for heterogeneous spatial distributions (MITASOVA, 1993).
[27] For a direct comparison between Kriging and thin plate interpolation see Saveliev et al., 1998: "Modeling of the Daily Rainfall Values Using Surface Under Tension and Krigin", Journal of Geographic Information and Decision Analysis, vol. 2, no. 2, 58-71, 1998. (ftp://ftp.geog.uwo.ca/SIC97/Saveliev/Saveliev.html)

plication and the results of this comparison are presented in the next chapter.

5 Cross Validation

In order to assess and quantify the grade of quality of the spatial precipitation interpolation method a cross validation approach had been chosen.

As an outcome of this validation process it was possible to determine which interpolation method provided the best spatial estimates. A simple cross validation approach was applied to the two spatial interpolation methods (Inverse Distance Weighting and Regularised Splines with Tension) GRASS offers. However, despite GRASS is also able to interpolate point data with the Ordinary Kriging method this method was not chosen due to the fact that Kriging depends critically on a prior estimation of the spatial covariance function. This estimation is a very laborious process and can hardly be determined by an automated process (HUTCHINSON: Modelling Spatial and Temporal Variability of Climate and Terrain, 1997). See chapter 4 (Spatial Interpolation of Rainfall: Methods and Theory) on page 53, for a detailed discussion of the applied interpolation methods.

The results of the cross validation are presented in the following chapters, furthermore a comparison with other estimation models is provided in Chapter 5.6.2 (Case Studies).

5.1 Cross Validation

A good measure for the fidelity (or the estimated error) of an estimation method is to estimate known values through other known values. This is achieved by discarding the sample value at a particular location from the sample data set ("leaving-out-one-sample" method). The remaining samples are then used to estimate the removed observed value. The difference between the observed and the estimated values can be used as a direct measure for the quality of the applied estimation method. Isaaks and Srivastava described two main interpretation methods of Cross Validation: a quantitative and a qualitative approach (ISAAKS and SRIVASTAVA, 1989: 352).

In a quantitative approach the standard statistical tools such as summary statistics, quantile-quantile plots, summary statistics of residuals are used to describe the relationship between estimated and true values. A qualitative approach, on the other hand, helps to reveal the spatial features of residuals by drawing any overestimated or underestimated values at their respective spatial location. A qualitative approach can show geographical regions where one has to treat the results of a used interpolation technique with caution.

Both described approaches were applied to the Upper Ewaso Ng'iro datasets.

5.1.1 Quantitative Cross Validation

In order to show the degree of dependence between the true and estimated variables, scatterplots and linear regression for all analysed situations were plotted and computed. Summary statistics such as mean error (ME), mean absolute error (MAE), mean relative error (MRE) and the mean squared error (MSE) helped interpreting the results of the Cross Validation.

A reasonable goal of every estimation is to produce an unbiased estimate that is a balance between overestimation and underestimation should be achieved. Any ideal error distribution should therefore show a symmetrically distributed error (or residual) histogram, which is centred on the median and the mean error. An unbiased distribution will have a mean error of 0 (overestimation and underestimation are balanced) Any other value than 0 of the mean error let us argue that the estimation may be biased. Not only the skewness of the bias of the residual histogram is of importance, but also the spread of the error distribution. A good yardstick for assessing the spread of the error distribution is the variance or the standard deviation.

However it is not always possible to achieve an unbiased error distribution together with a minimum spread.

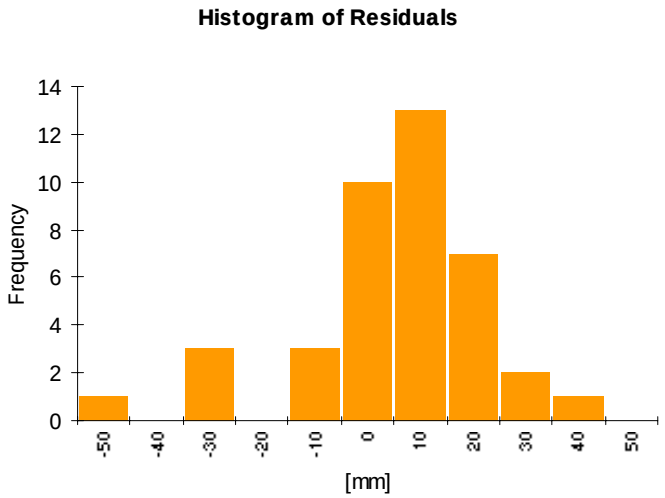


Figure 5.1: Histogram of the residuals. This interpolation model tends to slightly overestimate data points.

Therefore we will have to trade one off against the other.

Isaaks and Srivastava (1989) proposed to use the mean absolute error (MAE) and the mean squared error (MSE) as the two summary statistics to incorporate both the bias and the spread of the error distribution.

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|x_e - x_t\right|$$

Mean Squared Error (MSE)

$$MSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(x_e - x_t\right)^2}$$

where:

n = number of rain gauges in the dataset

x_e-x_t = Residual for gauge i

x_t = True or observed value

x_e = Estimated value

Additionally the mean error (ME) and the mean relative error (MRE)
were calculated in order to gain further information of the behaviour of the residuals.

Mean Error (ME)

$$ME = \frac{1}{n}\sum_{i=1}^{n}\left(x_e - x_t\right)$$

Mean Relative Error (MRE)

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{100\left(x_e - x_t\right)}{x_t}$$

where:

n = number of rain gauges in the dataset

x_e-x_t = Residual for gauge i

x_t = True or observed value

x_e = Estimated value

After: BAHRENBERG 1990:51 and ISAAKS, SRIVASTAVA, 1989

Further quantitative analysis include scatterplots and the drawing of the linear regression line. An assessment of the regression parameters, such as interception and the coefficient of determination, are commonly used to gain insight about the degree of dependence of the estimated against the true values. Chapter 5.3 (Statistical Analysis) describes the used summary statistics in detail.

5.1.2 Qualitative Cross Validation

As stated above a qualitative assessment can show the spatial behaviour of the applied estimation technique. In the presented work all residuals were plotted at their geographical location using untransformed UTM coordinates[28].

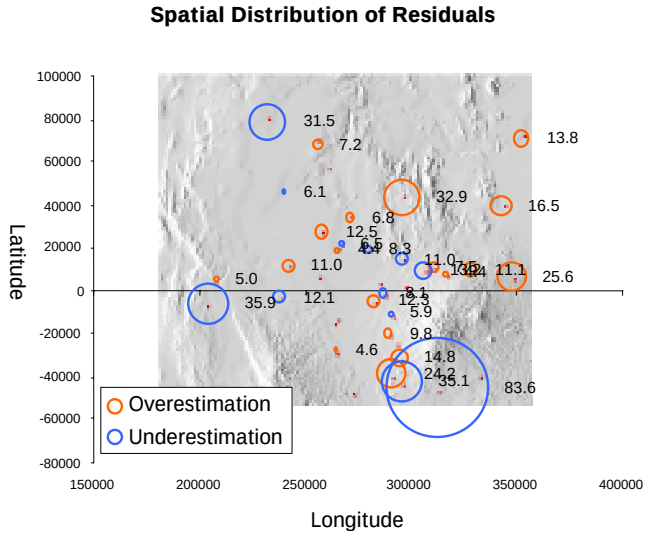


Diagram 5.1: Example diagram of the spatial distribution of residuals for monthly estimates. Red circles mark overestimated points, and blue connotes to underestimates. All values in mm and locations in UTM coordinates.

[28] Due to the equator UTM co-ordinates had to be transformed into negative values. This was only necessary for the GIS in order to process the interpolation correctly.

This spatial distribution was plotted for each estimation cycle and an interpretation helped to reveal any geographical weakness of the interpolation model.

5.2 Application

For this work cross validation was applied to the monthly mean values from 1.1.1967 - 31.12.1997 (31 years), thus resulting in 12 data sets.
For the n rain gauges in the study area an interpolated rainfall raster map was computed by using n-1 rain gauges. For each cycle one gauge was cancelled out and its amount of rainfall was estimated by using the observed values from the remaining stations.

In order to determine a best fit for the interpolation parameters this calculation was repeated for each method and each parameter change.
One can clearly see that this approach itself is a very time consuming expedition. The number of computations is defined by the number of gauges, the time resolution of each data set, the number of interpolation methods and the parameters, which control these methods.

As an example the validation of the Inverse Distance Weighting (IDW) method used:

40	Rain gauges
12	Months
1	Interpolation method (IDW)
4	Different parameter changes
40 x 12 x 1 x 4 = 1200 Calculation cycles	

Depending on the available computer power this calculation will take approximately 72 CPU hours (AMD K6-266Mhz)[29].
In order to automate the described cross validation process a Perl routine (crossval) was written which controlled the necessary GRASS GIS modules and provided, as an output, the observed and estimated values for each gauge. In order to preserve flexibility for future analysis, the routine is able to adapt to different data sets, and can easily be changed to future GRASS interpolation modules (see chapter 6.2.6 Modules on page 93 for more details).
The output of crossval was imported and analysed with MS Excel. The next chapter will describe this process in detail.

[29] Computation time can be saved by masking areas on the output raster map, this procedure is described in the GRASS online manuals for s.surf.rst, but is only applicable for Regularised Splines with Tension: Mitasovas Helena, November 1999, GRASS 5.0beta, Online Manual, s.surf.rst

5.2.1 Process Flow of crossval

Here we will describe the used cross validation process in detail. A process flow diagram will help to understand the Perl script. The source code can be found on the accompaigning CD-ROM or in chapter 13 (Perl Scripts) on page 173, and is freely distributable.

Crossval Version 0.09b

Start

Enter: Filename or list of filenames with rainfall point values.

Count the number of files to be processed ⇨ *List*
Count the number of gauges found in the rainfall file. ⇨ *stat_count*

For *j*=1 to *list*

For *i*=1 to *stats_count*

Save row *i* of *file j* as a definition file
This is the station for which the rainfall value will be estimated
Save all other rows of *file j* as an interpolation file
The stations in the file are used for the interpolation
Import the interpolation file into grass as site data (s.in.ascii)
Interpolate raster map with site data (s.surf.idw / s.surf.rst)
Query the raster map with the definition file in order to get the estimated values for the omitted gauge (r.what)
Remove the raster and the site map (g.remove)

Next *i*
Join and save the results of the queries in a comparison file

Next *j*
Remove all temporary files

End

Diagram 5.2: Flow Diagram of the Perl script 'crossval'

5.3 Statistical Analysis

After the successful interpolation crossval outputs a comparison file that can be used for statistical analysis with the EXCEL macro "CrossValAnalysisMacro.xls[30]" under Windows. The macro automatically processes all found crossval result (comparison) files and provides a worksheet which contains the following statistical summaries:

- Spatial Residual Plots
- Regression Analysis
- Residual Histogram
- Mean Error of Residuals (ME)
- Mean Relative Error of Residuals (MRE)
- Mean Absolute Error of Residuals (MAE)
- Mean Squared Error of Residuals (MSE)
- Median of Residuals

As described in chapter "Quantitative Cross Validation" (chapter 5.1.1) the goal of any estimation model is to produce residuals of 0 and a scatterplot of true and estimated values would plot us a straight line (ISAAKS and SRIVASTAVA (1989): 260). That is there is no difference between any observed and estimated rainfall value of a data set. As this is very unlikely to happen in a real-world scenario, we will have to find some criteria, which enable us to separate a 'good' from a 'bad' interpolation model.
For this we will study the univariate distribution of errors and the bivariate distribution of estimated and true values. (after ISAAKS and SRIVASTAVA, 1989)

5.3.1 Univariate Distribution of Errors

Any error is calculated using the following formula:

$$error = r = \hat{v} - v$$

where:

$\hat{v}$ is the estimated value by the model
v is the true value

This error is also referred as residual (r).
As stated above we would expect our error distribution to be bias free and show a spread close to 0. Further we have to make sure that the histogram of the distribution of errors is symmetrically centred around the median of the errors (which should be 0 in order to compensate any over- or underestimation).
A good check on the symmetry is the median of the computed error. We will therefore have to keep an eye on small median values in order to isolate good estimates. The Mean

[30] This macro can also be found on the CD-ROM provided with this paper.

Absolute Error (MEA) and the Mean Squared Error (MSE) will help us to get a small biased error distribution (MEA) and a small spread of the error distribution (MSE). So far we have discussed desirable properties of the entire error distribution. We would also like to see if these properties hold for any range of estimated values.

5.3.2 Bivariate Distribution of Estimated and True Values

By plotting a scatterplot of the true (observed) versus the estimated (predicted) values we will see how well an estimation method has performed. Ideally we would expect all points grouped on a single straight line, which would reflect our initially postulated zero residual condition.

In actual practice we will always have to live with some error in our estimates, and our scatterplots will always appear as a cloud of points. The mean error is correlated with the scatterplots in that way that a mean error of 0 for any range of estimates will lead to a conditional expectation curve of true values given estimated ones plotted on a 45-degree line.
The correlation coefficient of the linear regression gives us a good index for summarising how close the plots on a scatterplot are grouped on this 45-degree line.

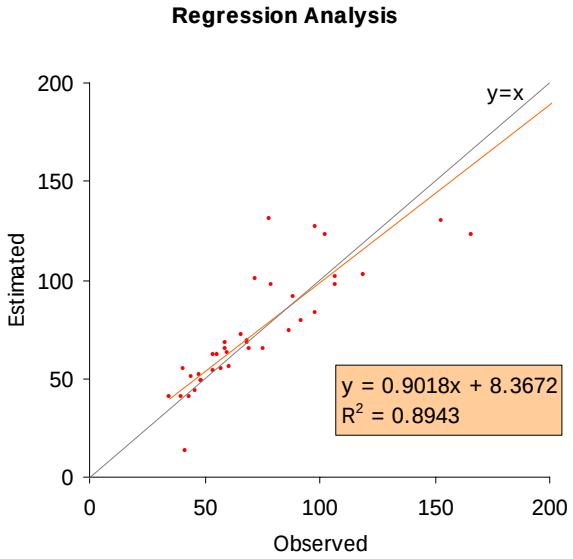


Diagram 5.3: Exampel of a regression curve for estimated versus observed data points. (Interpolation model: Regularised Splines with Tension for monthly mean values (October 1967 – 1997)

The cause of a global bias can be made apparent by analysing the spatial distribution of the residual plots and the scatterplots of the true values versus the estimated values. This will reveal where an estimation method has its geographical strength and weakness, or if the interpolation model tends to treat low values different than higher ones.

Besides the parameters of the line of regression, the coefficient of determination (B or R^2) is being calculated. As B explains the variance of the regression of the estimated through the true values, we will use B as a measure for the quality of our regression line. If the coefficient of determination is 1, all points are grouped on the line of regression and we can assume that there is a strong coherence between estimated and true values (BAHRENBERG (1990): 147).

5.4 How did it perform?

This chapter will discuss the outcome of the described cross validation approach. But before we jump to the results we will have to take a look at the parameters that were used for the interpolation methods.

5.4.1 Parameters used

① Inverse Distance Weighting Method (IDW)

As the interpolated value of a cell is determined by values of nearby data points and the distance of the cell from those input points, the IDW model in GRASS uses only one parameter to control the behaviour of the interpolation: the number of points which should be used for the interpolation.

By default this is set to the 12 nearest points. However, experiments prior to the cross validation of the IDW method showed that reasonable results can be achieved by using the 5 nearest points for the Upper Ewaso Ng'iro datasets.

In order to find the best fit for this parameter the following settings for the nearest neighbours were used for cross validate the Inverse Distance Method:

Nearest Neighbour Paramters used by module r.surf.idw	
5	Gauges
10	Gauges
12	Gauges (default)
15	Gauges

Table 5.1: Parameters used for the cross validation of the Inverse Distance Weighting model.

② Regularised Splines with Tension (RST)

The RST method used by GRASS can be controlled in various ways, and therefore offers more possibilities to define different parameter settings.
Crossval uses the following major parameter to steer the interpolation with the RST method:

- Tension
- Smooth
- dnorm
- dmin
- npmin

Tension

The GRASS5beta7 manual describes the tension parameter behaviour as follows:
"It is useful to know that the method is scale dependent and the tension parameter works as a rescaling parameter (high tension 'increases the distances between the points' and reduces the range of impact of each point, low tension 'decreases the distance' and the points influence each other over longer range). Surfaces with tension set too high behave like a membrane with peaks or pits in each given point and everywhere else the surface goes rapidly to trend. Surfaces with tension set too low behave like a stiff steel plate and overshoots can appear in areas with rapid change of gradient and segmentation can be visible."
Care was taken by selecting a set of good tension values. Good results were achieved by varying the tension parameter only slightly around its default value (40).

Smooth

The smoothing parameter can be used to ensure that the interpolated surface passes through the data points. With smoothing parameter greater than zero the surface will not pass exactly through the data points and the higher the parameter the closer the surface will be to the trend. Only parameters close to 0 were therefore used.

dnorm

The dnorm parameter rescales the coordinates depending on the average data density so that the size of segments with segmax=40 points is around 1 - this ensures the numerical stability of the computation. Dnorm works in relation to the tension parameter in that way that the given tension is applied to normalised data (x / dnorm..), that

means that the distances are multiplied (re-scaled) by tension/dnorm.

dmin
Sets the minimum distance in grid cell size between points. The default value is set to 0.5 grid cell size. In order to show the influence of a high dmin a distance of 50 grid cell size was used for one set of interpolation.

npmin
The number of pionts taken for the interpolation is controlled by npmin, this value must be greater than the value set for the maximum number of data points where no segmentation will occur (segmax). Both parameters (npmin and segmax) were set to a value of 700 in order to avoid segmentation.
Table 5.2 lists the parameters which had been used for the cross validation of the RST model.

Tension	Smooth	Dnorm	Dmin	Npmin	Segmax
40	0.5	N/A	1	700	700
60	0.1	Applied	1	700	700
44	0.05	N/A	1	700	700
36	0.1	N/A	0.5	700	700
45	0	N/A	50	700	700

Table 5.2: Parameters used for the cross validation of the Regularised Splines with Tension model.

5.5 Analysis and Results

After the post-processing of the crossval output files with the EXCEL macro Cross-ValAnalysisMacro.XLS a sound analysis of the results could be performed. For each validation step a result sheet containing the summary statistical key values was printed (see Diagram 5.4).

5.5.1 Quantitative Analysis

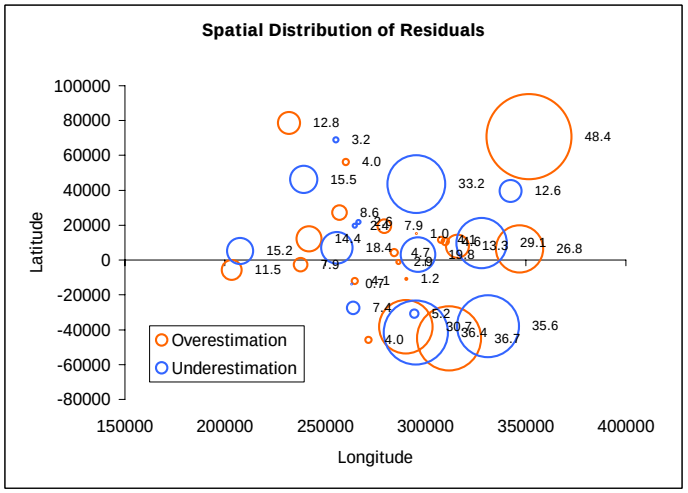
As Discussed in chapter 5.1.1 (Quantitative Cross Validation) summary statistics based on the residuals were calculated. These statistics were used to determine the grade of quality of the interpolation model. In a first step we will take a look at the results of the linear regression:

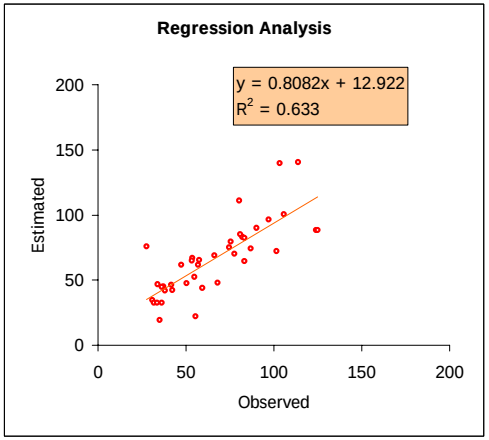
Linear Regression
① Inverse Distance Weighting Method (IDW)
The Inverse Distance Weighting Method unveiled quite a consistent image: the best regressions can be found where the nearest neighbourhood parameter was set to 5. The mean slope value for all months is at 0.5523, and the highest slope was achieved for February (0.7765). By studying the coefficient of determination (B) the image blurs a bit: the highest value for B (0.7214) can be found for January where interpolation with the 15 nearest stations was performed.

Cross Validation

6797MAR_RST

Parameter Setting		
	Tension=	44
	Smooth=	0.05
	Dnorm=	not used
	Dmin=	1
ME		**0.29881**
MAE		**12.2162**
MRE		**4.8074**
MSE		**17.579**
Median		**0.72746**





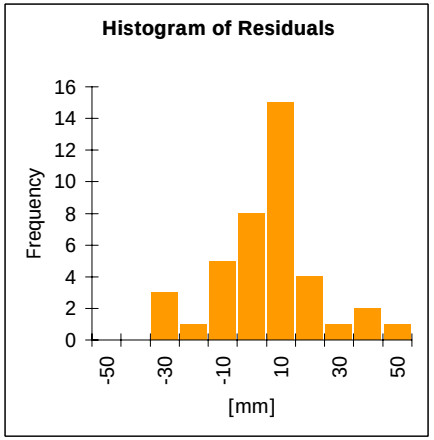


Diagram 5.4: Example of a post processing sheet produced by "CrossValAnalysisMacro.XLS". The above results stem from the monthly mean values of March from 1967-1997, Interpolation method is Regularised Splines with Tension.

However, by taking the 12 nearest neighbours into account, the IDW method shows the highest mean value for B over all performed interpolations at 0.596525. Therefore, the best-case scenario leaves us with an explained variance of only 59.7% through the regression of estimated and observed values. Such a low degree of dependence between estimated and observed data points may be a sign of wide scattered residuals.

Near. Neighbours	Slope (a)	Interception (b)	Coeff. of Determination
5	0.7765 (FEB)	7.5275 (JAN)	0.6641 (FEB)
10	0.7278 (FEB)	8.9534 (JAN)	0.6911 (FEB)
12	0.7117 (FEB)	9.3748 (JAN)	0.7093 (JAN)
15	0.6863 (FEB)	10.133 (JAN)	0.7241 (JAN)
Mean	**0.5523 (5)**	**29.9091 (5)**	**0.5965 (12)**

Table 5.3: IDW method best fit values of the line of regression over all calculations including mean values. Values in brackets indicate months or the number of the nearest neighbours.

Indeed, by comparing the mean squared errors (MSE) we find the best mean value for the MSE to be 25.944 (5 Nearest Neighbours).

② Regularised Splines with Tension (RST)

The RST method performed much better concerning linear regression. The best slope can be found at 0.9127 (December) and a tension parameter of 44. The interpolation where the tension parameter was set to 44, was also successful by providing the best explained variance of 73.8%. On the other side of the scale we find an explained variance of only 0.26% where the dmin parameter was set to 50 grid cell size. By discarding all the results of dmin=50, the explained variance (B) ranges from 34.1% (September and tension at 60) to 89.4% (October and tension at 44).

Tension	Slope (a)	Interception (b)	Coeff. of Determination
36	0.8655 (OCT)	5.997 (JAN)	0.8366 (APR)
40	0.8651 (OCT)	5.997 (JAN)	0.8828 (OCT)
44	0.9127 (OCT)	5.1269 (JAN)	0.8943 (OCT)
45	1.0623 (JAN)[31]	-1.2278 (AUG)	0.1529 (FEB)
60	0.4784 (FEB)	17.465 (JAN)	0.7549 (JAN)
Mean	**0.8282 (44)**	**8.675 (45)**	**0.7378 (44)**

Table 5.4: RST method best fit values of the line of regression over all calculations including mean values. Values in brackets indicate months or the tension parameter.

Histogram Analysis

① Inverse Distance Weighting Method (IDW)

First we will have a look at the worst interpolation run, which was performed by including the 15 nearest neighbouring gauges. For the month of May we find a mean squared error of 62.0mm and a mean absolute error of 32.4mm. These values indicate a high bias and a wide spread histogram. This is con-

[31] This result should be treated as an outlier because of a very low B=0.0821! Which suggests almost complete independence of the two variables.

firmed by visually examining the histogram for this run:

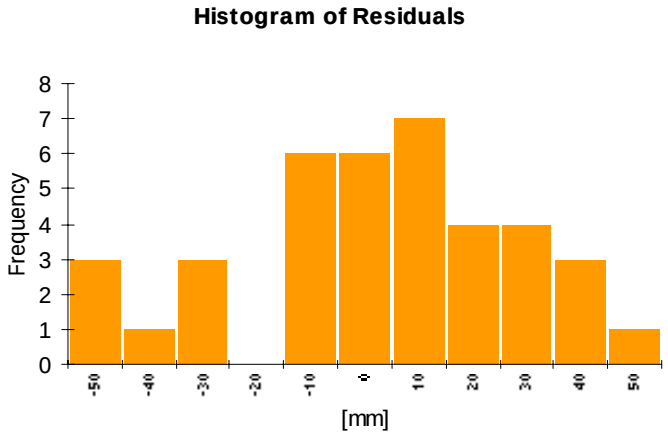


Diagram 5.5: Histogram of residuals for IDW interpolation, May 1967-1997, Nearest Neighbours = 15.

Diagram 5.5 reflects clearly the high summary statistics for the frequency of residuals. From the histogram diagram we can conclude that this interpolation will tend to overestimate precipitation.
The best interpolation is also found at the same parameter settings with nearest neighbours of 15: the estimation for February gave a MSE of 8.9mm and a MAE of 6.4mm:

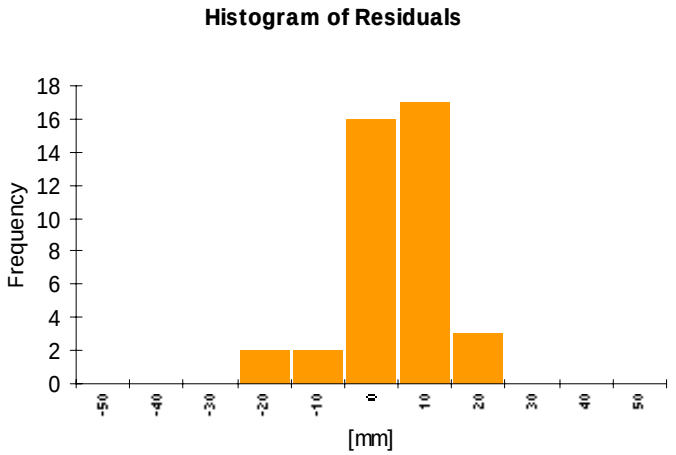


Diagram 5.6: Histogram of residual for IDW interpolation, February 1967-1997, Nearest Neighbours = 15.

Although February is a month with a high variability of expected rainfall in the Ewaso Ng'iro Basin, the inverse distance model with the 15 nearest neighbours estimates this rainfall very well as the calculated median of the residuals is 0.06mm!
However, despite the fact that we find the worst and the best performance of the IDW interpolation where the nearest neighbours parameter was set to 15, we find the overall best results again at the runs where the 5 nearest neighbours were used:

Near. Neighbours	MSE [mm]	MAE [mm]	Median [mm]
5	**25.944**	**15.625**	**1.202**
10	26.673	16.193	1.352
12	26.924	16.232	1.316
15	27.293	16.481	1.490

Table 5.5: Comparison of the residual mean values of MSE, MAE and the median of the IDW interpolation model.

② Regularised Splines with Tension (RST)

By studying the histograms produced by the RST model, we can find a far wider spectrum of spread and bias values compared to the IDW method. Again the worst results can be found where a minimum distance between the data points was set to 50 grid cell size. For the month of May we get a MSE of 157.409mm and a MAE of 87.562mm (see: Diagram 5.7) and a median of –0.673mm. Here we have a slight tendency to underestimate the point data.

By discarding all dmin=50 results, we find the worst estimation at dmin=1 and tension set to 60 for the month of May (MSE=68.016mm MAE=37.594 median=9.623mm)

The best performance of the RST model unveils a mean squared error of 8.566mm and a mean absolute error of 6.292mm (median=0.886) for the month of February and a tension parameter of 36.

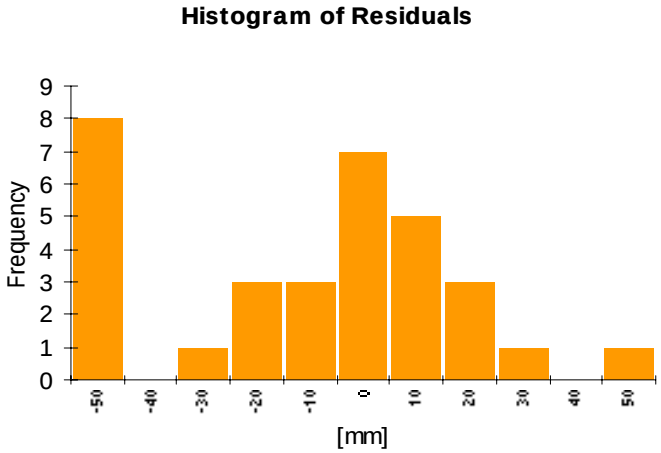


Diagram 5.7: Histogram of residuals for RST interpolation, May 1967 – 1997, Tension = 45, Smooth = 0, Dnorm = disabled, Dmin = 50.

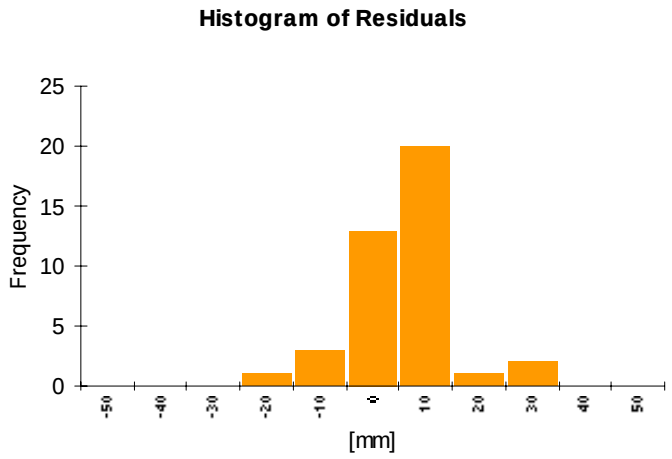


Diagram 5.8: Histogram of residuals for RST interpolation, February 1967 – 1997, Tension = 36, Smooth = 0.1, Dnorm = disabled, Dmin = 0.5.

As with the IDW model, the extreme values are not reflected in the analysis of the mean values and the best mean results are to be found where the tension parameter of 44 was applied (smooth=0.0, dnorm=disabled, dmin=1):

Tension	MSE [mm]	MAE [mm]	Median [mm]
36	19.704	12.870	0.764
40	19.373	12.706	0.507
44	**19.279**	**12.700**	**0.367**
45	111.158	62.626	-1.056
60	29.802	19.394	5.124

Table 5.6: Comparison of the residual mean values of MSE, MAE and the median of the RST interpolation model.

5.5.2 Qualitative Analysis

As described in chapter 5.1.2 (Qualitative Cross Validation) a qualitative analysis was performed by plotting the estimated errors at their geographical location. In this chapter we will show for the best interpolation how the spatial pattern of the estimated errors is represented.

① Inverse Distance Weighting Method (IDW)

By comparing the monthly distribution of the interpolation made with the 5 nearest neighbouring stations one can distinguish two different patterns:

1. The first pattern appears at the edges of the area. Here we find remarkable high divergent residuals. This is mainly apparent at the north-eastern and southern borders of the estimated station. Whereas the gauges in the centre and the western part of the study area are well estimated by the inverse distance weighting model. Overestimation occurs mainly in the drier part of the Basin. Notably at the Archers Post gauging station and partly at Colcheccio. On the other hand the southern slopes of the Mt. Kenya are underestimated by the model, this is evident for the gauge at Castle Forest Station.

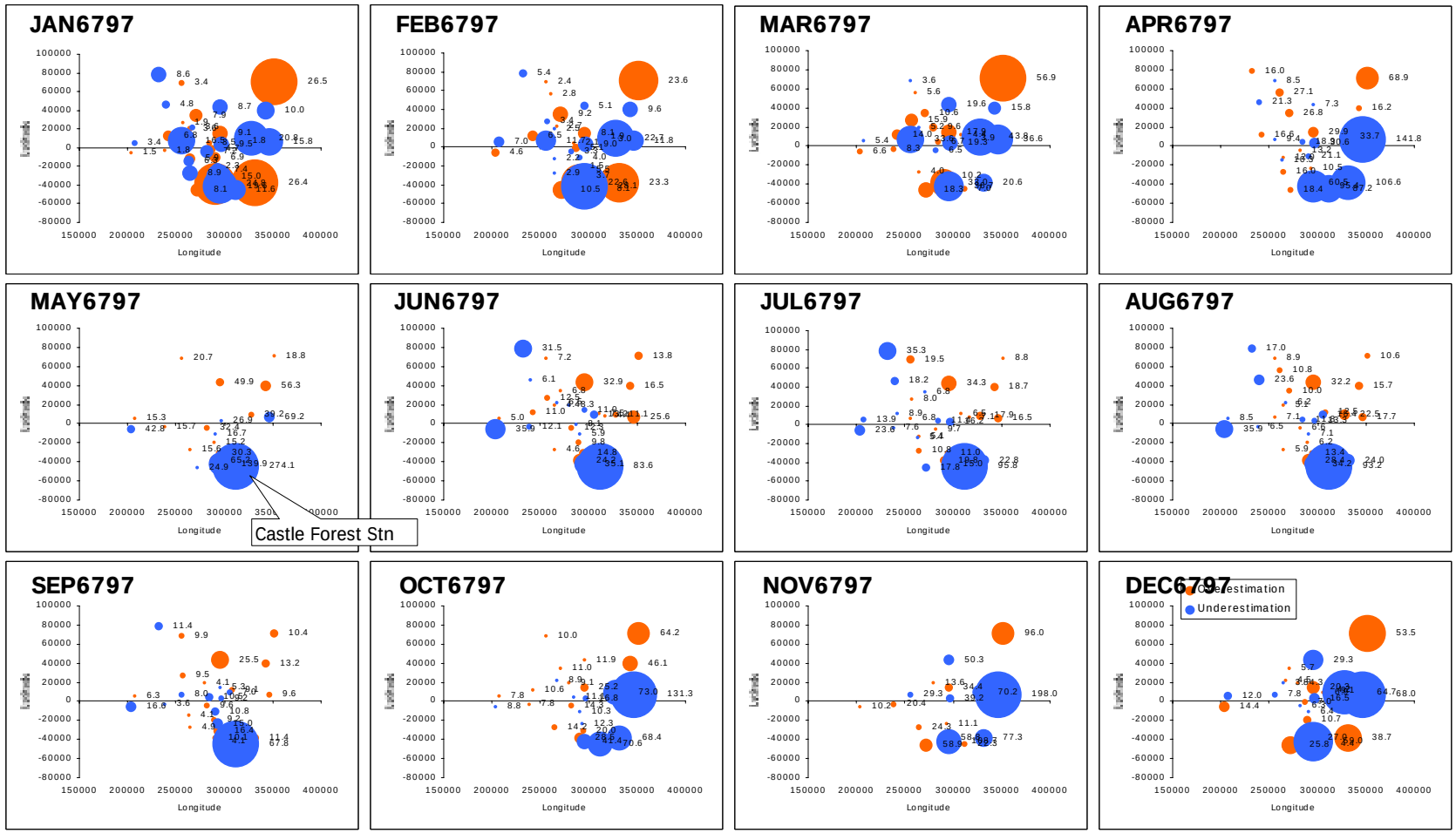


Diagram 5.9: Monthly geographical residual distribution of the inverse distance weighting estimation method (5 nearest neighbouring stations). Red circles mark overestimated values, blue represent underestimation. Note that the diagrams are not to scale to each other.

2. The second pattern can be seen by comparing the temporal development of the residual distribution: During months of spatially highly variable rainfall[32] there are also bigger differences in the estimated errors visible. This effect is very obvious for the month of May where the error for Castle Forest is -274.1mm. On the other hand we find a much more homogenous image for the drier months such as January, February, March and December where we note a maximum deviation of the residuals of 26.5mm for January and –96.6mm for March[33].

② Regularised Splines with Tension (RST)

For the regularised splines with tension model the interpolations made with tension set to a value of 44[34] was used to compare the spatial distribution of the residuals.
By studying the diagram on the next page the picture changes a bit. The overall impression is that the RST model produces much more homogenous estimation errors than the IDW method. Also we can no longer conclude that the method underestimates southern gauges and overestimates northern stations. Although we still find evidence of overestimated values at Archers Post, but the southern region is not dominated by underestimated residuals. As an example the estimations for the Irangi Forest Station are overestimated for most of the months.

The spatial distribution of the pattern does correspond with the pattern produced by the IDW method: gauges located in the centre of the Basin are calculated with a very small error, but the quality of the estimation degrades towards the edges of the area.
But, opposite to the IDW model, the pattern is more balanced: The worst error can be found for the Month of May at Irangi Forest Station with 149.1mm overestimation. The data points estimated through the RST model are by the order of magnitude of 2 better than the residuals produced by the IDW model.

[32] From April to November, where a distinctive precipitation pattern can be seen: higher rainfall in the Aberdare Mountains, and around the Mt. Kenya, but drier parts in the lowlands.
[33] Note that the overall precipitation pattern for March is already influenced by the onset of the 'long rains' that start according to Berger at the end of March (BERGER, 1989: 44).
[34] Parameters used: Tension=44, Smooth=0.05, Dnorm=disabled, Dmin=1

Diagram 5.10: Monthly geographical residual distribution of the regularised splines with tension method (tension=44, smooth=0.05, dnorm=disabled, dmin=1). Red circles mark overestimated values, blue represent underestimations. Note that the diagrams are not to scale to each other.

5.6 Discussion

As showed in the previous chapter (5.5 Analysis and Results) the both methods do perform differently and therefore the resulting estimates may lead to divergent precipitation maps. In this chapter we will discuss and compare the two methods. This discussion will be completed by reviewing some case studies of other spatial interpolation methods.

5.6.1 How well do they estimate?

Inverse distance weighting gave the best results where the 5 nearest neighbours were used to estimate the data points. We will therefore use the results of this run for the comparison of the two methods. The regularised splines with tension model proved to provide reliable estimates where the tension parameter was set to a value of 44 (tension=44, smooth=0.05, dnorm=disabled, dmin=1). The resulting summary statistics for both methods can be found in the following table:

Mean[35]	RST	IDW
n	40	40
Median	0.3666	1.2021
MAE	12.6998	15.6248
MSE	19.2790	25.9443
Correlation Coefficient	0.83	0.55
R^2	0.73	0.59
Best	**RST**	**IDW**
n	40	40
Median	0.0350 (AUG)	-0.0059 (JUL)
MAE	6.5547 (FEB)	7.0301 (FEB)
MSE	9.2450 (JAN)	10.0340 (FEB)
Correlation Coefficient	0.91 (DEC)	0.78 (FEB)
R^2	0.89 (OCT)	0.66 (FEB)
Worst	**RST**	**IDW**
n	40	40
Median	2.5461 (DEC)	3.3240 (NOV)
MAE	23.9393 (NOV)	27.9419 (MAY)
MSE	38.5694 (NOV)	54.6530 (MAY)
Correlation Coefficient	0.64 (SEP)	0.40 (SEP)
R^2	0.40 (SEP)	0.42 (SEP)

Table 5.7: Residual summary statistics for mean, best and worst cases for inverse distance weighting and regularised splines with tension method.

According to Table 5.7 we can expect a good representation of the actual measured rainfall by choosing the RST method as the correlation coefficient is at 0.83 and we have an explained variance of 73% through the linear regression line. On the other side the IDW method cannot model the rainfall as good as the spline under tension model: a

[35] This is the computed mean over all monthly summary statistics.

correlation coefficient of only 0.55 and an explained variance of 59% is not enough to trust this method. However, we see by comparing the worst with the best cases that the inverse distance weighting model is more consistent in its results as there is a smaller difference between the worst and the best cases. The difference of the explained variance between the worst and the best case is for the RST model 49% and for the IDW model only 24%.
When comparing the bias and the spread of the residual distributions, we still have to vote for the RST model, as both indicators (Mean absolute error and mean squared error) outperform the IDW model. In general both model tend to overestimate rainfall, the IDW model underestimates lower rainfall values, and overestimates higher values.

By analysing the mean distribution and the mean frequency distribution of the residuals we find further evidence that the RST model yields better estimates.

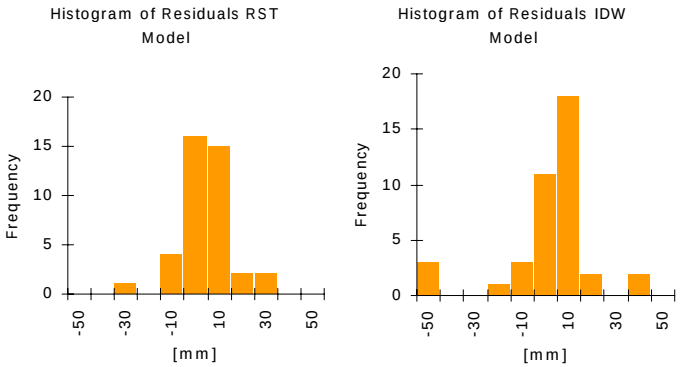
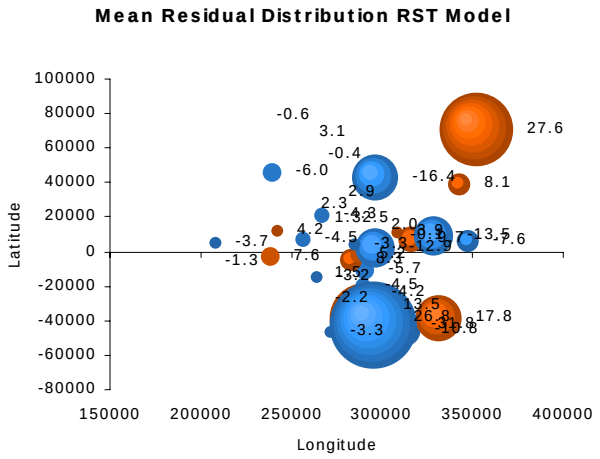


Diagram 5.11: Frequency distribution of the mean residuals of the IDW and RST model



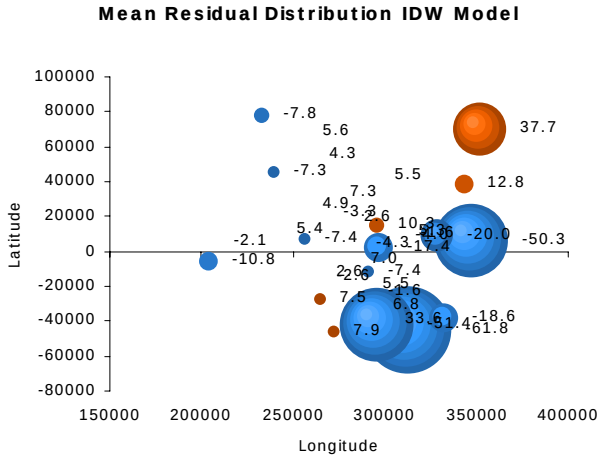


Diagram 5.12: Mean annual residual distribution for the IDW and RST model. (Note that the two diagrams are not to scale to each other!)

Diagram 5.11 underlines the better performance of the RST model as there is a much narrower spread and a far more balanced distribution compared to the IDW model.

Finally we will take a look at the scatter plots in order to analyse the quality of the regression line (Diagram 5.13).
With a correlation coefficient of 0.89 and an explained variance of 88% the regularised splines with tension model proofs a better estimation than the inverse distance weighted model with an explained variance of only 63%.
Please note, that the statistics provided by Table 5.7 (Residual summary statistics for mean, best and worst cases for inverse distance weighting and regularised splines with tension method.) are the mean values of monthly regression values, the regression lines presented in Diagram 5.13 (Scatterplots and regression analysis of the mean values from the two estimation methods.) are the regressions computed from the annual mean of the estimated and observed values.

Diagram 5.13: Scatterplots and regression analysis of the mean values from the two estimation methods.

5.6.2 Case Studies

This sub section will deal with the comparison of different estimation techniques. As the interpolation with splines is quite a modern method, not many comparative studies exist, and it is difficult to compare the results of the various studies with the results presented in this work. Climatic differences, altered estimation models, non homogenous data sets, and other differences makes it very hard to directly compare the studies with each other. Nevertheless a selection of comparative studies found in the literature are presented as case studies.

Case 1: Inverse Distance vs. Kriging

1985 Tabois and Salas published a comparative analysis of different spatial interpolation techniques[36]. They applied commonly used methods to 30 years of annual precipitation data measured at 29 stations located in the Region II[37] in the North Central Continental United States (TABIOS and SALAS, 1985: 372). The comparison included 12 different commonly used interpolation methods such as Thiessen Polygon Method, Inverse Distance, and kriging. They conclude that *"[...] the Kriging techniques are the best*

[36] Tabois, III G.Q and Salas, J.D., 1985: "A Comparative Analysis of Techniques for Spatial Interpolation of Precipitation" Water Resources Bulletin, 21, 3, 365-380

[37] An area of 52'000km² which is located between 96.0° and 99.5°W longitude and 40.0° and 42.5°N latitude.

among all techniques analyzed considering the various performance criteria used [...]" (TABIOS and SALAS, 1985: 379) The inverse distance method was found to be *"[...] significantly better than the Thiessen polygon method."* (TABIOS and SALA, 1985: 374) but does not produce as good results as the kriging model.
In their study they provide the coefficient of determination in percent for each interpolation technique. The best results were yielded with the kriging method which achieved a coefficient of determination of 82.7%[38] (Inverse Distance: 82.3%).

Case 2: 5 Methods
Isaaks and Srivastava compared 5 estimation models based on 780 data points[39]. They computed a correlation coefficient for the inverse distance method of 0.78 and for the ordinary kriging method of 0.82 (ISAAKS, 1989: 317). Whereas the triangulation and the polygonal model performed at a correlation coefficient of 0.80 respectively 0.69. Although ordinary kriging was designed to minimise the error variance, the *"[...] ordinary kriging estimates are also very good according to many other criteria."* (ISAAKS and SRIVASTAVA, 1989: 319)

Case 3: Thin Plate Spline Method
During the SIC97 (Spatial Interpolation Comparison 1997) in Switzerland, different interpolation methods were examined by the participants. Hutchinson used a thin plate smoothing splines model in order to estimate 367 rainfall data points in Switzerland[40]. He achieved a correlation of 0.87. However, better results were achieved by using the square roots of the data points (correlation of 0.89), and Hutchinson therefore concluded that *"The thin plate spline model, applied to the square roots of the observed rainfalls, has provided point estimates which show good agreement with the withheld data, as measured by both summary comparative statistics and order statistics. The estimated values are consistent with the standard error analysis"* (HUTCHINSON, 1998)

Case 4: Splines vs. Kriging
A direct comparison of results produced by a spline algorithm with the kriging algorithm was also presented at the SIC97 by Anatoly Saveliev, Svetlana Mucharamova, and Gennady

[38] Mean of coefficients of determination for each technique (TABOIS and SALAS, 1985: 376)
[39] Edward H. Isaaks and R. Mohan Srivastava, 1989: "An Introduction to Applied Geostatistics". Oxford University Press, New York, 1989.
[40] Michael F. Hutchinson, 1998: "Interpolation of Rainfall Data with Thin Plate Smoothing Splines: I Two Dimensional Smoothing of Data with Short Range Correlation". Journal of Geographic Information and Decision Analysis, vol. 2, no. 2, pp. 152-167

Piliugin[41]. Unfortunately they did not present exhaustive regression statistics, but we can compare the median of the residual statistics. Both methods are equally good in estimating the missing 367 data points by estimating a median of 2.2 (surface under tension) and 2.3 (kriging). The authors conclude in the discussion of their results that *"[...] the magnitude of the estimated values obtained by the first method **[surface under tension]** is wider than that of the source values, whereas the second method **[kriging]**, as common for kriging methods, is more "conservative" and tends to "average" estimated values."*[42] (SAVELIEV et al., 1998)

Case 5: Simple Good Sense Prediction

An interesting approach was also presented at the SIC97: the interpolation of rainfall by simple good sense by Marc G. Genton and Reinhard Furrer[43].

Because of the different models and variations of statistical methods, the authors decided to compare spatial estimates based on the simple good sense of a human with the results of a statistical approach. They used their subjective knowledge of the topography and of the micro climatic conditions to estimate the 367 rainfall points.

The calculated root mean squared error is RMSE=72, compared with the results of other participants (Saveliev: RMSE=71 for the kriging model, Alaa: RMSE=64 for adaptive kernel estimation[44]) this is quite a remarkable result, however the authors conclude that *"[...] the results of this method **[estimation by simple good sense]** are worse than those from a robust statistical methodology used by the authors in another analysis. This would lead to support that spatial statistics are worth the trouble."*[45] (GENTON and FURRER, 1998)

5.6.3 And the Winner is...

At the end of every competition there must be a winner. Fortunately the competitors produced relatively clear results, which makes our decision for a winner easier. By taking the presented case studies into account we would have to decide between a kriging method or a spline related approach. The two methods are quite similar in the ac-

[41] Anatoly A. Saveliev, Svetlana S. Mucharamova, Gennady A. Piliugin, 1998: "Modeling of the Daily Rainfall Values Using Surface Under Tension and Kriging". Journal of Geographic Information and Decision Analysis, vol. 2, no. 2, pp. 58-71.
[42] Bold remarks in brackets by the author (Bernhard Sturm)
[43] Marc G. Genton, Reinhard Furrer, 1998: "Analysis of Rainfall Data by Simple Good Sense: Is Spatial Statistics Worth the Trouble?". Journal of Geographic Information and Decision Analysis, vol. 2, no. 2, pp. 11-17.
[44] Alaa Ali, 1998: "Nonparametric Spatial Rainfall Characterization Using Adaptive Kernel Estimator". Journal of Geographic Information and Decision Analysis, vol. 2, no. 2, pp. 34-47.

curacy of their results (Hutchinson, 1999) but the application of them implies some critical differences: kriging is bound to a pre-calibration of a variogram with three parameters – range, nugget and sill. And as Hutchinson puts it: *"it can only offer empirical suggestions for the form that the spatial covariance should take."* (HUTCHINSON, 1999).

By using a spline model one doesn't need to estimate a spatial covariance prior to the interpolation process. The only parameter that has to be estimated is the smoothing or tension parameter that determines an optimal balance between fidelity to the data and smoothness of the fitted spline function. This parameter is normally estimated automatically by minimising the generalised cross validation (GCV). The GCV can be calculated quite efficiently (Hutchinson, 1999). Hutchinson therefore sees the main *"[...] advantage of splines is their operational simplicity."* (HUTCHINSON, 1999).

The correlation Hutchinson computed for the thin plate spline models presented at the SIC97 (Case 3) of 0.87 (for the non-transformed data points) is close to the correlation coefficient of 0.89 calculated for the regularised splines with tension used for this work.

Although the inverse distance weighted method is quite widely used as an interpolation model, we will not favour this method for the application presented in this paper. This is mainly due to its poorer ability to estimate precipitation (correlation coefficient of 0.52) and because of its fairly wider spread and bias of the error distribution. In the literature the inverse distance methods and its derivatives were classified of producing results similar to the Thiessen polygon methods[46], but are generally inferior to a kriging method (Tabios and Salas, 1985).

Therefore the winner of the cross validation is the regularised splines with tension method with the following parameter configuration[47]:

Tension	Smooth	Dnorm	Dmin	Npmin	Segmax
44	0.05	N/A	1	700	700

Table 5.8: And the winner is... The parameter setting for the regularised with splines model with the best estimation performance.

[45] Bold remarks in brackets by the author (Bernhard Sturm)

[46] Tabios and Salas conclude in their study that the inverse distance method gave smaller errors of interpolation than the Thiessen model. (Tabios and Salas, 1985)

[47] As showed in chapter 5.5 (Analysis and Results)

6 Interpolation: Step-by-Step

6 The Interpolation: Step-by-Step

A goal of the presented work was to provide a flexible tool for the spatial interpolation of rainfall. As we have seen in the previous chapters (chapter 5 Cross Validation, page 63) that there is no such model that can be labelled as the 'absolute best model', we have to weight the pros and cons against each other, and then choose the method that will serve us the best for the given situation.
However, this chapter will present one way how to calculate and draw precipitation distribution maps. This step-by-step explanation is targeted to those who need to get a quick overview over the rainfall situation in a particular area. The following points may form the border conditions of such a task:

- Incomplete rainfall series[48]
- Limited budget situation (in terms of time and money)
- The need for a quick overview over the rainfall situation in a region by analysing the given rainfall for different time intervals such as 10-day or monthly periods.
- Only limited computer power available

The above list shows that the method suggested in this chapter is tailored to 'in-the-field' situations, where one has only limited computer resources available. The method can also be used in developing countries where know-how and manpower is abundant, but money restrictions often put a limit to the application of such extensive GIS modelling tasks.

Before we start with the step-by-step introduction, we need to take a look what we need in order to get a precipitation map. By discussing this we need to be aware of, that there are two issues that we have to carefully examine before we start:

- The quality of the available raw data
- The technical requirements

6.1 The Quality of the Raw Data

As the quality of any interpolated map directly depends on the quality of the used raw data, we have to examine the quality of our observed precipitation data.

[48] Observed rainfall time series consisting of a certain amount of gaps.

This involves quality control measures at any stage of the rainfall data acquisition process, as well as during the analysis and post-processing of the collected rainfall data. This issue is discussed in chapter 2.3.2 (Sources of Error) on page 29.
However, it is inevitable that the data basis consists of a number of errors such as systematic and random errors[49] (HERWEG and OSTROWSKI, 1997: 2). Some of these errors can be eliminated or shown by applying standardised checks on the data with homogeneity tests for precipitation data like the Standard Normal Homogeneity Test (SNHT) developed by Andersson (ANDERSSON: 1985)[50].

6.2 The Technical Requirements

The technical requirements can be subdivided into software and hardware requirements. We will start with the hardware requirements:

6.2.1 Hardware

For the presented work all models were computed on a low-cost, low-end Personal Computer consisting of the following main parts:

- AMD K6-2 266Mhz CPU
- 64MB RAM
- 2 Hard-discs: 2GB/4GB
- Internal CD-ROM (4x speed)
- 16MB Video Board
- 17" CRT Monitor

These hardware requirements show that the presented method and models can be calculated in sufficient time on almost any notebook or computer available today.
The mentioned two hard-discs in the above list are only necessary because the computer was set up with two different operating systems (Linux/Windows98), but it is not a need to have both OS on different hard-discs installed.

6.2.2 Software

On the software side the following applications were used:

- Operating System 1: Linux (SuSE 7.1 / Kernel 2.4)
- Operating System 2: Microsoft Windows 98SE
- GIS System: GRASS 5beta11
- Database: Microsoft Access97
- Statistics: Microsoft Excel97
- Graphic Tools: NVIZ, GIMP (Linux), Micrografx Designer 7 (Windows)

When studying the above list we can see that the main tool used for the interpolation is

[49] Following the error theory we need also to take care of parameter estimation errors (methodical and application errors) as well as model errors (structural and process errors). (HERWEG and OSTROWSKI, 1997)
[50] This test was also applied for this work in order to verify the consistency of the sampling gauges. A VBA module has been developed for Access97/2000 to provide an automated version of this test.

6

the GIS-system GRASS Geographical Resources Analysis Support System) which is under the GPL (GNU General Public License), released and hence freely available. GRASS was designed to run under a UNIX/Linux environment, but it is also available for WindowsNT/2000 the Apple Macintosh operating system, and the SunOS system.

6.2.3 And How Much does it Cost?

The overall costs may vary, but generally speaking are very low compared to other GIS systems. In most office environments the Microsoft Windows Operating System and the Office97 or Office2000 suite are already available. The quite powerful Micrografx Graphics package (Version 7 from 1998) can be bought for US$ 20.00 (Spring 2001)). Because most of the software (GRASS, GIMP, NVIZ) runs under the open-source[51] operating system Linux, it is freely available.
The costs invested in the hardware and software (excluding manpower costs) for the presented work is as follows:

Hardware	**US$**
AMD K6-2 Computer	500.00
17" Monitor	200.00
Total:	**700.00**

Software	**US$**
Microsoft Windows 98SE (upgrade)	100.00
Microsoft Office Professional (upgrade)	400.00
Micrografx Graphics Suite 7	20.00
SuSE Linux 7.1 Professional Package	60.00
GRASS5beta11	0.00
GIMP (included in the SuSE package)	0.00
NVIZ (included in the GRASS package)	0.00
Total:	**580.00**
Grand Total:	**1280.00**

Table 6.1: Hardware and Software costs.

US$ 1280.00 for a full featured GIS system is not a lot of money, and as Richard Shepard, Director of Applied Ecosystem Services in Troutdale (Oregon, USA), states that one could get the *"[...] comparable analytic capabilities from ARC/Info and some of the expensive add-ons to it, but you'd spend >$30'000 for a single-user license on NT."* (Dr. Richard B. Shepard, GRASSLIST Thread No. 1665 (grasslist@baylor.edu), 23.3.01)
However, despite the low costs one should not forget that setting up a Linux based system can be a time consuming task, and one

[51] The open-source system depicts a way of developing and distributing software, protected by the GPL License, by publishing the source code in order to enable anybody to freely adapt change develop, and distribute the software for his/her own needs. The term 'free' has to be understood in the sense of 'freedom'. (See also: http://www.gnu.org/copyleft/gpl.html)

should be prepared to face the few hurdles which are to be taken for the successful use of such a high-performance operating system.

6.2.4 Why the Use of GRASS?

There is one widespread GIS that is being used by many authorities and institutions: ARC/INFO by ESRI. But ARC/INFO represents an expensive commercial GIS where each installed package or even additional module needs to be licensed, and it was one goal for the presented work to provide a simple and in-expensive method for interpolating rainfall. There is one alternative GIS to ARC/INFO: GRASS.

GRASS (Geographical Resources Analysis Support System) is a hybrid GIS, which enables users to analyse geographical data (organised as raster or vector data) with an underlying database structure. GRASS is a modular GIS. Each module is treated independently, and adds therefore transparency and structure to the daily work of data analysis with GRASS.
According to Baylor University there were around 40'000-45'000 users world wide using GRASS[52] in the year 2000.
Another argument for the use of open-source software is that less developed countries get a chance to gain independence from the western it-industry. This independence must not only be seen in terms of money but also in terms of knowledge. With open-source software, such as GRASS, less developed countries will be able to produce and develop their own software without being hooked to the dominant and expensive western software technology. According to Nazir Perez, TU-University Berlin, open-source software needs urgently to be propagated by development agencies and NGOs to achieve a balance of knowledge between industrialised countries and developing countries (GROTE: 2001, 106).
For instance, in Ernakulam, India, GRASS was used together with a relational database (PostgreSQL) to build a resource management database on a grassroots level thus enabling rural authorities to use and maintain the system as a decision making tool[53] (PRATHAPAN, 2000).
Major companies, such as Kenya Airways and the Power & Lighting company in Kenya and other East African countries have already realised, that the open-source operating system Linux offers them a great deal of freedom and stability, and have adopted the

[52] According to personal communication with Dr. Richard Shepard, Applied Ecosystem Services Inc., USA. (email from GRASS maillist 23.3.2001)

[53] This has in fact even cultural implications as Prathapan puts it: "Another major success of the project has been the successful development of a Malayalam phonetic keyboard overlay for GUI in Linux environment" (PRATHAPAN, 2000)

system for the benefit of their own needs (GROTE: 2001, 105).

6.2.5 The History of GRASS

Initially GRASS was developed by the U.S. Army corps of Engineers (CERL: Construction Engineering Research Lab) for the purpose of military use. Since the end of the eighties in the last century, GRASS was made available to the public. GRASS is a typical 'child' of the Internet generation: the emerging World Wide Web made it available to anybody who was looking for a free GIS system. Because GRASS was not perfect at all, many researchers around the globe began to add new and specialised modules to the GRASS package. Since 1994 CERL stopped the support and development of GRASS. In 1997 the Baylor University of Texas, USA, began to steer the development of the free GIS.

Since 1999 two GRASS headquarters exist: a very strong European branch, lead by Markus Neteler, University of Hannover, Germany and the Center for Applied Geographic and Spatial Research at the University of Baylor, Texas[54]. These two headquarters co-ordinate the worldwide development of GRASS. The source code for GRASS and compiled binaries are available from the two main servers:

- http://grass.baylor.edu/~grass
- http://www.geog.uni-hannover.de/grass/

In 2000 GRASS was put under the GPL (GNU General Public License), in order to prevent any commercial, and hence payable, distribution of GRASS.
GRASS is widely used in the United States, Europe, South America, and Asia. Two main servers in Europe and the USA host the current developing versions. An excellent support and help system via the GRASS mailing-list, provides fast and competent[55] help through the Internet.

6.2.6 Modules

The GRASS modules can be used for the analysis of:

- Point data (site)
- Vector data
- Raster data

[54] GRASS Development Team
Institute of Physical Geography and Landscape Ecology at University of Hannover, Germany; Schneiderberg 50; 30167 Hannover; Germany
Center for Applied Geographic and Spatial Research Baylor University; P.O. Box 97351; Waco, Texas 76798-7351; General Email: grass@baylor.edu

[55] This mailing-list is used by all users and developers of GRASS. In many cases a problem with GRASS leads therefore directly to a fixed module. Because of this tight link between users and developers GRASS evolves constantly and much quicker than any other commercial GIS.

There are hundreds of specialised modules and small tools, which provide support in the realms of:

- Image Enhancement (Geo- and radiometric rectification; Filtering; Re-sampling; Orthophoto; Classification)
- Calculation of/with digital elevation models
- Digitising/Vectoring of maps
- 3D animated visualisation (fly-by)
- Interface to databases (Postgres)
- Erosion modelling
- Hydrological modelling (run-off models)
- Environmental structure analysis
- Import/Export Modules (ARC/INFO, Erdas, ASCII, Idrisi)

Despite all these modules GRASS has one special discipline: remote sensing image analysis. When it comes to remote sensing, GRASS offers possibilities like hardly any other GIS (NETELER, 1998: 11)
Because of its modular and open character GRASS can easily be adapted to the users need. For this work a variety of Perl and Shell scripts were written in order to automate the GIS analysis workflow.
The GRASS Development Team is currently working on a complete 3D capability of GRASS which will users give the ability to work fully in a 3D environment, something that does not exist in any other GIS package. Users will then be able to work on raster elevation data as well as on vector and sites data in the 3D environment[56].
All this makes GRASS an ideal and very powerful GIS for the 'out-in-the-field' use and for small local offices in developing countries as well as developed countries.

6.2.7 Is GRASS worth the Investment?

The answer to this question depends strongly on the aims one plans to achieve with a GIS application. In terms of money, GRASS is more than worth the investment (because GRASS is a free GIS). Concerning invested manpower, the answer whether GRASS is worth the time, depends on the skill and knowledge of the users of GRASS. GRASS is primarily a command line controlled GIS (like ARC/Info), however since version 4.x GRASS can also be run under the Tcl/TK GUI environment. This offers a very user friendly, windows like drop-down menu styled interface. As stated in chapter 6.2.6 (Modules, page 93) new modules and functions can easily added to GRASS, and this is one of the strongest points of this GIS. GRASS could even be used to develop own commercial GIS modules, as

[56] In the current beta versions of GRASS some of these 3D capabilities are already included.

Shepard puts it: *"*[a company in Boulder, Colorado has]*... adopted GRASS* [...]. *They've modified it themselves and have become experts with it. They sell spatial analyses and visualization based on GRASS to developers and they are making a ton of money. Because they can do what others cannot do. This company doesn't care where GRASS is going because they have what they need and they have the ability to make it into whatever the need in the future, too."* (Dr. Richard B. Shepard, GRASSLIST Thread No. 1665 (grasslist@baylor.edu), 23.3.01)

To the issue whether GRASS is a sustainable investment, Richard Shepard mentions the long tradition of GRASS: *"GRASS has been around since the late 1980s, and I expect it will be around longer than I. Why not? You get the source code and can do with it as you wish."* (Dr. Richard B. Shepard, GRASSLIST Thread No. 1665 (grasslist@baylor.edu), 23.3.01).

6.3 From Raw data to Gap-Free Data

After having discussed the technical requirements for a spatial interpolation, we will move on to the first step: the elimination of missing data in the raw dataset.

Starting point are the stored precipitation data in the NRM-database. As described in the chapter 2.6 (Quality Control of Raw Data) on page 38, we assume these data as being quality controlled. All precipitation values are organised and stored in a single database table in the following format:

StationID	RDate	RAmount	Flag
1	02.02.57	0.00	

Table 6.2: Example for the table 'Rainfall' in the NRM-database

The field [StationID] contains a reference ID-number to the description of the rain gauge, where [RDate] is the date of the observation, and [RAmount] contains the observed value in mm for this date. The field [Flag] is used to mark uncertainties, interpolated or averaged values.

As discussed in chapter 3.3 (Gaps Infilling) on page 46, it is inevitable that the stored values are incomplete and therefore contain gaps. Any null values in the NRM database are treated as 'gaps'. Because spatial interpolation is based on complete and homogenous data series a first task will be to eliminate any gaps from this raw data sets.

For this purpose linear regression according to a method described by Lindsay McMillan (MCMILLAN, 1998) was used (for a detailed discussion refer to chapter 3.3.1 (Methodology) on page 47.

Before a surface interpolation within GRASS can be performed, the gap free rainfall values need to be grouped into groups of the same time period like, for example, the same

10-day period. The whole workflow is shown in Diagram 6.1: Workflow for the interpolation of gaps.

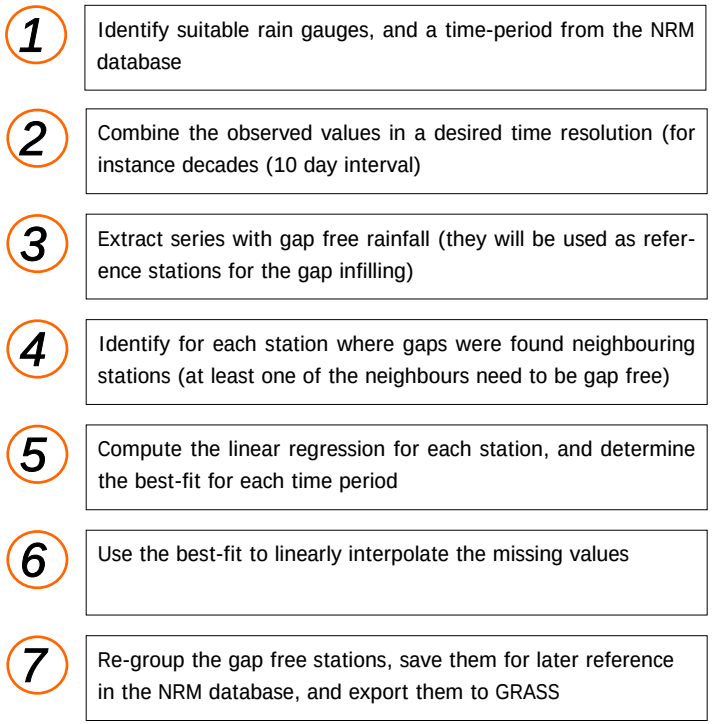


Diagram 6.1: Workflow for the interpolation of gaps

At each stage of this workflow semi- or fully automated tools will help the user to perform the needed tasks.

6.3.1 Step 1: Identify Suitable Rain Gauges

Depending on the time period and the type of rainfall map we will want to draw, we have to select the rain gauges (e.g. rainfall measurement points), which suit best for our needs. In most cases we will try to get as many as potential stations as possible for the interpolation, and these stations should be of the best available quality[57], and when analysing longer time periods, we will certainly try to make sure all selected stations cover the widest possible time span. Obviously we cannot always optimise these requirements, in most cases we will have to negotiate between them and find a common ground for the set of stations we want to analyse.

For the presented study two different sets of analysis were performed[58]:

- A long-term analysis (30 years) with decade data resolution (10 day periods)

[57] The quality of a station is also stored in the NRM database, and was determined during the fieldtrip in Kenya in 1998. However, quality does not only include quality in terms of the robustness or the overall physical state of the gauge, but also the geographical location of the observation point. Depending on the interpolation to be performed, one may consider certain locations (in relation to other selected gauges) as less favourable and hence assessing such gauges as of lower quality.

- A short-term analysis (8 years) with decade data resolution (10 day periods)

Analysis	Years	Resoulution	Stations n	Gaps
Long-Term	31 (1967-1998)	10 days	40	28 Stations
Short-Term	9 (1990-1998)	10 days	63	24 Stations

Table 6.3: Summary of the two used data set in this study.

Table 6.3 shows that we have to trade-off between long time periods, gaps and the number of stations used for the interpolation. The amount of gaps found per station can be a crucial criterion for the grade of quality of the resulting precipitation maps. However, the used linear regression interpolation method proofed to be robust enough to interpolate even complete missing years (see Chapter 3.3.2 (Quality of Interpolated Gaps) on page 51, for a detailed discussion). Despite the fact that the presented interpolation software for this study is able to automatically interpolate such long periods of missing values, extreme care should be taken when trying to do so.

6.3.2 Step 2: Collecting Precipitation for a Desired Time Resoultion

Chapter 3.2 (Selecting Sampling Locations and Time Span) on page 43, discussed the importance and implications of selecting a particular time resolution. Depending on the station density a too small time resolution will result in a too high standard deviation of the error when determining the mean precipitation (WMO, 1985: 16-17). For the Ewaso Ng'iro Basin a station density of 656.25km^2 and 423.39km^2 (for 40 stations and 63 stations) can be achieved, which allows, after the World Meteorological Organisation, a 10-day precipitation period with a standard deviation of the error of about 22%-25% (WCP-100 Report: WMO/TD-No. 115, 1985: 16).
After selecting the optimal set of gauges, we can start to collect the associated observed rainfall values:

[58] For a complete description of the used gauges refer to chapter 10 (Rainfall Gauges) on page 159.

Group the Precipitation in an Interval

1. Open the frmIntervalQuery in the NRM database or use the frmIntervalQuery provided on the data CD-ROM of this study:

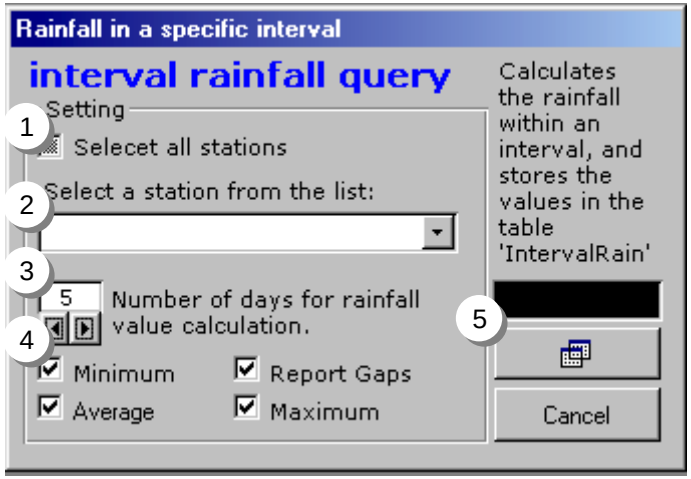


Figure 6.1: Interface of the Interval Rainfall Query dialog in the NRM access database.

1) This will select all stations found in the NRM database.
2) Or select one station from the list provided.
3) Enter a desired interval (in days)
4) It is possible to calculate minimum, maximum, average and the number of gaps found in the specified interval. (It is recommended to leave all options ticked).
5) Start the calculations by pressing this button.

The results of the interval calculation are being stored in a new table 'tblIntervalRain' in the NRM database.

Save the tblIntervalRain as an EXCEL spreadsheet
Copy and Paste the contents of the 'tblIntervalRain' from the Access database to an empty EXCEL spreadsheet:

Station	FromDate	ToDate	Rainfall	Mean	Max	Min	Gaps
ARCHERS POST	01.01.90	10.01.90	0	0	0	0	0
ARCHERS POST	11.01.90	20.01.90	2	0.2	2	0	0
ARCHERS POST	21.01.90	31.01.90	0	0	0	0	0
ARCHERS POST	01.02.90	10.02.90	6.9	0.69	4.9	0	0
ARCHERS POST	11.02.90	20.02.90	27.1	2.71	27.1	0	0

Table 6.4: Example of the first entries of the new EXCEL spreadsheet containing the rainfall in the specified interval.

The column [Gaps] lists the number of gaps found in the specified interval (10 days in the above example).
Note: Before you copy the table to the EXCEL sheet make sure that the dates are sorted by the field [FromDate] in ascending order!

1. Mark the Date columns in the EXCEL sheet and format them to the date format of 'DD.MM.YY'.

2. Save the EXCEL spreadsheet according to its station name (e.g.: 'ARCHERS POST.xls').
3. Store all stations (you will get for each station a single EXCEL spreadsheet) in a directory labelled 'RawData'. Make sure that you have the following directory structure on your hard disc:

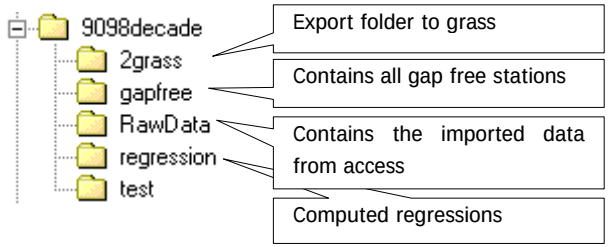


Figure 6.2: Suggested directory structure

After that all stations are stored as EXCEL spreadsheets in the folder 'RawData', and we will continue with the next step:

6.3.3 Step 3: Identify Neighbouring Stations and Group Them

In a first step all EXCEL spreadsheets with stations that are gap-free (you can easily check this by calculating the sum on all values found in the column [Gaps]), are to be moved to the folder 'gapfree'. These stations represent reference gauges; we will need them to compute the missing values for the other gauges.

Identification of neighbouring stations

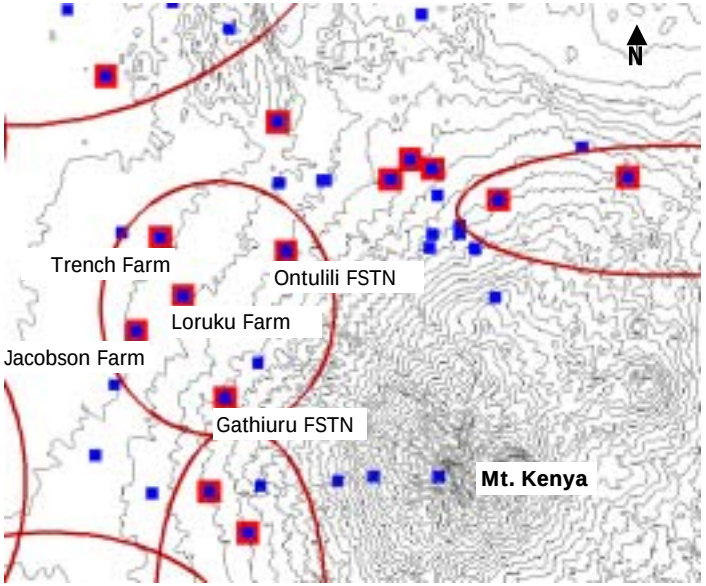
This is a crucial part of the interpolation process, as the applied method will use the values of surrounding stations in order to fill in the gaps of a particular station, a not appropriate selection will result in unreliable results and low correlation coefficients. Stations in the neighbourhood of a gap-station must therefore lie within the same or similar precipitation regime. By analysing the time-series of the precipitation one can identify such similar regimes (see chapter 7.1.3 (Annual Precipitation Patterns) on page on page 128).

Care should be taken that at least one of the selected station shows a complete time series.

For each station containing gaps, a set of neighbours has to be defined. The quality of the infilled gaps depends also on the amount of used neighbouring stations, but as stated above, if they belong to non-related rainfall regimes the quality will degrade.

For the purpose of quality control it is best to create a control EXCEL spreadsheet where all gauges are listed and the grouped neighbouring stations are marked. (see Figure 6.3: Excerpt from an example of a station control list)

Figure 6.3: Excerpt from an example of a station control list



Key

- Gauges stored in the NRM3 database
- Gauges used for the time series from 1967-1998 (10 day interval)
- 100m altitude contours
- Neighbourhood areas containing gauges used for the regression anlysis.

Figure 6.4: Areas of neighbouring rain gauges in the vicinity of Mt. Kenya

If an analysis covers only a short period of time it is possible that the time series of the incomplete stations may miss entire time intervals[59]. Because the algorithm used for the regression analysis will rely on complete time series it is necessary to produce a so-called 'period reference' file (Figure 6.5). This 'period reference' file is a dummy station, which contains the complete time series to be used for all other stations. For instance if an analysis covers the period from the 1.1.1990 to the 31.12.1998 (in a 10 day interval), a 'period reference' file containing all intervals for this period has to be produced. To activate the 'period reference' file, it needs to contain nothing but gaps! This is necessary to make sure that the regression analysis algorithm is synchronised with the complete time series.

[59] This is common as it is very unlikely that an occurrence of a gap is a singular event. In most cases if gaps occur they will consecutively cover several dates, such as an entire month. As a consequence this month may not be included in the extracted data from the database, and will not be available for any following regression analysis.

6

	Station	FromDate	ToDate	Rainfall	Meas	Max	Min	Gaps
2	PERIOD REFERENCE	01.01.90	10.01.90	0	0	0	0	1
3	PERIOD REFERENCE	11.01.90	20.01.90	0	0	0	0	1
4	PERIOD REFERENCE	21.01.90	31.01.90	0	0	0	0	1
5	PERIOD REFERENCE	01.02.90	10.02.90	0	0	0	0	1
6	PERIOD REFERENCE	11.02.90	20.02.90	0	0	0	0	1
7	PERIOD REFERENCE	21.02.90	28.02.90	0	0	0	0	1
8	PERIOD REFERENCE	01.03.90	10.03.90	0	0	0	0	1

Figure 6.5: First entries of a 'period reference' EXCEL spreadsheet (PERIODREFERENCE.XLS) (note that each interval contains one gap!). This reference file is only needed to ensure a continuous time series.

Start the [Collect Data] Macro

For the regression analysis we need to collect the observed rainfall from all stations in the same period. For example for the period 1-1 (1.1.xx-10.01.xx) all data from all stations within the same neighbourhood must be collected and grouped in this period. This will result in such as list:

Time	JACOBSON FARM	NICOLSON FARM	MATANYA (NRM)	PERIOD REFERENCE	GATHIURU FORESTSTN
01.01.90	7.6	23.8	25.9		0
01.01.91	0	0	12.4		20.5
01.01.92	0	3.5	4.2		12.1
01.01.93	92		82.6		62.9
01.01.94	0		0		2.6
01.01.95	0	0	1.5		0
01.01.96	23.876	4.6	12.2		18
01.01.97	2.032	3.4	3.8		20.6
01.01.98	73.152	80.3	132.1		100.8

Table 6.5: Collected and grouped rainfall data ready to be analysed with linear regression. Note the gaps for NICOLSON FARM in 1993 and 1994. The 'PERIOD REFERENCE' column is intentionally left blank to ensure time series integrity.

Table 6.5 is an example of a table produced by the macro [Collect Data], which will be used in this step. Please note that the dummy station 'PERIOD REFERENCE' will not be used for the later regression analysis, and the 'PERIOD REFERENCE' column is therefore left blank.

1. Open the EXCEL file 'GapCollect.xls'. This file contains all the necessary macros for the automatic gap infilling process.
2. To run CollectData open the EXCEL files of all neighbouring stations (including the station you want to complete the gaps) and the PERIODREFERENCE.xls. CollectData will then analyse the gaps contained in these files, and group them in one single EXCEL worksheet organised by intervals (each sheet in a EXCEL file represents an interval). This file is named in the form of XXX_YYY_ZZZ.XLS (XXX = Rainfall Station 1, YYY=Rainfall Station 2, ZZZ=Rainfall Station 3[60]).

[60] It is no problem to denote more than 3 stations the proposed naming can be extend to any number of stations in the form of: ST1_ST2_ST3_ST4_STn.XLS.

3. After you have opened all the necessary files, press the [CollectData] button and the following window will open:

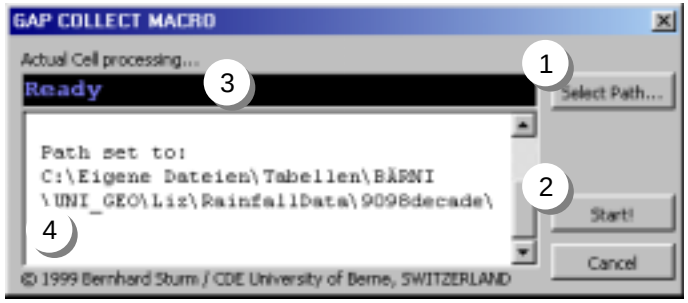


Figure 6.6: Collect Data dialog

   1) Sets the path where the temporary result files are stored
   2) Initiates the collect algorithm
   3) Status line: Provides information about the current cell that is being processed by the macro.
   4) Log window: This window logs all completed tasks, the content can be copied with copy + paste for later reference.

4. The completion of the macro will take a while. The progress of the macro is displayed in the status line, and the log window of the macro.
5. When finished the macro can be closed (press the [Cancel] button) and all used EXCEL files (including the temporary 'GapCollection' EXCEL spreadsheet), but except the new file can be closed without saving.
6. Save the resulting EXCEL spreadsheet in the folder 'regression' according to this example:

   XXX_YYY_ZZZ_A.XLS

   Where:
   XXX,YYY,ZZZ denote to the first 3 letters of the collected stations.
   A is a reference to the number of the neighbourhood area.

7. Repeat this process from step 2 onwards for all non-gap free stations.

6.3.4 Step 4: The Regression Analysis

To run the regression analysis, step 3 (chapter 6.3.3 Step 3: Identify Neighbouring Stations and Group Them, page: 99) must be completed, and all collected stations must be organised in EXCEL spreadsheets, and saved in the folder 'regression':

6

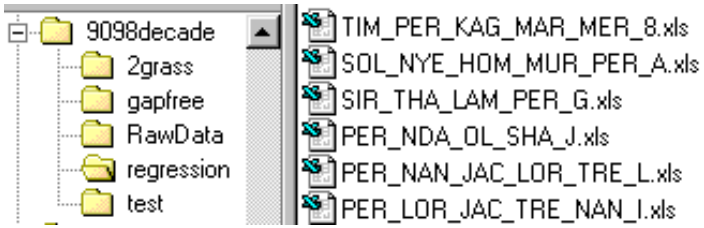


Figure 6.7: Content of the folder regression after the Collect Data macro was run.

Step 4 will analyse each interval with the help of the linear regression method in order to identify the station that represents most accurate the incomplete station. This is achieved by computing the regression coefficient for each interval[61]. For a detailed discussion about the used infilling model refer to chapter 3.3.1 (Methodology) on page 47.
The macro 'Regression Analysis' can be run in a full automatic mode in which the macro computes all necessary regressions and fills the missing values (gaps) automatically according to the best-fit model. In practice this proofed to be a very reliable mode, but it is strongly suggested to check the used regression coefficients manually, as it is always possible that the identified best fits by the macro are erroneous, and hence need to be adjusted. The macro is designed to use these manual adjusts instead of the pre-computed coefficients. Manual adjustments can take place at a defined stage of the regression analysis. For later reference all manipulations are stored in the EXCEL files produced at step 3.

[61] Hence each time interval can have different 'best-fit' neighbouring stations.

Start the [Start Analysis] Macro

1. Press the [Start Analysis] button in the macro 'GapCollect.xls' EXCEL macro workbook:

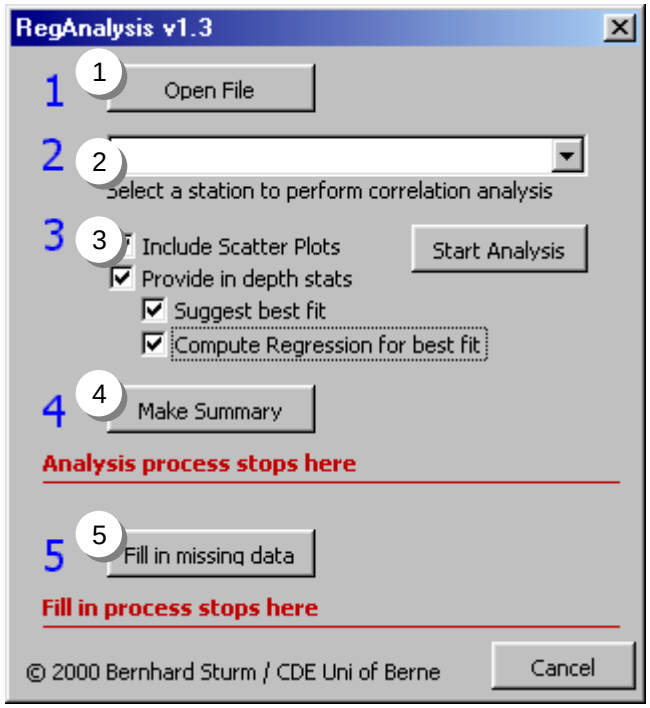


Figure 6.8: Regression Analysis dialog

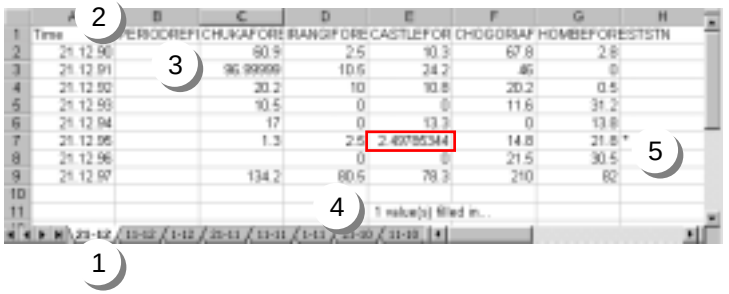
   1) Select the file on which the regression analysis shall be performed. The file stems from step 3.
   2) Choose the station on which you would like to perform the gap infill process.
   3) Tick the needed statistical parameters that shall be computed. Disabling the 'Include Scatter Plots' option will make the calculation faster. Also note that with Office97 you will get errors when using this option[62]. For a useful analysis tick at least 'Provide in depth stats', 'Suggest best fit', and 'Compute Regression for best fit'.
   4) Pressing this button will add another worksheet to the workbook. This worksheet contains the best fits and their respective regression coefficients for all examined intervals.
   5) Fills the gaps according to the marked best fits in the regression workbooks.

2. After finishing step 4 the regression analysis process can be stopped in order to manually check the computed statistics

[62] This problem is recognised by Microsoft, and is related to the fact that EXCEL97 (service release 1) contains errors, which do not allow EXCEL to add more than 124 charts to a workbook. In order to fix this, Microsoft has released the Service Release 2 package. However, the error seems not to disappear. Microsoft suggests correcting manually the registry key HKEY_CURRENT_USER\Software\Microsoft\Office\8.0\Excel\Microsoft Excel\AutoChartFontScaling. After editing this key EXCEL97 is able to process more than 124 charts in a workbook. This error should not be present in EXCEL2000 (Microsoft Knowledge Base, Q168650-XL97: "Not

for all intervals (see point 4 in Figure 6.10 for how to set manually the station which will be used for the gap completion process).

3. Safe the workbook and continue with the next station with missing values. Note the already processed station in your station control EXCEL table to avoid calculating two times the same station.
4. After you have determined the best-fit neighbouring station for all pending gauges, you can continue by automatically eliminating the gaps.
5. Press the [Fill in missing Data] button. A file open dialog will appear, select the first file you would like to complete. After pressing the [Open] button the selected file is loaded and all gaps are eliminated according to the pre-calculated best-fit Regression Coefficient. After completion any inserted value is marked with an asterisk and a comment how many values were inserted (Figure 6.9, page 105) is added.



1) This EXCEL Worksheet '21-12' represents the 10-day interval starting on the 21st of December (and ending on the 31st of December).
2) The time column indicates the sampling period of the values (here from 1990 to 1998).
3) The PERIODREFERNCE column denotes to the dummy station used to achieve a complete series of intervals, and is therefore left blank (there is no need to interpolate any values for this column)
4) The number of inserted values is written to the column for CASTLEFORESTSTN (this station had a gap on the 21.12.95)
5) An asterisk (*) marks the row where a gap was replaced by a computed value. (For illustration purposes the value is outlined by a red box.)

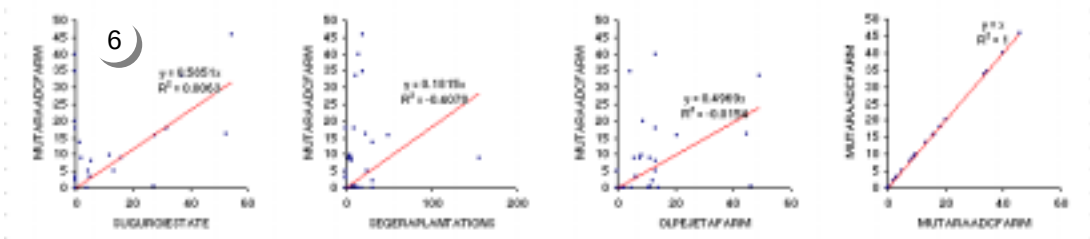
Figure 6.9: A completed EXCEL worksheet with one replaced gap at gauge Castle Forest Station.

6. Save the workbook under the same name in the same directory, and continue the infill procedure with all other stations.

Enough Memory" Error Adding Chart to Workbook, 29.09.1999, www.microsoft.com)

PEARSON CORRELATION ANA S: MUTARAADCFARM vs.:

	Corr.	Best	Cases	Median	Min	Max	Avg	Skew	RSQ	t-Test	alpha=5%	Best Fit (MAX=8)	REG. COEFF y=bx
UGUROIESTATE	0.46418477	*	28	4	0	54.6	10.8357143	1.65995802	0.21547	2.6722192	2.04840944	7 <--	0.58511823 no gaps to fill in
SEGERAPLANTATIONS	0.08229744		30	10.95	0	157.2	17.2366666	4.0394548	0.00677	0.4369593	2.04227035		
OLPEJETAFARM	0.33405655		26	9.75	0	49	12.5438462	1.87129979	0.11159	1.7362802	2.05653079		
MUTARAADCFARM	1		31	5.1	0	45.8	10.4258065	1.4886365	1	0	2.03951458		



1) The gauges to which the regression analysis is performed. The name of the station for which eventually existing gaps shall be completed has a correlation coefficient of 1.

2) Lists the computed Pearson correlation coefficients. If the two variables are completely dependent the Pearson correlation coefficient is 1.

3) In the column [Best] the best computed correlation coefficient is marked with an asterisk (*) character.

4) In column [Best Fit] the overall best result is marked with an arrow and the number of points this station achieved. A maximum of 8 points is possible, and is calculated according to the following weights: best Correlation Coefficient, T-Test, and Coefficient of Determination contribute each 2 points, the most cases and smallest skewness contribute each 1 point to the best fit. Manual corrections can be done by pointing the arrow to any desired station. When infilling missing values the Regression Coefficient of the station where the <-- arrow points to will be used.

5) Line of Regression, defined by y=bx. Where b indicates the steepness of the curve. A missing value is computed by multiplying the amount of rainfall of the best fit station with the Regression Coefficient.

6) Scatterplots are plotted in order to visualise the dependency between the analysed gauges. Any clustering or unusual distributed values can be easily identified in these scatterplots.

Figure 6.10: The result worksheet of a regression analysis for a single interval.

6

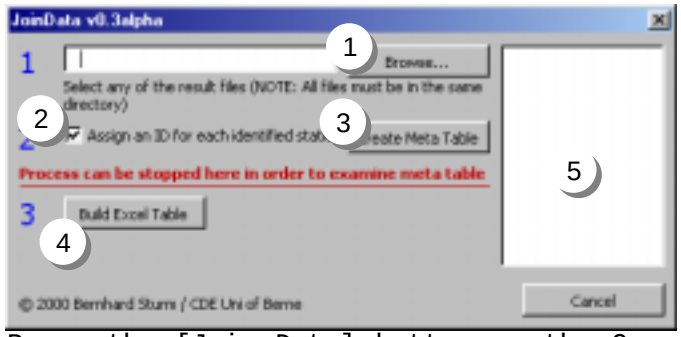
6.3.5 Step 5: Export to ACCESS and GRASS

At this stage we have eliminated all gaps (which was the primary goal of this chapter, and one of the most important conditions to carry out a spatial precipitation interpolation), but the values are hidden in various EXCEL workbooks under different directories. What we need is a method to join all rainfall values ordered by date and station into one single table. This is necessary, because we will save this table as a separate 'gap-free' table in the NRM database (or any other database).
For this purpose the last two macros will be used: [Join Data] and [Join Data (NO GAPS!)].
The first will join all stations with eliminated gaps in a single table; the later will join all originally complete stations (they never had any gaps!) in another table.
Due to the nature of most relational databases (like the NRM database) each record in a table has to be identified by a unique key. In the NRM database each rainfall gauge is identified by an integer number, which is referred to as the ID of a station. During the gap infilling procedure this ID was not used and is therefore not assigned to the station names, the two macros will have to make sure that the rainfall gauges are again linked with the correct station ID.
The two macros are split into two sections:

- Identification of the gauges
- Joining the values of the identified gauges

After the identification it is possible to manually alter the identified gauges. This can be useful in order to check whether the correct station IDs had been assigned to the gauges, or just to file the identified stations for later reference.

Start the [Join Data] Macro



1. Press the [Join Data] button on the Gap-Collect EXCEL worksheet:

Figure 6.11: Dialog of the Join Data macro

1) Select the folder from which the stations shall be joined.
2) Tick in order to assign a station ID to each station.
3) Builds a meta table containing the location and IDs of the stations to be joined.
4) Joins all stations mentioned in the meta table in one single table.
5) Progress and status information is displayed in the status list.

2. Make sure that all the completed files are stored in the folder [regression]. Also make sure that there are no other files located, otherwise the macro will try to identify these files as rainfall stations, too.
3. Press the [Browse...] button and select one of the files in the [regression] folder. Press [Open].
4. The status window displays now the number of files found in the folder. Tick the 'Assign an ID for each identified station' box, and start the identification process by clicking on the [Create Meta Table] button.
5. The macro will now load the first file on the list, and it will analyse the content found in this file. If a valid station is found (e.g. a station which will be joined), a dialog will pop-up asking the user for the ID number of the currently open station[63]. You will have to provide an ID otherwise the macro will not continue.
6. When finished the created meta table (saved in the same directory as previously selected) can be opened with a text editor and examined. Care should be taken when editing the table, as any altered entry may render the joining of the stations impossible.
7. Push the [Build Excel Table] to join the stations and their values. Select the meta table file 'metatable.txt' and click [Open].
8. The join process starts and the EXCEL table with all consecutive listed data is being built:

Station	StationID	Date	Rainfall
KAGURU	26	21.12.90	26.2
KAGURU	26	21.12.91	93.8
KAGURU	26	21.12.92	10.4
KAGURU	26	21.12.93	13.8

Table 6.6: The first four entries of the joined rainfall values. Note the assigned 'StationID' that will be used to identify the gauges in the NRM database.

[63] In practice this proofed to be a source of errors. Any mistake (wrong ID) will not immediately result in a failure. The mistake may be evident when building queries on the data, or worse, when running the spatial interpolation models. This is because a wrong ID is not related to a station name, and the data will therefore not selected in any SQL-query on station names.
It is also possible that a wrongly assigned ID will never be discovered, as a swapped ID of two closely related gauges may not result in dramatic changes of the precipitation pattern.

6

9. Select the three columns [StationID], [Date], and [Rainfall] then copy and paste them to an empty table[64] in Access with the following data fields:

Field	Data Type	Primary Key
StationID	Number (Double)	Yes (not indexed)
Date	Date	Yes (not indexed)
Rainfall	Number (Double)	No

Table 6.7: Data field definition of the Access rainfall table

10. Repeat the join procedure with the stations where no gaps existed. Use the [Join Data (NO GAPS!)] macro for this purpose, and remember that the EXCEL files were stored in the folder 'gap-free'. Add the joined data to the same Access table that you have previously created.

6.4 Spatial Interpolation

This section will discuss the way of interpolating point data using GRASS. In order to follow this chapter it is necessary that the reader is aware of some of the fundamentals of the Linux operating system and GIS applications.

[64] It may be necessary that you have to create a new table in access with the properties given in Table 6.7.

So far we have a complete dataset of 'gap free' rainfall data stored in a separate table in an Access database. Any time domain analysis of these data can easily be done using any statistical or spreadsheet software package such as MS Excel. This is useful in order to check the reliability and consistency of the available rainfall data.

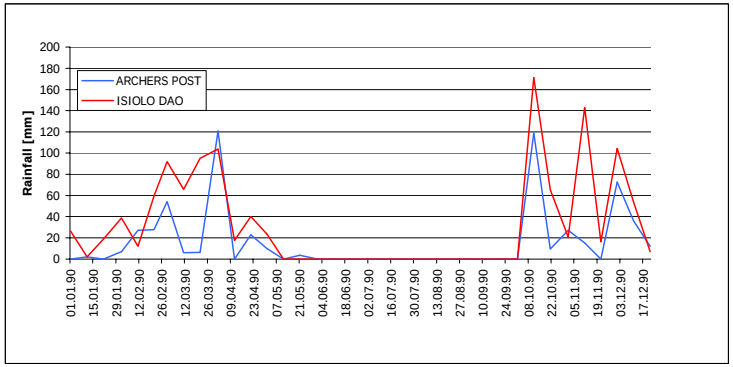


Diagram 6.2: Example of a time domain analysis for the rainfall gauges ARCHERS POST and ISIOLO DAO, showing the rainfall of the year 1990 (1.1.1990 – 31.12.1990, 10-day interval data)

One of the goals of this work is to show the distribution pattern of the rainfall in the Upper Ewaso Ng'iro Basin in Kenya, and in order to achieve this a spatial interpolation of the available observed data was performed. For this purpose a time resolution of 10 days was chosen (as discussed in chapter 3.2 (Selecting Sampling Locations and Time Span) on page 43. As the available data

covers a time horizon of 30 years we could map each 10-day interval within these 30 years resulting in 1080 different maps, but from a practical point of view, it is much feasible to use summary statistical data for the spatial interpolation. To get the 'big' picture, it is most convenient to calculate the mean values from all the 10-day intervals, which will result in a 'mean annual' map showing the mean spatial distribution of the rainfall of the last 30 years in the study area. Because of the 10-day resolution we will be able not only to observe the overall situation of rainfall distribution, but also the detailed behaviour of the pattern at the rainy season onsets.
This analysis demands for 36 interpolated maps (12 months, and each month consists of 3 10-day decades), which have to be produced with the aid of GRASS.

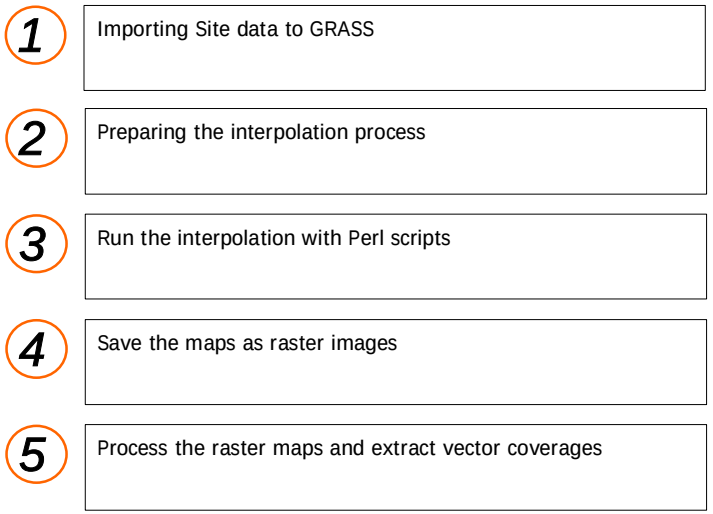


Figure 6.12: Work-Flow process of the spatial interpolation with GRASS.

Figure 6.12 illustrates each step of the work-flow process of the spatial interpolation.

6.4.1 Importing Site Data in GRASS

In GRASS point data is being treated as 'site' data. And import of site data can only be arranged in the ASCII format. The following ASCII structure is used by GRASS:

```
X-value1 Y-Value1 Z-Value1 desc1
X-value2 Y-Value2 Z-Value2 desc2
X-Value3 Y-Value3 Z-Value3 desc3
...
```

Figure 6.13: Site data ASCII structure (after: NETELER: 1998, 25)

A header is not necessary and the description is optional. Note that the 'space' character is the default delimiter, therefore the description should not contain any space character.
The z-values are represented by the observed amount of rainfall, where x- and y-values denote the geographical position of the rainfall gauge.
The GRASS command s.in.ascii will import such an ASCII file into the internal GRASS sites database.

6.4.2 Surface Interpolation of Site Data

For a surface interpolation of the site data GRASS offers three different models: Inverse Distance Weighting (IDW), Kriging and Regularised Splines with Tension (RST)[65]. It strongly depends upon the application and quality of the site data which model produces the best interpolation. For the used precipitation data, and topography a cross validation of all data revealed that the RST model was reliable enough for the interpolation (see chapter 5 (Cross Validation) on page 63, for a detailed discussion). All introduced methods and scripts are therefore optimised for the use with the RST model.
The GRASS site command s.surf.rst will produce an interpolated surface raster file based on an input site file by using the regularised splines with tension model (refer to chapter 4.2.5 (Thin Plate Splines and Relatives) on page 59, for more details about the RST model). Each raster point (or pixel) of the interpolated raster file represents the interpolated precipitation for the particular geographical location. This value is represented by indexed color information (Figure 6.14). However, this indexed color information is automatically generated by GRASS and defines arbitrary class boundaries. Because standardised class boundaries are necessary in order to compare the different maps, a manually generated color class rule was introduced. The following rules were applied to the colormap of the interpolated raster files:

[65] Future releases of GRASS will include modules to interpolate 3D site data with the regularised splines with tension model (s.vol.rst). It is then possible to compute elevation dependent precipitation maps based on 3D grid point data. Current beta versions of GRASS already support the 3D data structure (s.vol.rst was developed by Jaro Hofierka, www.geomodel.sk, 2000)

Rainfall in mm	Red Value (R)	Green Value (G)	Blue Value (B)
0	241	125	68
10	236	169	86
20	232	203	101
30	229	232	113
40	207	223	135
50	162	200	174
60	115	175	215
70	64	132	235
80	44	100	220
90	25	80	193
100	20	93	161
150	14	110	118
200	6	131	64

Table 6.8: Color rules used for raster maps interpolated with 10-day interval rainfall data.

For easier handling these color rules can be stored in a color definition file, which can be used to alter the color table of any desired raster map by pipping the file to the raster map:

```
cat rules.file | r.colors map=rainfall color=rules
```

If applied to all interpolated raster files we end up with standardised and therefore comparable precipitation maps like the one shown on Figure 6.15 (page 113).

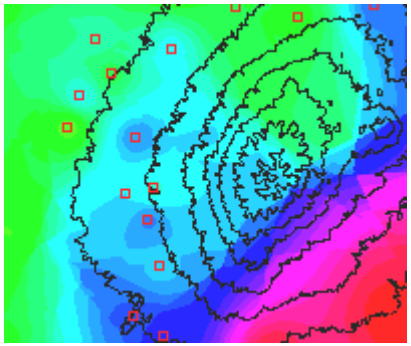


Figure 6.14: Example of an interpolated raster map with GRASS default color coding (IDW-interpolation, mean values for the month of October from 1961-1970, Mt. Kenya, green=low rainfall, red=high rainfall)

Additional analysis and map extraction based on these maps is possible. GRASS offers many tools to further analyse raster maps[66].Export from GRASS to any graphic tool (such as GIMP) is possible through various ways:

- Use the virtual CELL monitor driver in GRASS to export large-size images.
- Export raster files directly as bitmap images (TIFF)
- Create screenshots and save them as bitmap images.

[66] For instance, the command r.mapcalc allows users to logically compare and re-compute raster maps through the use of a powerful mathematical GRASS internal language. (Refer to the GRASS r.mapcalc manual for more information: http://www.geog.uni-hannover.de/grass)

6

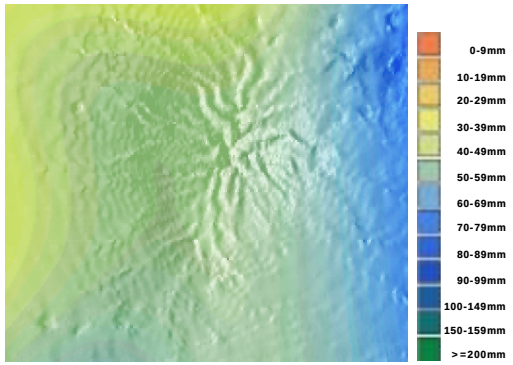


Figure 6.15: Example of an interpolated raster map with manually added color indexing rules (RST-interpolation, 10-day mean value for the decade 21.3. - 31.3. from 1990 to 1997, Mt. Kenya)

As explained in chapter 6.4 for a mean annual analysis 36 such precipitation maps have to be drawn, and despite the fast GRASS interpolation algorithms it would take a couple of hours to generate only the interpolated raster files for these 36 maps.
To save time and manpower, a couple of perl scripts were developed which control the necessary GRASS commands automatically for any number of input data.
The next section will explain the use of these scripts.

6.4.3 Prepare and Run the Interpolation Process

As described in Figure 6.12 (page 110) before the above mentioned perl scripts can be run a few things have to be prepared:

- Site list files which contain the list of site files to be processed.
- If necessary, colormap lookup table
- The work directory within the perl script

Surface interpolation can be done by using the script **rsurfrst**.

rsurfrst – The Surface Interpolation
rsurfrst is designed to automatically import and interpolate rainfall data with the Regularised Splines with Tension model.
Figure 6.16 describes the working procedures of the script. The site list and the colormap definition files are stored in the same directory as the rainfall ASCII files, whereas the parameters for RST interpolation can be entered at the beginning of the script by the user.

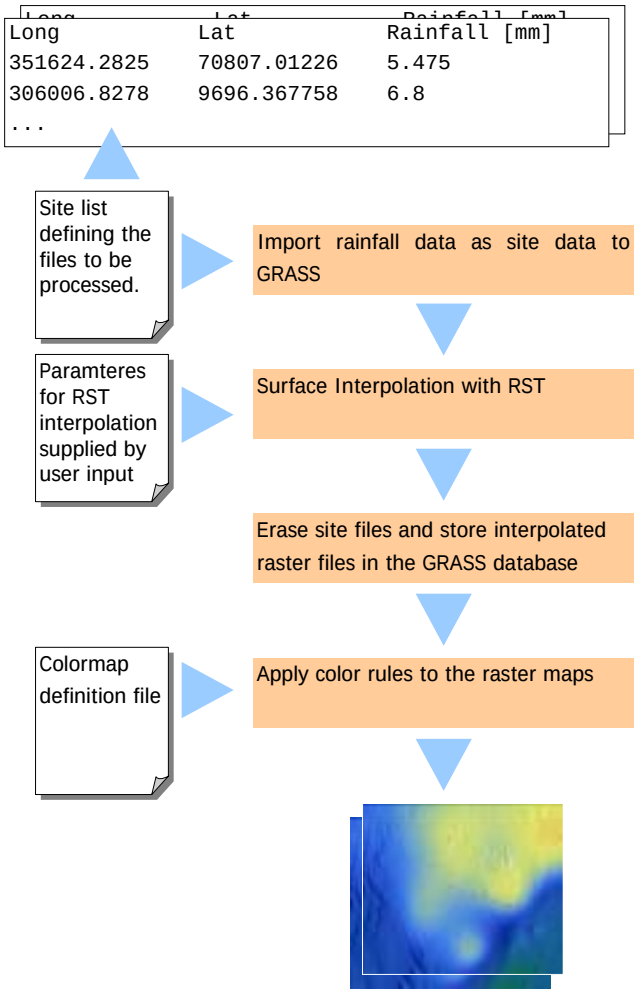


Figure 6.16: Working procedures of rsurfrst

The site list file includes a list of all files which will be processed by the script:

```
/ site control file for precipitation decadal data
1990-1997 (c) '01 Bernhard Sturm
0101.txt
0111.txt
0121.txt
0201.txt
```

The first line is introduced by a slash (/) character and contains a description. Each file is then listed by its filename, no end statement is required.
It is not necessary to generate a site list file, if only a single file has to be processed it is sufficient to enter the name of this file when prompted by the script.
The colormap file contains a set of rules for the standardised color scales of the interpolated raster maps:

```
-20 255 100 60
0 241 125 68
10 236 169 86
20 232 203 101
30 229 232 113
40 207 223 135
50 162 200 174
60 115 175 215
70 64 132 235
80 44 100 220
90 25 80 193
100 20 93 161
```

6

```
150 14 110 118
200 6 131 64
end
```

The four columns refer to the class limit in mm, the red, green and blue color values. A trailing 'end' statement terminates the list.
These two files have to be located in the same directory as the site ASCII files. If the colormap file is missing, rsurfrst will generate its own default colormap file with a scale spanning from 0mm to 180mm in a 5mm interval.
The working path of the script must be edited manually within the Perl source code. The script can be edited with a standard Linux editor (vi, nedit) and the following line can be changed to reflect the data path (where the ASCII site list files are stored):

```
$path="/home/ego/grass/Data/9097decade/";
#sets the current data path
```

Further changes may be necessary in these lines:

```
# now we interpolate with regularised splines with tension
            $input=$sites;
            $elev="9097".$sites."_RST";
```

The "9097" generates a pre-fix to the name of the raster maps for identification purposes[67].
If an input site ASCII file name was "0101.txt", then the interpolated raster map will be named by rsurfrst as: "90970101_RST". And this reads as "Decade starting at the 1st of January from 1990-1997, interpolated with Regularised Splines with Tension".

Run the Script
In order to run the script, change to the directory where the script is located by using the GRASS shell:

```
cd grass/scripts/
```

And start the script with:

```
perl rsurfrst
```

You will then be asked to enter the site list file and the colormap file.
While the script is processing the raster maps you can still work with GRASS as only the shell is occupied by rsurfrst.

[67] '9097' refers to the time horizon of the decadal precipitation data sets: 1990-1997.

6.4.4 Export and Contour Vector Extraction

After interpolation has successfully completed the raster maps can be exported for the purpose of map generation. Unfortunately GRASS offers only very limited possibilities to generate genuine maps one has to rely on other tools to produce proper maps. One way is to export raster images as TIFF bitmaps to a graphic tool such as GIMP. GIMP (GNU Image Manipulation Program[68]) has very powerful image processing features and is often compared to Adobe's Photoshop. GIMP was used for this work to superimpose different raster maps and layers for the map production process (Figure 6.17). The GRASS command for raster export to TIFF is r.out.tiff[69].

When producing maps with rainfall isolines, vector coverages (Figure 6.18) can be extracted from the interpolated raster files with r.contour:

r.contour [-qn] input=name output=name [levels=value,value,...,value] [minlevel=value] [maxlevel=value] [step=value]

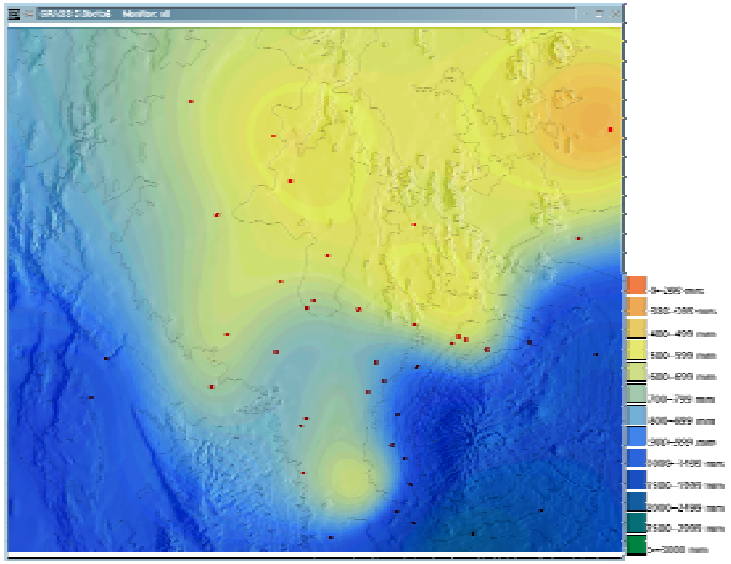


Figure 6.17: Superimposed raster, site and vector coverages with GIMP. Mean annual rainfall distribution 1967-1997

[Output name] in r.contour is the resulting vector coverage. This vector file can either exported directly to ARC/INFO or it can also be exported to a vector graphics package such as Micrografx Designer in the Autocad's DXF format:

1. Convert the binary vector file to an ASCII vector file with **v.out.ascii** (this is because export to the DXF format works only with ASCII vector files).

[68] GIMP: www.gimp.org
[69] For multiple processing another perl script is available: process. Any GRASS command can be applied to a list of raster, site or vector files. For more information refer to the comments in the 'process' source code.

2. Use **v.export** to export the ASCII vector file to various formats:

   - ASCII DLG file from GRASS Vector Format
   - ASCII DIGIT file from GRASS Vector Format
   - ASCII SCS-GEF file from GRASS Vector Format
   - ASCII ARC/INFO file from GRASS Vector Format
   - ASCII DXF file from GRASS Vector Format

3. To export to the DXF format, choose the ASCII DXF file option.
4. After export and conversion has terminated the new dxf file (it is stored in the users home directory) can be opened with any vector graphic tool that supports Autocad DXF files. Vector arcs and associated labels are organised in two separated layers.

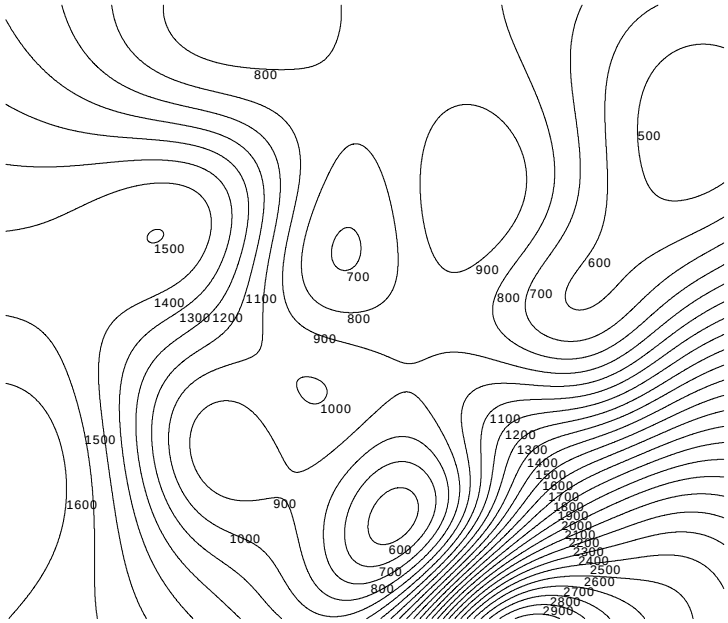


Figure 6.18: Extracted rainfall isoline vector coverage from a raster file with r.contour. Mean annual values in mm for the year 1977.

7 Presentation of Results

7 Presentation of Results

This chapter presents the results and conclusions of the interpolated rainfall in the study area. The results will be presented by covering two topics. The first topic deals with the spatial features of the interpolated precipitation, the second topic is formed by the temporal distribution of the observed patterns.
The chapter will be closed by a brief summary of the results gained through the interviews conducted during the fieldtrip in the study area.

7.1 Spatial Distribution

The spatial distribution analysis is based on the two sets of precipitation estimations performed for this study:

- Long-Term analysis: Mean decadal analysis from 1.1.1967-31.12.1997, 40 gauges used.
- Short-Term analysis: Mean decadal analysis from 1.1.1990-31.12.1997, 63 gauges used.

For a detailed discussion of the two sets refer to chapter 3.2 (Selecting Sampling Locations and Time Span) on page 43.
For both sets various detail calculations had been conducted. For instance, there are mean annual, monthly, and decadal analysis available. Although it would have been possible to present all decadal rainfall pattern 'snapshots' for both sets of analysis, this would have formed a laborious and time consuming task[70], therefore the presented results are mainly based on mean calculations.

[70] Such an extensive analysis would have covered more than 1116 single spatial interpolations for the long-term analysis sequence.

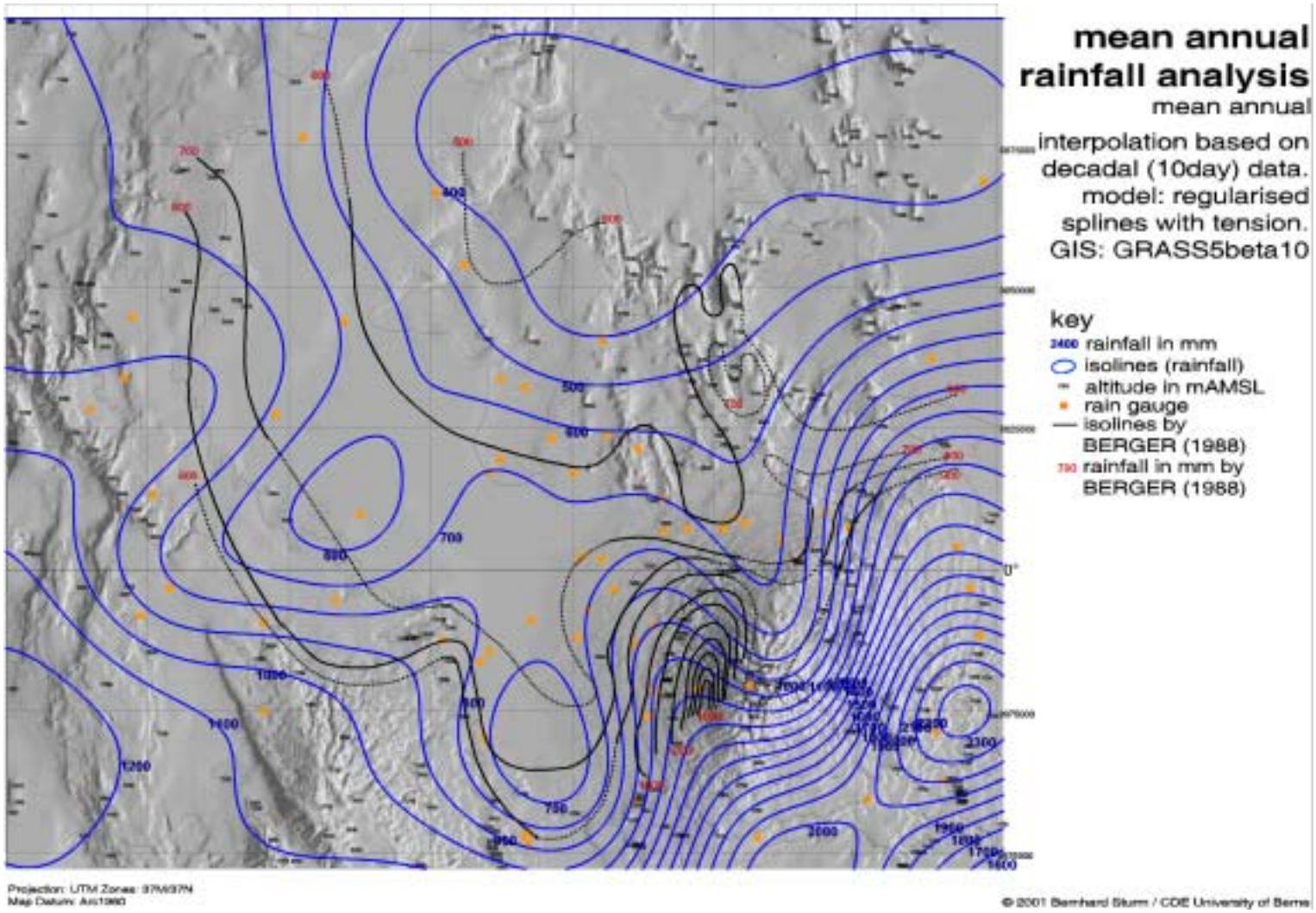


Figure 7.1: 1990-1997 mean annual rainfall distribution compared with Bergers calculations from 1988. (map by B. Sturm)

7.1.1 Annual Distribution

For the annual distribution the short- and long-term situations will be presented in two different ways. For the long-term analysis a colour gradient map illustrates the general distribution pattern. For the short-term analysis, isolines[71] marking regions of the same rainfall regime were used.

Long-Term Analysis (1967-1997)

Figure 7.2 shows the spatial rainfall distribution from 1967-1997. The wettest regions with more than 1000mm rainfall per year can be observed around the Mt. Kenya and the Nyandarua Range. A maximum of more than 3000mm rainfall per year can be found on the southern slopes of the Mt. Kenya between Castle Forest Station and Irangi Forest Station. The north-eastern parts of the Basin receives an annual rainfall of less than 500mm.

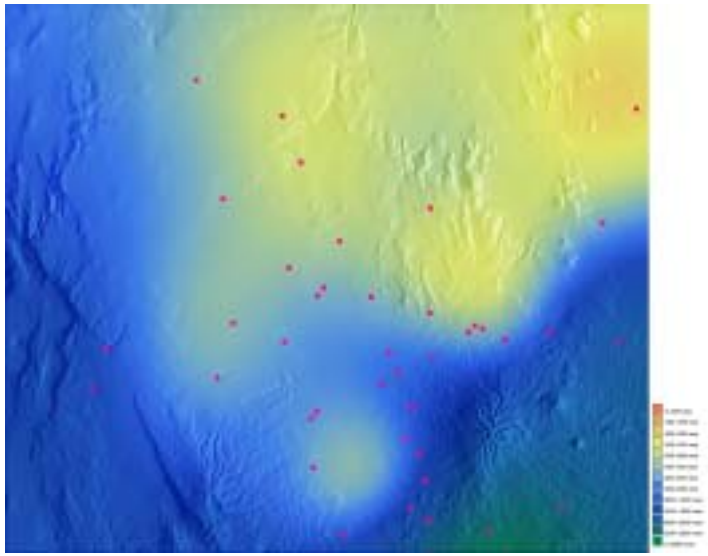


Figure 7.2: Mean annual rainfall distribution from 1967-1997. Scale ranges from 0mm – 3000mm.

Short-Term Analyis (1990-1997)

As already stated the short-term analysis comprises the use of 63 rainfall gauges for the estimation of the spatial rainfall. The majority of the additional stations – in comparison to the long-term analysis – are situated in the lowlands between the Mt. Kenya and the Nyandarua range. These additional gauges enhance the accuracy of the interpolation algorithm in this area. A big drawback of the long-term analysis is the lack of gauges directly in mountainous to-

[71] Isolines mark lines of same values, in this case, lines representing the same amount of estimated rainfall.

pography. This is partly solved by the additional gauges which form a transect on the western slope of Mt. Kenya. With the aid of these gauges, a more detailed view of the behaviour of the rainfall around the hill slopes is possible.

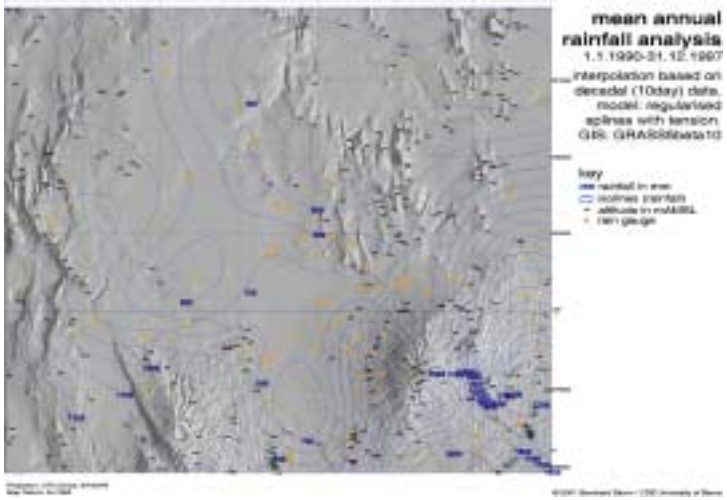


Figure 7.3: Mean annual rainfall distribution from 1990-1997. Rainfall isolines are plotted as blue curves. (Map by B. Sturm)

On Figure 7.1 (which is the same as Figure 7.3) one can clearly see the impact of the use of more gauges compared to the long-term analysis (as shown in Figure 14.2. In appendix 4) Especially the pattern on the western slopes of Mt. Kenya is more differentiated than it appears on Figure 7.2 on page 121: the mountain represents an obstacle in the rain field around the Mt. Kenya. The region in the vicinity of Naro Moru and Nanyuki is favoured by higher annual rainfalls (800mm – 1200mm), where on the other hand, the region north of the mountain lies in the leeward oriented part of Mt. Kenya, and receives substantial less rainfall than the other parts of the mountain (600mm-800mm).

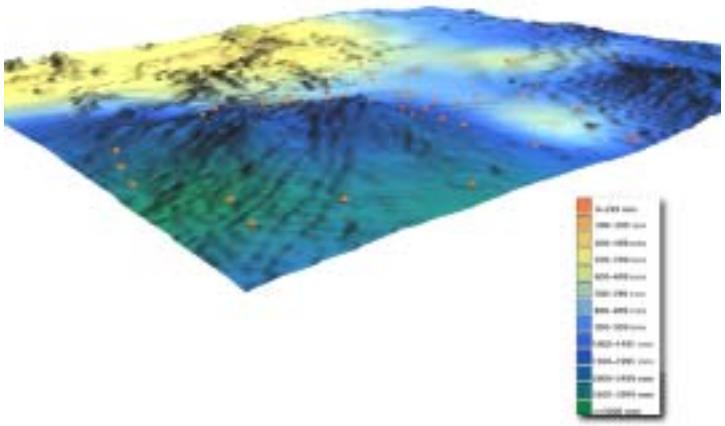


Figure 7.4: 3-dimensional view of the mean annual rainfall distribution of the entire Upper Ewaso Ng'iro Basin from 1990-1997. Orange polygons represent rainfall gauges. (3D imaging by B. Sturm with NVIZ2.2)

As the means of visualisation are playing an important role in understanding the general distribution pattern, a visualised 3-dimensional DTM (digital terrain model) with a draped estimated rainfall pattern was used. This modified DTM is presented through the Figure 7.4 and Figure 7.5. Both images

visualise the amount of the mean annual rainfall by means of color coded gradients. Orange polygons mark rain gauges used for the estimation. On Figure 7.5 the observer is located somewhere above the Nyandarua range and looks towards the westward side of Mt. Kenya. Superimposed isolines help to identify regions receiving similar amount of rainfall. It is noteworthy that the rainfall gradient along the previously mentioned transect gauges is not describing a normal orographic gradient where rainfall increases with increasing altitude, but more a steady curve with the same amount of rainfall with increasing altitude.

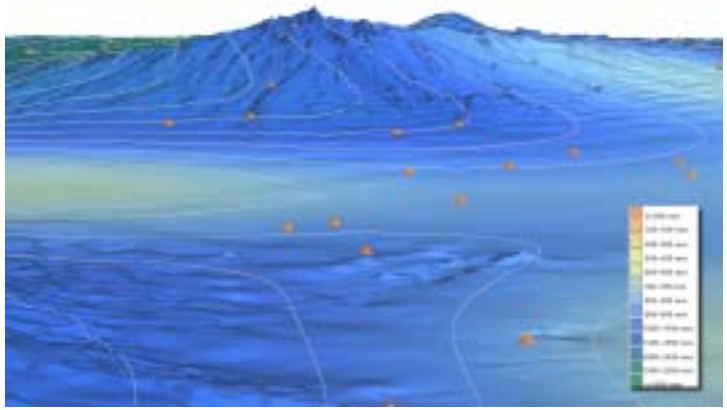


Figure 7.5: 3-dimensional view of the mean annual rainfall distribution around the western slopes of the Mt. Kenya from 1990-1997. Orange polygons represent rain gauges, grey curves mark rainfall isolines (3D imaging by B. Sturm with NVIZ2.2)

7.1.2 Decadal Distribution

Long-Term Analysis (1967-1997)

This is a brief overview of the precipitation pattern obtained from the mean decadal interpolation based on data from 1.1.1967 to 31.12.1997. Please note that the herewith-presented decades are grouped into monthly intervals, for illustration purposes one typical decade is shown.

Decades 1-3 (January 1.1.- 31.1.)

1.1.-10.1.

The study area is mainly under very dry conditions. Only small regions on the eastern slopes of the Nyandarura range and on the southern slopes of the Mt. Kenya receive more than 25mm of rain per decade.

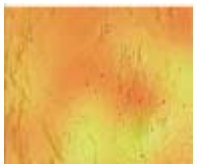
Decades 4-6 (February 1.2. - 28./29.2.)

1.2.-10.2.

Like during January, the Upper Ewaso Ng'iro basin doesn't receive much rain. February can also be designated as a dry month, towards the end of February (Decade 6 (21.2. - 28./29.2)) the small wetter regions located on the south oriented slopes of Mt. Kenya

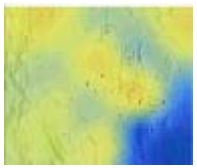
receive remarkable more rain (30mm - 40mm) than the rest of the basin.

Decades 7-9 (March 1.3. - 31.3.)


1.3.-10.3.

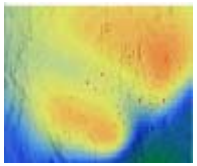
March is the month where the 'long rains' set in, this change in the atmospheric pattern is clearly reflected in the distribution of the precipitation pattern of the region. Starting with decade 7 (1.3.-10.3.) the western slopes of Mt. Kenya, the eastern slopes of the Nyandarura range receive more than 30mm, but less than 50mm of rain. However the pattern changes for the next 10 days only slightly: the Nyandarura range does no longer receive that much rainfall, instead the overall rainfall pattern is more shifted towards the north-east of the study area (the 25mm decadal precipitation line is located just south of Isiolo). The onset of the 'long rains' is remarkably visible during decade 9: except the central and northern parts of the district, all regions receive more than 30mm rain during the last 10 days of March.

Decades 10-12 (April 1.4. - 30.4.)


1.4.-10.4.

The long rains have reached the entire basin: only small cells of less than 25mm rainfall remain in the central and northern part of the study area. The slopes of the Mt. Kenya receive more than 80mm during the first 10 days of April. During the last two decades of April the following pattern establishes over the region: a small belt obtaining less than 30mm rain develops from Colcheccio over Mukogodo to Dol Dol. Where on the other hand a zone of more than 60mm rainfall per decade grows in the lowlands starting from the north-western slopes of the Mt. Kenya towards Loldoto Farm, and stretching at the end of April over Ol Mysor Farm. In the last decade (21.4. - 30.4.) the wettest region (>80mm) spans from the south-eastern slopes of Mt. Kenya (Irangi Forest Station with more than 200mm) to the Aberdare Mountains (south of Ol Bolosat Forest Station).

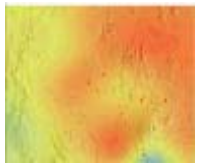
Decades 13-15 (May 1.5. - 31.5.)


1.5.-10.5.

The pattern changes remarkably. Where the central and the northern parts of the study region begin to dry out, the regions receiving more than 60mm per decade can be found in the west over the Aberdares and around the Mt. Kenya region. During the first two decades, cells with almost no rainfall develop around the footslopes of the Nyandarura range and in the northern region. These cells move north- respectively southwards during May until they meet in the central

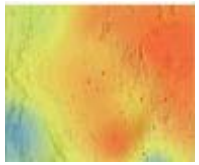
part of the Upper Ewaso Ng'iro basin at the beginning of June.

Decades 16-18 (June 1.6. – 30.6.)


1.6.-10.6.

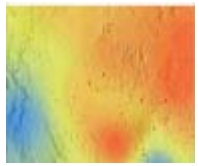
During the month of June most parts in the Upper Ewaso Ng'iro Basin receive less than 50mm per decade.
Especially the lowlands are suffering from an almost complete absence of rainfall (in the vicinity of Isiolo and Archer's Post the estimated amount is around 0mm-10mm per 10 days). At the end of June the continental rains begin to set in, and start to spread northwards.

Decades 19-21 (July 1.7. – 31.7.)


1.7.-10.7.

The pattern during July represents a more marked situation: continental rains are prevalent in the western part of the study area with the local occurrence of 60-80mm rainfall at the western slopes of the Nyandarua Range. Higher rainfall is to be expected in the southern part of the Mt. Kenya between Castle Forest Station and Irangi Forest Station.

Decades 22-24 (August 1.8. – 31.8.)


1.8.-10.8.

August represents the study area as divided into two parts: the Nyandarua Range marks the border of this division. West of the Range wetter conditions, typical for the continental rains, in the magnitude of 40mm-80mm rain per 10-days prevail. Almost dry conditions dominate east of the Range the area. Only the southern parts of the Mt. Kenya benefit from rains delivering approximately <70mm water per decade to the slopes.

Decades 25-27 (September 1.9. – 30.9.)


1.9.-10.9.

As the ITCZ moves southwards over the Upper Ewaso Ng'iro Basin rainfall is almost suppressed by divergent circulation conditions. The first 10 days of September mark the end of the continental rains in the western part of the Basin. During the rest of the month a transitional rainfall regime reigns over the area: most parts receive less than 20mm rain per decade.

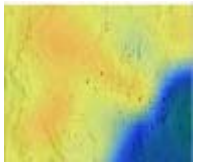
Decades 28-30 (Oktober 1.10. – 31.10.)


1.10.-10.10.

The onset of the short rains can be observed in the course of October. As the meteorological equator moves slowly south-

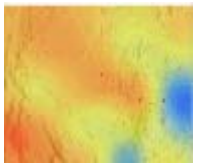
wards (and so does the high-pressure zone). The change from divergent to convergent circulation brings the onset of the short rains in starting at the end of August and the beginning of September. Mainly the south-eastern parts of the basin benefit from rains with more than 100mm per decade.

Decades 31-33 (November 1.11. – 30.11.)


1.11.-10.11.

During the first two decades in November the short rains intensify for the fertile eastern and southern parts of the Mt. Kenya slopes. This band of high rainfall dissolves towards the end of November into single patterns of higher rainfall. This is caused by the stronger domination of the north-east trade winds, which mark the end of the short-rain period.

Decades 34-36 (December 1.12. – 31.12.)

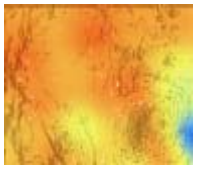

1.12.-10.12.

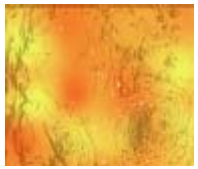
At the end of the year dry conditions reign over the Upper Ewaso Ng'iro Basin, only small cells of 60-70mm rain per 10 days can be observed on the north-eastern and south-western slopes of the Mt. Kenya. The rest of the Basin receives less than 20mm per 10 days.

Short-Term Analyis (1990-1997)

As expected the analysis of the decadal rainfall estimates for the short-term set reveals similar features as the long-term set. Although there are slight variations in details, that seem to originate from the more rain gauges used for the short-term estimates, the overall pattern remains the same as described for the long-term analysis. Therefore, only a very brief overview shall be provided. Please note, that the following small rainfall distribution maps show decadal distribution and illustrate the precipitation features discussed in the text. The white dots on the maps show the locations of the rain gauges.

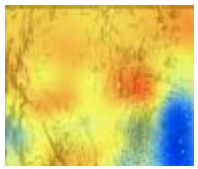
Decades 1-8 (1.1. – 20.3.) Dry Season

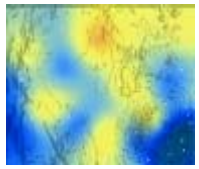

1.1.-10.1.


1.2-10.2.

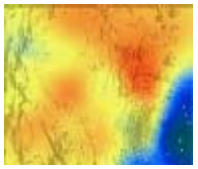
During January and February the Basin suffers from the very dry conditions. Especially the lower parts of the Upper Ewaso Ng'iro basin receive less than 10mm of rainfall per decade. The only areas where 10mm-30mm of rainfall can be expected are between the Mt. Kenya and the Aberdares (Decades 3-8). At the end of March (shortly before the 20th of March) the long-rains set in.

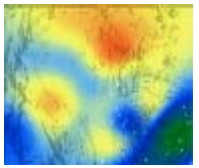
Decades 9-15 (21.3. – 31.5.) Long-Rains


21.3.-31.3.


1.4.-10.4.

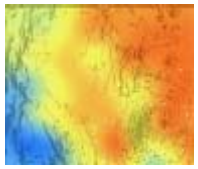
These decades can be defined as the period of the long-rains (BERGER, 1989: 42). Most rainfall can be expected in the southern part of the basin at higher altitudes. The southern slopes of Mt. Kenya receive a maximum of rainfall of more than 200mm per decade (decade 13 and 14). Remarkable is that the pattern marking the long-rains isn't consistent: after the onset of the rainy season most parts of the study area are receiving more than 50mm per decade, but just after the 10.4. this pattern collapses, and only a small band of high rainfall can be noticed at the eastern slopes of the Mt. Kenya, whereas the rest of the Upper Ewaso Ng'iro basin doesn't receive more than 70mm for the decade of 11.4.-20.4. 10 days later the former pattern is again established.

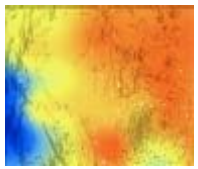

11.4.-20.4.


21.4.-30.4.

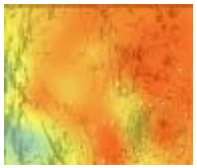
The long-rains terminate by the end of May by leaving the study area in a south-eastward direction.

Decades 16-27 (1.6.-30.9) Transition


1.7.-10.7.

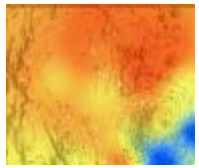

1.8.-10.8.

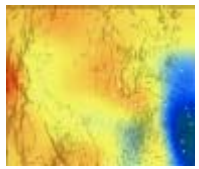
Continental rains set in ranging from the 1st of June to the end of August and the beginning of September. It is remarkable that only the western part of the area receives slightly more than approximately 40mm rainfall. The pattern is very disturbed and can not be compared with the much stronger appearance of the long-rains. The Nyandarua range and a small

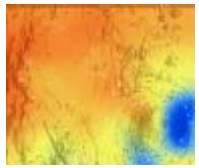

1.9.-10.9.

region around Kinamba and Lariak Forest Station receives the most of the rainfall (approximately 50mm – 80mm per 10 days).
However at the end of September nowhere in the entire basin falls more than 40mm per 10 days. A dry period can be identified for more than 20 days (11.9.-30.9.).

Decades 28-36 (1.10.-31.12) Short-Rains


1.10.-10.10


21.11.-30.11.


11.12.-20.12

From October to the beginning of the December a more moderate and balanced pattern can be observed: the onset from the Short-Rains can be traced from the south-eastern slopes of the Mt. Kenya moving slowly westwards and northwards until the end of November (with rainfall of more than 150mm per decade)
The amount of rainfall increases also slightly for the lower parts of the Upper Ewaso Ng’iro: from mid October to mid December between 20-40mm decadal rainfall can be observed.
The transition to dryer conditions and the short-rains ends during the second half of December (Decades 35 and 36).

7.1.3 Annual Precipitation Patterns

According to the work of Berger, three distinctive features in the annual rainfall distribution can be isolated: (BERGER, 1989: 42)

- The Long-Rains from March until May
- The Continental Rains from June until August
- The Short-Rains from September to December

In his study Berger has pointed out, that these rainy seasons are not distributed evenly over the Ewaso Ng’iro Basin. There are regions that are favoured by all three rainy seasons; others are only influenced by the long- and short-rains.

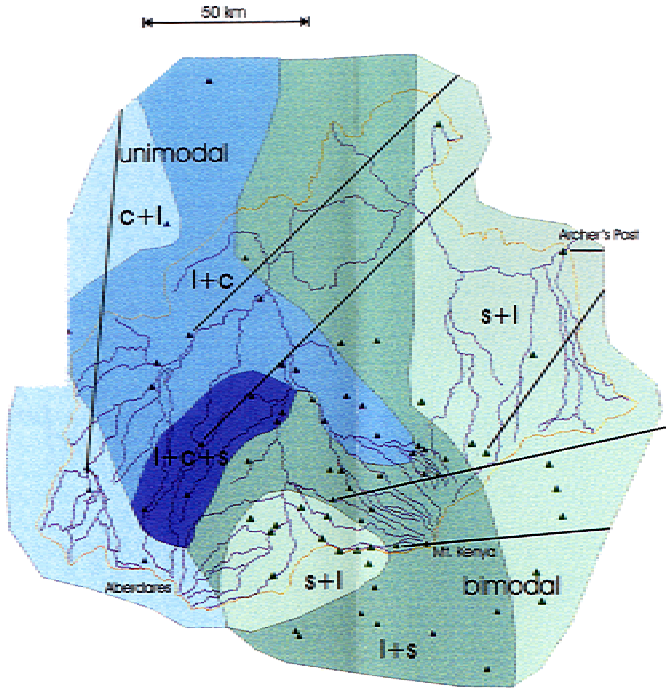


Figure 7.6: Annual rainfall distribution in the Upper Ewaso Ng'iro Basin. L= "long rains" totalling more than 25% of annual rainfall, C="continental rains" totalling more than 25% of annual rainfall, S="short rains" totalling more than 25% of annual rainfall. The order of l,c, and s given in accordance with the relative importance of the rainy phases. (Map compilation by Liniger and Schwilch, 1998)

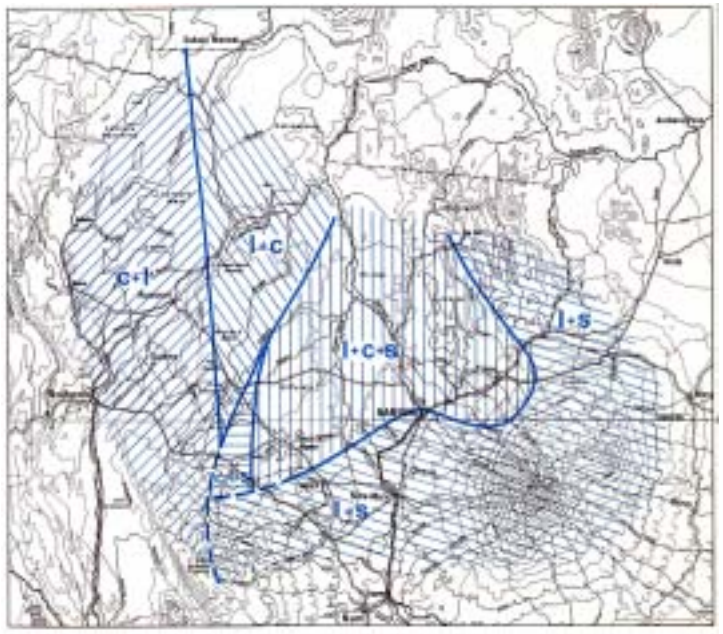


Figure 7.7: Rainfall regimes on the Laikipia Plateau, 1971-82. "l" denotes "long rains" totalling more than 25% of annual rainfall, "c" denotes "continental rains" totalling more than 25% of annual rainfall, "s" denotes "short rains" totalling more than 25% annual rainfall. The order of l,c,s is given in accordance with the relative importance of the rainy phases (Map compilation by Peter Berger, 1989)

Berger has studied monthly totals from 1971-1982 and developed the simple classification presented in Figure 7.7 where the three rainy seasons are abbreviated as 's','c', and 'l'. As a limit, Berger has proposed to use the regions, which receive more than 25% of the total annual rainfall to be classified as belonging to one of the rainy sea-

sons. In Figure 7.8 on page 130 the same classification was used, but based on a visual assessment on mean decadal rainfall data form 1990-1997. A newer classification had been developed in 1998 by Hans-Peter Liniger and Gudrun Schwilch (LINIGER, 1998: 17). See Figure 7.6 on page 129 for this new compilation. The following discussion will compare Berger's classification with the classification defined by the findings of the herewith-presented study. The new classification reveals some differences to the Berger classification, but is more detailed because of the higher resolution of the interpolation. However, only slight deviations between the two maps are visible: according to Berger the long rains dominate (partly in association with the short rains) much of the northern and western slopes of the Mt. Kenya. Based on the new estimates, Berger seems to have underestimated the importance of the short-rains that are of a high importance for crop production. He designates the north-eastern parts towards Isiolo as being dominated by long- and short-rains with the relative importance of the long-rains. The pattern based on the new estimates show, that the short rains are much more dominant in this north-eastern part as assumed by Berger. This is also illustrated by the map (Figure 7.6) reproduced by Liniger et al. on page 17 of the Eastern and Southern Africa Geographical Journal. (LINIGER et al., 1998: 17). This underestimation of the short-rains may be due to the non-availability of data, and coarse resolution in terms of time and rainfall gauges used by Berger.

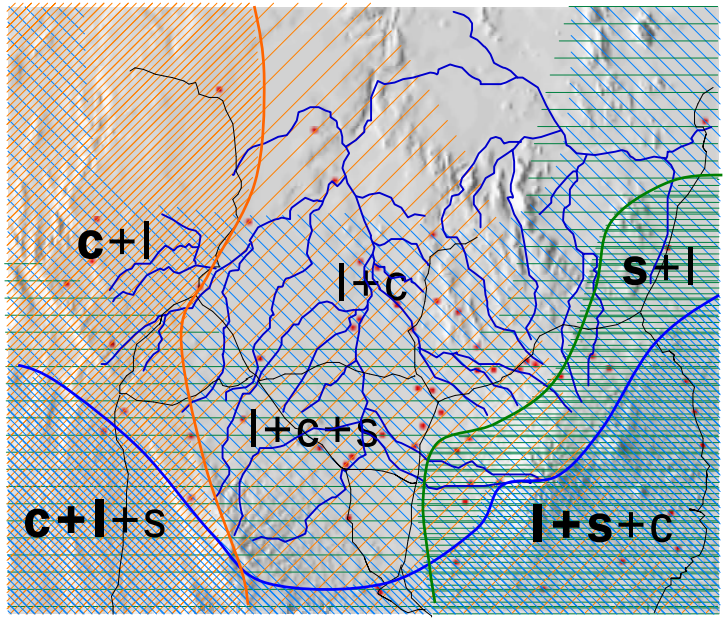


Figure 7.8: Appearance of the long-rains (blue), continental-rains (orange), and short-rains (green) in the Upper Ewaso Ng'iro Basin according to the analysed decadal data from 1990-1997. The order and style (bold or regular) of l,c,s is given in accordance with the relative importance of the rainy phase. (Map by B. Sturm).

Concerning the dominance of the different rainy seasons, we can divide the study area in three regions:

- An eastern section that is dominated by the long- and short-rains during April-May and October-December.
- A western section that receives most of its rainfall during April-August (long- and continental-rains).
- A middle section that can also be labelled as a region of transition between the rainfall regimes. With the exception of the southmost part of the study area, most of the rain this region receives, originates from all three rainy cycles, but with much less rain than the other parts.

Further Berger states that: *"With the exception of Solio in the extreme south of Laikipia and of Marania which lies outside the District on the northeastern slope of Mt. Kenya, all stations show a three-peaked rainfall distribution."* (BERGER, 1989: 42). While analysing the annual distribution of the rainfall in the Upper Ewaso Ng'iro Basin, the above statement proved to be not entirely correct, as the quoted 'three-peaked' regime does solely exist, but is geographically distributed and not all stations show such a distinctive annual distribution. Diagram 7.1 on page 131 shows the annual rainfall distribution for each rainfall station used for the long-term decadal analysis (1967-1997).

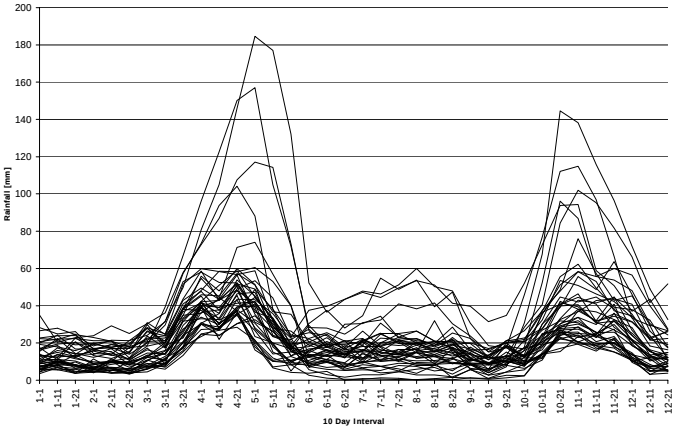


Diagram 7.1: Annual rainfall distribution. Each line represents the mean 10-day rainfall from 1967-1997 observed at one gauge. Source: B. Sturm.

Most of the stations are situated in a two-peaked annual regime with an almost equal distribution of received rain during the long- and short-rainy periods (approximately 40mm at the peaks). Only a few stations show the clear feature of continental rains in the middle of the year[72]. All of these stations are located either on the southern slope of the Mt. Kenya (IRANGI, CASTLE, and RAGATI FOREST STATIONS) or in the Nyandarua Range (OL BOLOSAT and SOUTH MARMANET FOREST STATIONS). This is in accordance with the

distribution of the rainy seasons shown by Figure 7.8 on page 130.

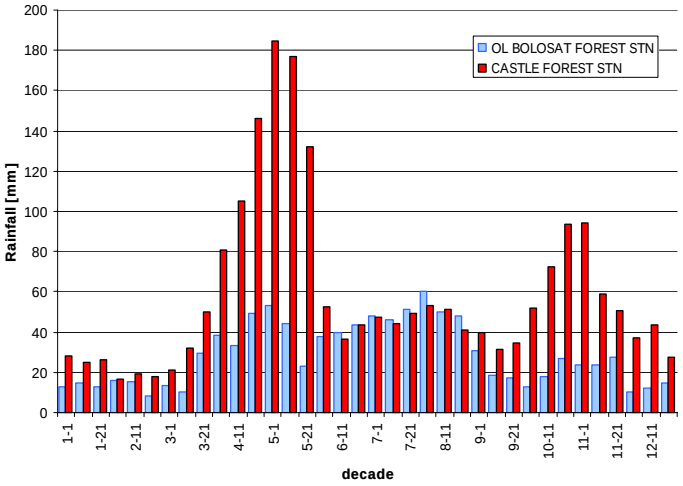


Diagram 7.2: Rainfall regimes for OL BOLOSAT FOREST STN and CASTEL FOREST STN. Mean decadal rainfall from 1967-1997. Source: B. Sturm.

Diagram 7.2 directly compares the rainfall regimes of two stations showing a distinctive continental rainy period. Both stations reflect the distribution described by Berger as 'three-peaked'. Where the higher observed rainfall at Castle Forest is due to the windward orientation at the southern slopes of the Mt. Kenya, and hence under the influence of orographic rainfall during the south-eastern trade wind circulation.

7.1.4 Extremes: Wet and Dry Years

The above described annual rainfall patterns are not persistent over longer time spans. And it should be noted that the presented statistics and diagrams describing the annual and seasonal features are based on mean calculations. However, mean values are only useful if the data are normally distributed about the mean. Unfortunately rainfall data are rarely normally distributed, and according to Graham Summer there are two rules of thumb concerning the statistical spread of precipitation data (after SUMMER, 1988: 350):

1. The shorter the duration under consideration, the less likely is that the resulting distribution will be normal. Rainfall data is mostly positively skewed.
2. The drier the site the more variable is the precipitation from decade to decade.

For the first rule the WMO recommends a minimum sequence of 30-year data (SUMMER, 1988), which is maintained by our long-term precipitation analysis from 1967-1997. For the second rule we will use the wettest decade for the OL BOLOSAT FOREST STATION, the decade from the 1st - 10th of May in order to avoid analysing a too variable decade. By

[72] These stations are: OL BOLOSAT FOREST STN, SOUTH MARMANET FOREST STN, IRANGI FOREST STN, CASTLE FOREST STN, MUGIE RANCH, RAGATI FOREST STN.

superimposing a 5-year running mean trend curve it is apparent that there are in fact year-to-year fluctuations that characterise the long-term rainfall regime for this station (Diagram 7.3). The 5-year running mean smoothes the most extreme irregularities, and hence making the task of identification the cycles or trends a little easier, at least on a visual non-statistical level.

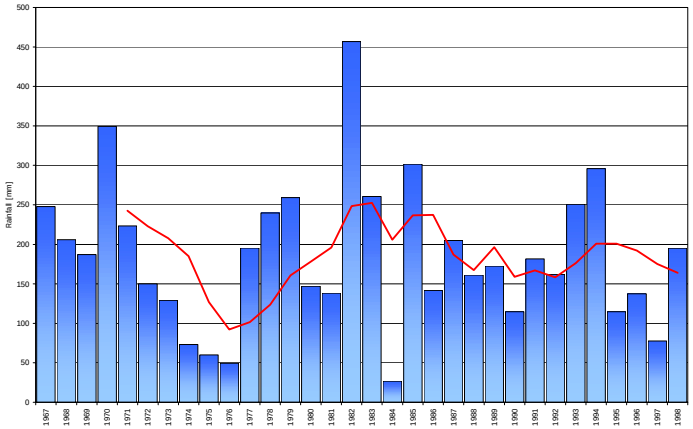


Diagram 7.3: Decadal rainfalls for CASTLE FOREST STN for the 10 days from 01.05-10.05. from 1967-1998 with 5-year running means superimposed as red trend curve. Source: B. Sturm.

Diagram 7.3 also shows that there are wetter and drier sequences marking the long-term precipitation distribution.

In his study about the rainfall and agroclimatology of the Laikipia Plateau in Kenya[73] Peter Berger has identified two years of extreme wet and dry fluctuations. These are the years 1977 (wet) and 1980 (dry). Table 7.1 on page 140 shows the mean annual rainfall for selected years, and the year 1977 can indeed be identified as an exceptionally wet year. For his study Berger has manually drawn the isohyets based on a 'simple good sense' method by using his own experience and the rainfall data obtained by the rain gauges (BERGER, 1989). We will compare Berger's map for the wet year 1977 with the estimates based on the regularised splines with tension model used for this work.
Figure 7.8 combines the map presented by Berger in 1989 and the isolines computed based on the long-term sequence from 1967-1997. As Berger did not have sufficient information about the rainfall in the area[74] he marked his uncertainties by drawing dotted black lines instead of solid black isolines.

[73] Berger, Peter, 1989. 'Rainfall and Agroclimatology of the Laikipia Plateau, Kenya'. Bern, *Geographica Bernensia*, A7

[74] Berger was not actually caluclating the entire areal rainfall because his method was based on the manual 'estimation' of the location of isolines between known rainfall gauge observations.

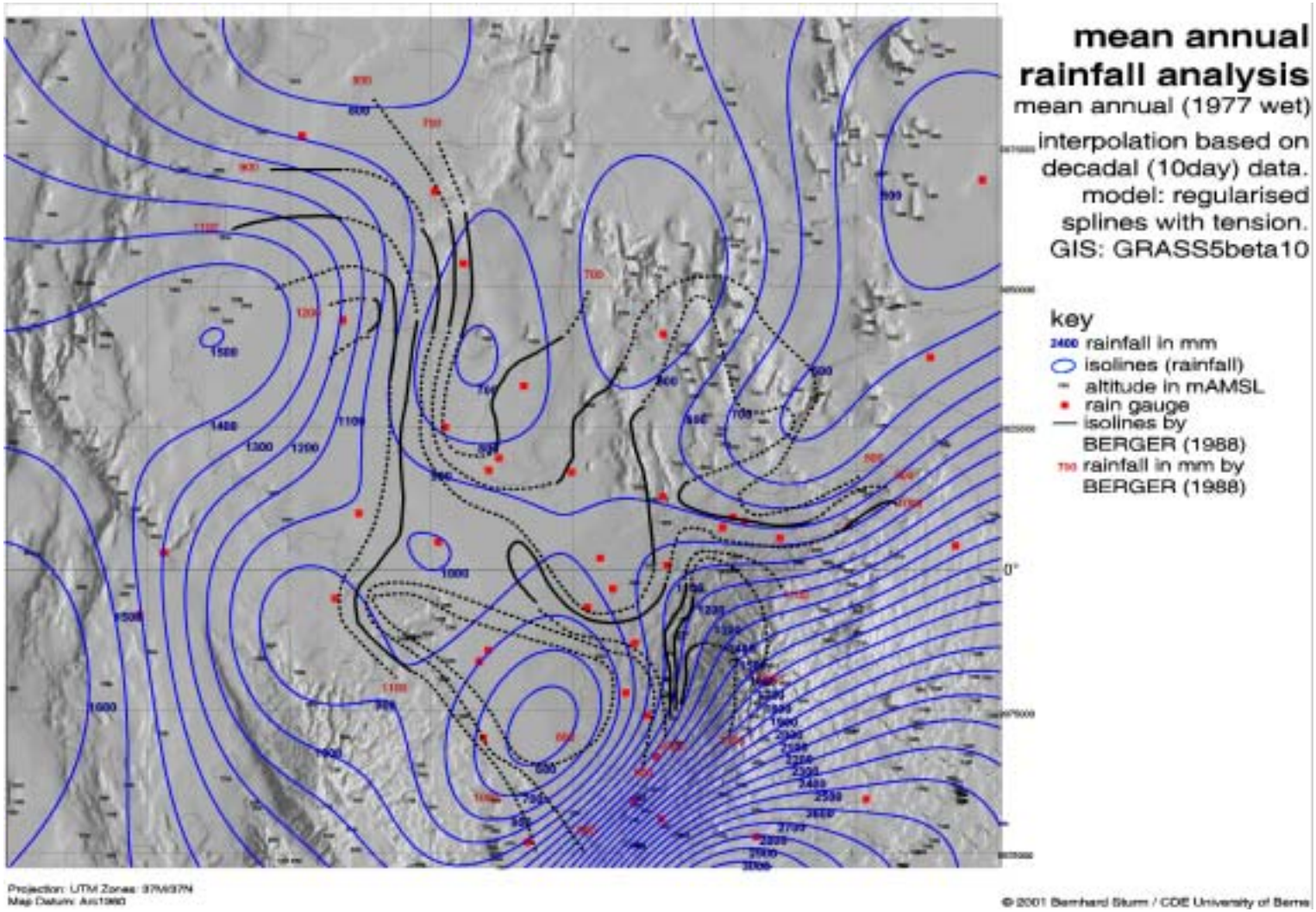


Figure 7.9: Mean annual rainfall distribution for 1977 (wet year) compared with the caluclated isolines by Berger (1988). (Map by B. Sturm)

By studying the two isoline patterns one can clearly see that they correspond well for most of the area. Especially the general distribution pattern is well represented by Bergers isolines, however, there are differences particularly in the west of Mt. Kenya where Berger estimated a total rainfall as high as 800mm per year. On the other hand, the RST estimates indicate a much lower rate of 600mm per year for the same region.
But, however, Bergers interpretation of the rainfall distribution at the northern slopes of the Mt. Kenya seem to be more reliable as the rainfall on the leeward side of the Mt. Kenya is clearly visible in Bergers isoline pattern, whereas the regularised splines with tension method suffers from the lack of additional rainfall gauges in this region, and tends to underestimate and simplify the more complicate rainfall regime for mountainous topography[75].
In a next step we will compare the pattern of a dry and a wet year[76] around the Mt. Kenya. Figure 7.10 shows the two extreme patterns with the superimposed 750mm isoline. Above this limit the establishment of temporary grass leys is possible which allow to carry some stock over the year (JACKSON, 1979: 191).

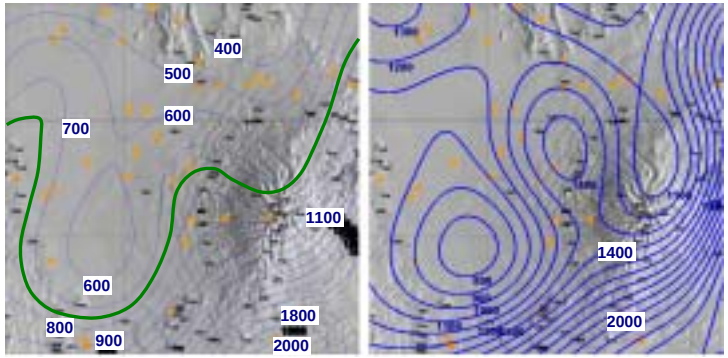


Figure 7.10: Annual precipitation for a dry and a wet year (left: 1991, a dry year; right: 1997 a wet year). The green curve marks the 750mm rainfall limit above which the establishment of temporary grass leys is possible (JACKSON, 1979: 191).

This limit illustrates clearly that during dry years it becomes very hard to maintain livestock in areas that receive less than 750mm annual rainfall.
Another interesting difference is that the mountain plays an important role in the rainfall distribution: during the wet year in 1997 the north-eastern part of the Mt. Kenya was favoured by a band of high rainfall (>1300mm) whereas during the drought year 1991 this band could not be observed.

[75] That this lack can be overcome by a denser gauge network is demonstrated by Figure 7.3 on page 122 where the leeward side of the Mt. Kenya is more articulated represented by the RST estimates.
[76] The two years 1991 (dry) and 1997 (wet) were chosen due to the fact that they are both available in the more accurate short-term analysis set.

7.2 Temporal Distribution

The previous chapter has already introduced some of the interannual features of the rainfall. In this chapter we will focus on the temporal distribution of the rainfall in the Upper Ewaso Ng'iro Basin.

7.2.1 General Temporal Analysis and Trends

The study of time series is of interest for several reasons. For a long-term development and planning it is of importance to know whether present rainfall totals are below or above average in order to avoid planning errors. The analysis of time series can also reveal any eventually underlying trend affecting the general temporal distribution pattern of rainfall.

As the main issue of this paper is on the spatial interpolation of rainfall, we will only briefly touch the question whether there is a trend affecting the present rainfall condition in the study area. Nevertheless, already a short analysis reveals some very interesting features, which give room for further and sounder time series analysis.

For a long-term analysis over more than the previously introduced 30 year period, a privately maintained gauge was chosen: Jacobson Farm, a farm located on the western slopes of the Mt. Kenya, located at 1905m AMSL has the longest rainfall logs in the study area with over 64 years of rainfall observations (from 1.1.1934 up to present time).

Figure 7.11: Rain gauge (red circle) at Jacobson Farm. Image by B. Sturm

When plotting the 10-year running mean trend curve for the monthly totals for Jacobson Farm, a slight positive trend towards more rainfall per month is visible (see Diagram 7.4). It is interesting to note, that between 1952-1972 there was an increase of monthly totals observed. This increase can be explained by a somewhat shifted northern sub-tropical high pressure cell closer to the equator during the years 1951-1969. This movement lead to reduced latitudinal migration of the equatorial belt and therefore some equatorial locations received more rainfall in this period (JACKSON, 1979).

After 1972 Jacobson Farm received less rainfall than during the previous periods, but the situation improved steadily up to present again. Whether the latest increase can also be seen in the light of a moved tropical high pressure cell should be analysed in further researches.

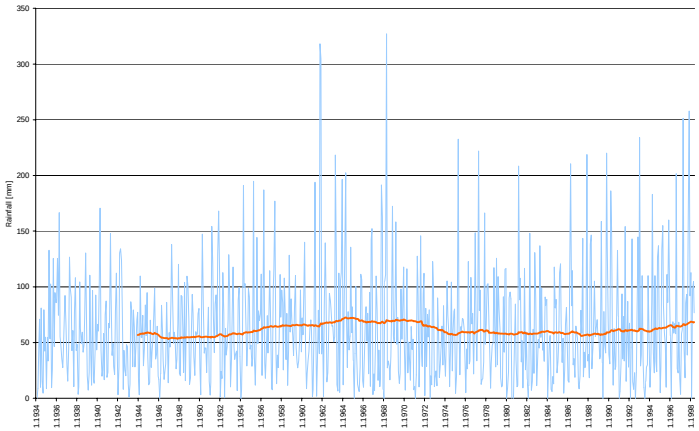


Diagram 7.4: Monthly totals for JACOBSON FARM from January 1934 – May 1998. Superimposed 10 year running mean as orange trend line. Source: B. Sturm.

During the interviews with the farmers in the study area, it was a general notion that there is maybe not less rain than in the past, but that the rains have become more variable. These observations can be supported by Diagram 7.5, which shows the rainy days per year[77] and the 10-year running trend line.

The trend line indicates a dramatic drop in the number of rainy days per year after 1960. It can be stated that although the amount of rain had been increased, the number of actual rainy days decreased and hence the rainfall variability and intensity increased. Which could explain the remarks of the farmers that the rainy days became less predictable than in former years. A higher intensity means that the dry periods between the single rains are lengthened which increases the water stress for plants and, probably, the incidence of soil erosion (BERGER, 1989).

[77] A day was counted as a rainy day if there was more than 0.0mm of rainfall observed.

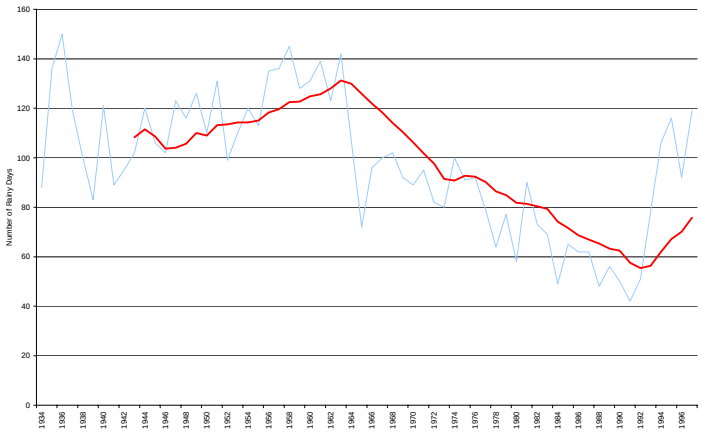


Diagram 7.5: Rainy days per year for JACOBSON FARM from 1934 – 1997. Superimposed 10 year running mean as a red trend line. Source: B. Sturm.

7.2.2 Inter-annual Variability

When studying the temporal distribution of the 10-day data, periods of high rainfall and low rainfall are distinguishable. The 2-year periodic average in Diagram 7.7 reveals an oscillation with a period of approximately 9-11 years. It is noteworthy that this cycle corresponds well with the 11-years sunspot cycle (see Diagram 7.7). This interannual cycle was already mentioned in a paper of Wood and Lovett, who proposed the 11-year sunspot cycle in order to forecast the Sahel droughts (HASTENRATH, 1991: 358). Other authors have observed cycles with similar periodicity for the East African region. Lumb (1966) has recognised the existence of three distinct 10-year cycles over the whole of East Africa (JACKSON, 1979: 65).

These cycles can both be identified in the Diagram 7.6 and Diagram 7.7.

However, it is of importance to note that the presented diagrams were only analysed on a visual non-statistical level. Care should be taken by drawing further conclusions from these statements, as Jackson puts it: "*[...] even experienced workers tending to see cycles in any data, even in a random series*" (JACKSON, 1979: 65).

Further, it should be emphasised that it is very dangerous to extrapolate such trends into the future since there may be no justification for assuming the trends persistence.

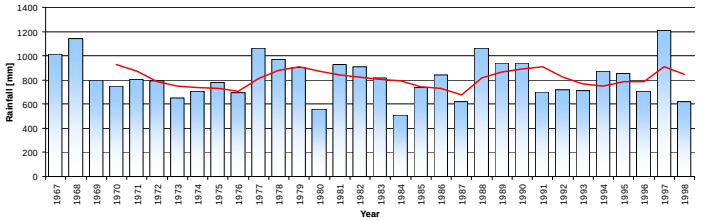


Diagram 7.6: Mean annual rainfall for all 40 gauges from 1967-1998. The red line marks the superimposed linear trend (4 year periodic average). Note that data for the year 1998 covers only the first 5 months! Source: B. Sturm.

7

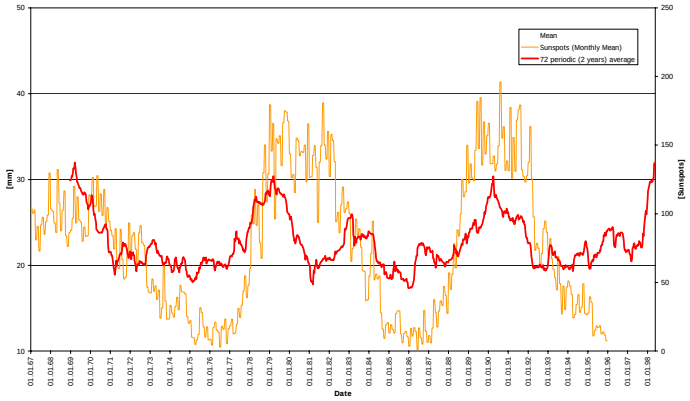


Diagram 7.7: Comparison of Sunspots (orange) and 2-year periodic average of all observed 10-day precipitation data (red). The sunspots are monthly average counts; the 2-year periodic smoothed average is derived from the mean 10-day precipitation data of all 40 stations in the study area. (Source: Sunspot Counts: National Geophysical Data Center, NOAA, Boulder - Colorado, USA, 2000, Precipitation Data: Centre for Development and Environment, Berne, Switzerland, 2000)

It is, however, important to study the temporal distribution of precipitation as this can help us to understand past or ongoing fluctuations or oscillations that are not apparent from the study of mean values alone.

7.3 Farmers Perception and Interviews

One objective of the fieldtrip in Kenya was to obtain information about environmental parameters of the rain gauges. This information is stored in the NRM3 database together with the exact position of the gauge itself. The rain gauge database is described in chapter 2.5 (Precipitation Database) on page 36.

In addition to entirely gauge related information such as exposition, position, type and condition of the gauge, interviews with the observers in charge were held. These interviews were mainly based on the perception of rainfall and the aspects of land use and harvesting regarding to the amount of fallen rain.

The questions of this questionnaire had been defined before the fieldtrip, and are grouped into open questions and questions with a predefined set of answers. The complete answers to the questionnaire can be found in chapter 16 on page 195.

Due to the nature of these questions only farmers or landowners were asked to participate at the interview. At some farms the manager was not available, and it was therefore impossible to conduct an entire interview at all stations. From 80 visited stations 16 observers answered the questionnaire.

Farm	Average	1961	1964	1977	1980	1984	1987	1988	1991	1993	1994	1995	1997
OL DONYO FARM	**634.90**	●1228.10	○426.60	●935.40	○341.90	○291.00	○415.90	●974.00	○450.40	○589.30	●784.50	●679.00	●1025.05
EMBORI FARM	**680.95**			●903.90	○379.10	○413.00	○491.10	●1055.90	○495.70	○465.90	●854.30	681.70	●1054.30
TIMAU MARANIA	**921.99**	●1818.70	●948.50	●1004.10	○784.10	●1116.10	○725.30	●1526.40	○867.10	933.6	●1325.50	●1096.20	●1013.00
OL MYSOR FARM	**625.31**			●1243.60	○331.50	○182.40	○350.50	●1002.10	○537.60	○545.40	●798.60	655.10	●995.21
LOLDOTO FARM	**651.28**	●1114.90	670.80	●842.50	○335.10	○280.80	○419.50	○599.00	○506.10	○573.00	●724.50	○526.00	●1073.00
NICOLSON FARM	**739.59**			●1049.80	○520.90	○560.30	○623.20	●1026.60	○569.40	-	○289.50	●958.70	●1023.60
SUGUROI ESTATE	**710.99**		767.10	●789.10	-	○426.80	○581.20	●839.90	719.70	702.30	●899.30	○638.80	-
OL PEJETA FARM	**713.74**			●1086.10	○480.50	○553.00	○453.20	●1001.80	○610.30	-	-	○693.10	●995.21
ARDENCAPLE FARM	**649.44**	●1337.80	●748.50	●907.50	○294.40	○285.50	○524.50	●897.20	○503.00	○474.60	●841.30	697.80	●1160.30
MOGWONI RANCH	**632.55**			●841.00	○243.00	○187.20	○429.20	●834.80	○423.80	○448.40	●702.90	●693.60	●1399.29
ENASOIT FARM	**582.21**			●903.20	○247.90	○187.00	○446.60	●901.50	○352.50	○446.50	●644.00	537.50	●941.50
MPALA FARM	**522.45**			●838.70	○352.60	○188.20	○335.10	●723.90	○205.80	○359.60	503.90	484.10	●1201.42
LORUKU FARM	**793.80**		○718.80	787.90	○347.00	○455.00	○638.80	●945.10	○601.80	○632.50	●846.70	●973.60	●1363.25

Table 7.1: Overview of the mean annual rainfall in mm for selected farms, compared with the drought, wet assessment of the farmers. A ● indicates a value above average, a ○ indicates a value below the average. Blue colored boxes are years where the farmers remembered rainfalls high above the average (answers to question 3.6). Yellow colored boxes mark years where the farmers remembered droughts

7.3.1 Discussion

Perception of Rainfall in General

The importance of rainfall for the farmers is reflected by the answers for the questions 3.2 and 3.3. As the period of 1997/98 was an El Niño year with extraordinary high amount of rainfalls and with almost no gap between the rainy seasons it is considered as a 'good' year. The reason why farmers assessed this period also as 'normal' and 'bad' was mainly because there was too much rain causing infrastructure damage on the farms.

Relying on their experience and pre-caution they expected the period 1998/99 to be average. But it seems difficult to predict the overall rainfall pattern, as most (11) of the farmers believe that the weather has significantly changed for the last couple of years (they mentioned a time period from 2-6 years). When asked in what way they believe the weather has changed, their answers were always 'the rain is no longer predictable'

or 'the rainfall distribution is somehow changed'.
It is interesting to compare the answers given at question 3.6 where the farmers had to remember years of high rainfall and drought years with the measured rainfall for the mentioned years.

Looking at the years, which ought to be 'drought' years, according to the farmers memories, we can find some inconsistencies. For the year 1994 only Nicolson Farm was suffering from less rain than average, all other farms had higher amounts of rainfall. This is even more interesting as many farmers believe that the weather has changed to the bad since approximately 1992. But the years 1994, 1995 and 1997 were for most areas above the average 'wet' years[78].
The most remembered drought year was 1984, it's important to note, that during the 1980's more drought years occurred than during the 1990's, however this was not mentioned by any farmer.
The reason for this biased assessment may be twofold. On the one hand the farmers do not use their own rainfall recordings[79] in order to compare the amount of water they received in former days - they do not have an objective tool to check their experience -, on the other hand, the farmers may really have experienced 'less' rain for the last years. This second reason stems from the fact that rain-fed agriculture has intensified for the last decade in the Laikipia district, and therefore more rain is needed in order to get the expected yield. Another reason for the experience of under average rainfall years, was mentioned by Jackson, as he states that due to degenerated soils and humus, rainfall becomes less effective (JACKSON, 1979). Because the farmers never used their own records for a proper statistical analysis, their only real measure is the productivity of their own land, hence the divergent observations from measured rainfall and farmers perception.
Table 7.1 on page 140 also shows that there are outstanding years which were remembered by almost all interviewed farmers. These are the 'El Niño' years 1961 (interestingly this year was also mentioned by farmers whose rainfall records did not go back to the 1960s), and 1988. The drought years are 1964 and 1991. These years are also consistent with the measured annual average rainfall.

If the farmers do not use their own rainfall records, do they have another possibility to forecast the rain? This is what question 3.7 and 3.8 wanted to know. The answers were

[78] The year 1996 was slightly below average for most farms.

[79] See also answers to question 4.4

confusing: according to the answers to the question 3.7 not a single farmer has rules to predict the onset of a rainy season. But question 3.8 reveals that they know in fact a set of rules that serve to this purpose. The most popular rule is the bloom of the Thorntree (the species that was mentioned was the 'whistling' Thorntree) 2-3 weeks before the beginning of the onset of the short rains. It was difficult to tell whether the farmers do use this knowledge or if they are simply aware of the existing rules. When asked they seemed to be very interested in the topic, but more out of curiosity than the fact that they may use this knowledge for their own benefit.

Most of the farmers believe that there is an ongoing change in the pattern within the rainy seasons, and also in the distribution of the rainfall periods within the year. Almost all farmers agreed that the onset and length of the rainy seasons are no longer the same as in former years. Independently they mentioned that the overall pattern is now delayed towards the end of the year, but they were unsure if the pattern within the rainy season is being affected as well (Question 3.12).

When asked about an eventually ongoing trend concerning the rainfall, they were at variance, but some of those asked mentioned that other climatic patterns than rain may undergo a change such as wind (1) and temperature (3).

At Hombe Forest Station observer Miriam Kamau made some interesting remarks concerning the predictability of the rains: *„In the old days the rains were constant and heavy, now they are less predictable, and the weather is more hotter, that's the reason why the trees grow faster!"*. Miriam Kamau sees a connection between the deforestation and the climate change: *„The old folks say: The reason why the weather changed is because of the cutting of trees."* (Personal interview by the author).

Land-use and Harvesting

As an overall assessment of the answers given to the questions 4.1 – 4.7 one can state that the farmers are aware of the impact any change of the water balance may have on their land, but they don't feel that they have to respond with appropriate means so far. Although they all measure the daily rainfall for generations, none of the asked farmers are actually using it as an aid for the start of the sowing or for the predictability of the onset of the rainy seasons. This can be explained by the fact that it is not sufficient enough to know only the amount of rainfall from one point in space in order to obtain a reliable image of the

7

rainfall pattern of an entire farm, this can only be achieved by including other measurements of other farms. Many of the visited farmers showed a high interest in the rainfall maps that were produced by the NRM[3] project so far. This leads to the conclusion that with more data available the farmer would take appropriate measures to be prepared for the impact that an eventually changed rainfall pattern might have on their soils productivity.

7.4 Conclusions

In this chapter the results gained from the estimation methods and models were presented.
Compared with previous works conducted in this area, the estimated rainfall maps show a high level of detail and can be used for further planning in the fields of hydrology and agroclimatology.
The following points summarise the findings from this chapter:

1. The geographical image of the onset and end of the rainy seasons can be obtained with much greater accuracy than in earlier studies from the estimated maps.
2. The quality and accuracy of the estimated rainfall distribution pattern can directly be compared with previously manually drawn rainfall maps, and are in good correspondence.
3. Inter-annual cycles should be further observed on a more detailed level, but first analysis reveal an increased variability of the general rainfall pattern for the region. This increased variability is also supported through the assessment of the farmer interviews conducted during the fieldtrip. Further evidence exists that inter-annual oscillations are related to sunspot activity, but thorough analysis on this field would be necessary.
4. Intra-annual oscillations can be studied with more accuracy with the use of the presented maps and models.

8 Outlook and Discussion

8 Outlook and Discussion

In this final chapter an outlook and a short discussion of the presented work shall take place.

8.1 Application of the Results

The presented results can directly be used for other studies in the Upper Ewaso Ng'iro Basin. Because the interpolated rainfall maps can easily be exported to other formats, such as the Arc/Info grid format, the results are available for further geostatistical analysis[80].
On the CD-ROM, included in this study, all presented maps and rainfall data can be found in various formats, and may be used for further studies.

8.2 Application of the Methods

Because of the step-by-step explanation of the applied method, it should not take too much additional work to apply the herewith-presented methods and models for other tropical regions than the study area. However, care should be taken when analysing non-tropical rainfall regimes, as there may be more sources of errors involved than the ones described in this work[81].
Because the rainfall maps are computed using the free GIS System GRASS, it is possible to adapt the introduced scripts and tools for any particular precipitation estimation task without much trouble. It shall also be emphasised that the used GIS system and the applied modules are freely available, and therefore free of charge, hence making a distribution of the maps and GIS database easy. This is of particular interest for low-budget projects located within developing countries.

8.3 Further Research

In the following we will point to some issues where further research would be necessary or could help to understand the dynamics of a changing tropical rainfall regime.

8.3.1 Topographical Dependent Estimations

As presented in the result section (chapter 7, Presentation of Results on page 119) an enhanced interpolation quality can primarily achieved by a denser network of gauging stations. By comparing the long- and the short-

[80] This has already happened as Hilde Vanleeuwe, a PhD researcher in Kenya, uses the interpolated rainfall maps for her study on elephant habitat management around the Mt. Kenya. (after personal email communication with H. Vanleeuwe, October 2001)

[81] Such as snow and different atmospheric circulation conditions, or different topographic conditions that influence the quality of the estimated precipitation. The biggest source of error is the doubtless the quality of the used raw data.

term interpolations it could be shown that, especially in mountainous topography, the rainfall pattern was more differentiated where more stations were used for the estimates.
But, especially in developing countries, we do not find many rainfall gauges located on mountains, and hence there is not much detailed information on mountainous rainfall distribution available.
However, there is one way to overcome this lack of additional gauges. Because precipitation is altitude dependent (see chapter 2.1 Condensation and Cloud Formation on page 25) the topographical information can be used as additional information for the spatial rainfall interpolation.
There are various studies available discussing the orographic dependent interpolation of rainfall. For instance Dingman et al. used the Krigging method to orographically estimate rainfall (DINGMAN et al., 1988), or Daly et al. used a digital elevation model to derive facets, which are then used to establish local altitude regression models in order to produce accurate precipitation maps (DALY et al., 1994).
At the University of Berne at the KLIMET group, a software package had been developed by Dimitrios Gyalistras in order to estimate rainfall depending on topography. This package (ClimShell) was successfully used for the rainfall map production based on the inverse distance method. However, the lately released GRASS 5.0 version has already 3D capabilities and the new 3-dimensional spline interpolation module has been introduced to this new release, thus making 3-dimensional rainfall estimates possible.
For a further study it would be of interest to see how well these new methods could be used for the estimate of rainfall with respect to a given topography.

8.3.2 Time-Series Analysis

It is of importance to identify seasonal variations or underlying trends in a rainfall time series, as this may help to understand any eventually ongoing change in the rain regime. Due to the restriction of this work to the development of a quick and easy spatial interpolation model, it wasn't possible to test the data for such underlying trends. Although chapter 7.2 (Temporal Distribution on page 136) revealed that there are in fact indications that the study region suffers from an increased variability in rainfall intensity, this finding could not be confirmed through the use of sound statistical methods. However, it would be of great interest to apply harmonic (Fast Fourier Transformation) or spectral analysis on the complete time series. Although these techniques require data covering a considerable period of time, it may be possible to run a Fast Fourier Analysis on the data ob-

served at gauges having long records such as JACOBSON FARM. Harmonic analysis entail the fitting of a mathematical sine or cosine function to observed data, and assume the 'stationarity' of the time series: that is, that the long term mean is constant.

If monthly variation within a calendar year is being studied, the first (or 'fundamental') harmonic has a period equal to the calendar year. The second harmonic corresponds to half the fundamental period, and so on.
According to Summer this technique may be "*used where regular fluctuations are being sought and where data are available at regular intervals, and is more generally applied when looking at monthly precipitation within the average year. [...] It may therefore be used to indicate broad seasonal, or other regular, trends.*" (GRAHAM, 1988: 357)

Applying filters to time series can also help to reveal dependencies between different environmental parameters. The goal of the application of filters is to exclude the more obvious and well-known periodicities, and enhance other periodicities present in the data. Graham shows an example where a correspondence between differently filtered running means of long-term precipitation data reveals a connection between sunspot activity occurrence and the rainfall periodicity[82] (GRAHAM, 1988: 359).

8.3.3 The Perception of Variability

We have seen that rainfall varies not only from place to place but also with time. This variation can have major implications if the growing season is affected by a changing rainfall pattern. As Jackson puts it: *"Traditional peasants activities were well adapted to the environmental conditions but changes, often induced from outside, can lead to a breakdown in the system or make it more difficult for communities to adapt to harsh conditions such as a series of dry years, particularly in semi-arid areas."* (JACKSON, 1979: 33).
Because the effects of rainfall characteristics are not only determined by climatic and atmospheric conditions but also by other physical conditions like soil type and relief, moreover economic, social and political conditions are playing an important role in the perception of the effects of rainfall.
It is of importance to be aware of the fact that humans who rely on a subsistence strategy that is strongly dependent on rainfall notice more the effect of rainfall than the gauged amount of rain. If these two differ-

[82] See also chapter 7.2.2 (Inter-annual Variability) on page 138.

ent views of rainfall are not mixed and properly understood, we may find other explanations for a dramatically changed rainfed agriculture than an eventually shifted rainfall regime.

According to the farmers perception in Kenya (chapter 7.3 (Farmers Perception and Interviews) on page 139), rainfall is less predictable and no longer reliable. Jackson (1979) has an explanation for such a view which is worth noticeable as in areas with fairly low rainfall farmers think that rainfall is decreasing, whereas often it is simply that the same rainfall has become less effective: *"As the soil loses its humus content it becomes less moisture retentive, and hence rainfall is less effective and nutrients become less available to plants."* (JACKSON, 1979: 33).

The variability of rainfall is crucial in marginal areas when a period of 'good' years is followed by long droughts. The intensification in the cultivation of land during years of 'good' rainfall conditions is especially dangerous and perhaps the main cause of desertification. Here again perception is a key issue: a study of the perception of pastoral farmers in western Queensland revealed, that the farmers tend to describe wetter years as 'normal', and the drier years as 'bad' (JACKSON, 1979: 52). If the focus of their attention is set on 'normal' years, then they will hardly develop strategies for the 'bad' years, even if their ancestors knew of such strategies, but the changing political, social and economic situation rendered those former strategies as no longer practical. (See also 7.3 (Farmers Perception and Interviews) on page 139)

This later point emphasises the need for an interdisciplinary strategy, which is not only based on the results of physical empirical observations and studies, but also on the understanding of ongoing changes in the socio-economic structures of an afflicted society.

9 Bibliography and Index

9 Bibliography and Index

9.1 Bibliography

Alexandersson, Hans, 1986. 'A Homogeneity Test Applied to Precipitation Data'. Uppsala, *International Journal of Climatology,* 6.

Ali, Alaa, 1998. 'Nonparametric Spatial Rainfall Characterization Using Adaptive Kernel Estimator', *Journal of Geographic Information and Decision Analysis,* Vol 2.

Arkin, Ph., et al., 1985. 'Review of Requirements for Area-Averaged Precipitation Data, Surfaces-Based and Space-Based Estimation Techniques, Space and Time Sampling; Accuracy and Error; Data Exchange'. Boulder, USA, *World Meteorological Organization (WMO),* WMO/TD-No. 115.

Bahrenberg, Gerhard, Giese Ernst, Nipper Josef, 1990. 'Statistische Methoden in der Geographie 1'. Stuttgart, *B.G. Teubner,* 1

Bahrenberg, Gerhard, Giese Ernst, Nipper Josef, 1992. 'Statistische Methoden in der Geographie 2'. Stuttgart, *B.G. Teubner,* 2.

Bastin, G., Lorent, B., Duque, C., Gevers, M., 1984. 'Optimal estimation of the average rainfall and optimal selection of raingauge locations', *Water Resources Research,* 20(4).

Beek, E. G., Stein A., Janssen L. L. F., 1992. 'Spatial variability and interpolation of daily precipitation amount', *Stochastic Hydrology and Hydraulics,* 6.

Berger, Peter, 1987. 'Niederschlag und Agroklimatologie des Laikipia-Plateaus (Kenya)'. Brügglen, *Selbstverlag, Peter Berger.*

Berger, Peter, 1989. 'Rainfall and Agroclimatology of the Laikipia Plateau, Kenya'. Bern, *Geographica Bernensia,* A7.

Berndtsson, Ronny, 1988. 'Spatial Hydrological Processes in a Water Resources Planning Perspective - An Investigation of Rainfall and Infiltration in Tunisia'. Lund, *Univerisity of Lund.*

Bigg, G. R., 1991. 'Kriging and intrare-gional rainfall variability in England', *International Journal of Climatology,* 11.

Cappel, M., Kalb, M., Schmidt, H., 1988. 'Klimatologische und statistische Grundlagen als Erläuterungen für Klimabearbeitungen'. Offenbach am Main, *Deutscher Wetterdienst / Zentralamt.*

Casella Catalogue 1983/935', 1983, London, *Casella London Ltd.*

Creutin, J. D., Obled C., 1982. 'Objective analysis and mapping techniques for rainfall fields: an objective comparison', *Water Resources Research,* 18(2).

Daly, C, Neilson R.P., Philips, D.L., 1994. 'A statistical-topographic model for mapping climatological precipitation over mountainous terrain', *Journal of Applied Meteorology,* 33.

Dana, Peter H., 1994. 'Global Positioning System Overview'. www.utexas.edu/detps/grg/gcraft/notes/gps/gps.html, *University of Texas.*

Dingman, S. L. D., Seely-Reynolds M., Reynolds III. R.C., 1988. 'Application of kriging to estimating mean annual precipitation in a region of orographic influence', *Water Resources Bulletin,* 24/2.

Feyerabend, Paul K., 1983. 'Wider den Methodenzwang'. Frankfurt am Main, *Suhrkamp Taschenbuch Verlag.*

Flury, Manuel, 1987. 'Rain-Fed Agriculture in the Central Division Laikipia District, Kenya'. Bern, *Geographica Bernensia,* A6.

Freydank, Erhard, 1986. 'Die quantitative Berücksichtigung der Zusammenhänge zwischen Niederschlagsverteilung, Orographie und Wetterlage bei der räumlichen Interpolation von Niederschlägen'. Dresden, *Fakultät für Bau-, Wasser- und Forstwesen,* 1.

Genton, Marc G., Reinhard Furrer 1998. 'Analysis of Rainfall Data by Simple Good Sensse: is Spatial Statistic Worth the Trouble?'. Internet, *Journal of Geographic Information and Decision Analysis,* Vol. 2.

Gichuki, Francis N., Hanspeter Liniger, Gudrun Schwilch, 1998. 'Knowledge about Highland - Lowland Interactions: The Role of a Natural Resource Information System'. Nairobi, *Eastern and Southern Africa Geographical Journal,* Vol. 8, page 5.

Golubev, V.S., 1986. 'On the Problem of Actual Precipitation Measurements at the Observation Site'. Zurich, *Verlag Geographisches Institut ETH Zürich,* Vol. 23: 61-64.

Goovaerts, P., 2000. 'Geostatistical approaches for incorporating elevation into spatial interpolation of rainfall'. Ann Arbor, *Journal of Hydrology,* Vol. 228(2000), 113-129.

Grote, Andreas, 2001. 'Tux für die Welt - Open Source macht Entwicklungsländer unabhängiger'. Hannover, *Verlag Heinz Heise GmbH & Co KG,* Heft 10, 104-106.

Gyalistrias, D. and Fischlin, A., 1999 'Towards a General Method to Construct Regional Climatic Scenarios for Model-Based Impacts Assessments', *Petermanns Geographische Mitteilungen,* 143, 1999/4, pp. 251-264.

9

Hanssen-Bauer, Inger, Foerland, Eirik J., 1994. 'Homogenizing Long Norwegian Precipitation Series'. Oslo*, American Meteorological Society,* 7.

Hastenrath, Stefan*,* 1991. 'Climate Dynamics of the Tropics'. Dordrecht*, Kluwer Academcic Publishers.*

Heinrich, Uwe*,* 1992. 'Zur Methodik der räumlichen Interpolation mit geostatistischen Verfahren: Untersuchungen zur Validität flächenhafter Schätzungen diskreter Messungen kontinuierlicher raumzeitlicher Prozesse'. Kiel*, Deutscher Universitäts Verlag.*

Henderson-Sellers, Ann, Robinson, J., Peter, 1996. 'Contemporary Climatology'. London*, Longman Group Limited.*

Herweg, Karl, Ostrowski, Manfred K., 1997. 'The Influence of Errors on Erosion Process Analysis'. Bern*, University of Berne, Switzerland,* Research Report 33.

Hevesi, J. A*,* Flint A.L*,* Istok J.D., 1992. 'Precipitation estimation in mountainous terrain using multivariate geostatistics. 2. Isohyetal maps'*, Journal of Applied Meteorology,* 31.

Hevesi, J. A.*,* Flint A.L*,* Istok J.D., 1992. 'Precipitation estimation in mountainous terrain using multivariate geostatistics. 1. Structural-analysis'*, Journal of Applied Meteorology,* 31.

Hutchinson, M. F.*,* 1997. 'A Locally Adaptive Approach to the Interpolation of Digital Elevation Models'*, The World Wide Web.*

Hutchinson, M. F.*,* Jennifer Kesteven 1997. 'Spatial Modelling of Climatic Variables on a Continental Scale'. Santa Fe*, http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/kesteven_jennifer/jlkpaper.html.*

Hutchinson, M. F.*,* 1999. 'Modelling Spatial and Temporal Variability of Climate and Terrain'. Canberra*, http://incres.anu.edu.au/hydweb/hutch/hutch2.html.*

Hutchinson, M. F.*,* 1995. 'Stochastic space-time weather models from ground-based data'*, Agricultural and Forest Meteorology,* 73.

Hutchinson, M. F.*,* 1995. 'Interpolating mean rainfall using thin plate smoothing splines'*, International Journal of Geographical Information Systems,* 9.

Hutchinson, M. F.*,* 1998. 'Interpolation of Rainfall Data with Thin Plate Smoothing Splines: II Analysis of Topographic Dependence'. Internet*, Journal of Geographic Information and Decision Analysis,* Vol. 2.

Hutchinson, M. F., 1998. 'Interpolation of Rainfall Data with Thin Plate Smoothing Splines: I Two Dimensional Smoothing of Data with Short Range Correlation'. *Journal of Geographic Information and Decision Analysis,* Vol. 2.

Isaaks, Edward H, R., Mohan Srivastava, 1989. 'An Introduction to Applied Geostatistics'. New York, *Oxford University Press,* 1.

Jackson, Ian Joseph, 1974. 'Inter-Station Rainfall Correlation under Tropical Conditions'. Giessen, *Lenz - Verlag Giessen,* Catena Vol. 1, 235-256.

Jackson, Ian Joseph, 1979. 'Climate, Water and Agriculture in the Tropics'. New York, *Longman Group Limited.*

Jensen, Holger, 1989. 'Räumliche Interpolation der Stundenwerte von Niederschalg, Temperatur und Schneehöhe'. Zürich, *Verlag Geographisches Institut ETH Zürich,* 35.

Kiteme, Boniface P., Urs Wiesmann, Erwin Künzi, Jospeh M. Mathuva, 1998. 'A Highland-Lowland System under Transitional Pressure: A Spatio-Temporal Analysis'. Nairobi, *Eastern and Southern Africa Geographical Journal,* 8.

Kreuels, R. K, Breuer L.J., 1985. 'Wind Influenced Rain Gauge Errors in Heavy Rain'. Zurich, *Verlag Geographisches Institut ETH Zürich,* pages 105-109.

Kumm, Helmut, 1983. 'Vergleichsmessungen mit windgeschützten und ungeschützten Niederschlagsmessgeräten'. Offenbach am Main, *Deutscher Wetterdienst / Zentralamt,* 162.

Lauer, Wilhelm, 1993. 'Das Geographische Seminar: Klimatologie'. Braunschweig, *Westermann Schulbuchverlag GmbH.*

Lebel, T., Bastin G., Obled C., Creutin J.D., 1987. 'On the accuracy of areal rainfall estimation: a case study', *Water Resources Research,* 23.

Liniger H.P. and Thomas D.B. 1998. 'GRASS: Ground cover for the Restoration of the Arid and Semi-arid Soils.' *Advances in GeoEcology* 31, 1167-1178, *CATENA Verlag, Reiskirchen.*

Liniger H.P., MacMillan Lindsay C., Schwilch Gudrun, 1998. 'Scarce Water: Exploring Resource Availability, Use and Improved Management'. Nairobi, *Eastern and Southern Africa Geographical Journal,* Vol. 8, page 15.

McMillan, Lindsay, Gudrun Schwilch 1997. 'Infilling Monthly Rainfall Data Gaps for period 1973-1995'. Bern, *Centre for Development and Environment, CDE.*

9

Menz, Gunter, 1996. 'Niederschlag und Biomasse in den wechselfeuchten Tropen Ostafrikas: neuere Methoden zur quantitativen Erfassung klimaökologischer Raumparameter aus digitalen Satellitendaten'. Stuttgart, *Franz Steiner Verlag Stuttgart.*

Microsoft Encarta '99, Weltatlas (CD-ROM), 1999, Redmond, *Microsoft Corporation*.

Mitásová, Helena, Mitás Lubos 1993. 'Interpolation by Regularized Splines with Tension: I. Theory and Implementation'. Champaign, *Mathematical Geology,* Vol.25, No.6, 1993, 641-655.

Mitásová, Helena, Mitás Lubos, 1993. 'Interpolation by Regularized Splines with Tension: II. Application to Terrain Modelling and Surface Geometry Analysis'. Champaign, *Mathematical Geology,* Vol. 25, No. 6, 1993, 657-669.

Moberg, Anders, Hans Alexandersson 1996. 'Temperature Variations in Sweden Since the 18th Century'. Stockholm, *The Department of Physical Geography,* No. 5.

Mwiti, Sammy, 1998. 'Ngushishi a poor soul's paradise'. Nairobi, *East African Standard,* 8th October 1998.

Nespor, Vladislav, 1996. 'Investigation of Wind-Induced Error of Precipitation Measurements Using a Three-Dimensional Numerical Simulation'. Zürich, *Verlag Geographisches Institut ETH Zürich,* 63.

Neteler, Markus, 1998. 'Introduction to GRASS GIS Software'. Hannover, *www.geog.uni-hannover.de/phygeo/grass/gdp/neteler/.*

Niederer, Peter, 2000. 'Classification and Multitemporal Analysis of Land Use and Land Cover in the Upper Ewaso Ng'iro Basin (Kenya) using Satellite Data and GIS'. Bern, *University of Berne, Department of Geography.*

Prathapan, B., 2000. 'Kerala evolving new vistas in e-governance (25.05.2000)'. www.indiaserver.com/thehindu/, *The Hindu & Tribeca Internet Initiatives Inc.*

Rhea, J. Owen, 1996. 'An Objective Orographically-Based QPF aid for California'. Sacramento, *Western Regional Technical Attachement,* No. 96-02.

Richter, Dieter, 1995. 'Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Messfehlers des Hellmann-Niederschlagsmessers'. Offenbach am Main, *Deutscher Wetterdienst / Zentralamt,* 194.

Rudolf, Bruno, 1995. 'Die Bestimmung der zeitlich-räumlichen Struktur des globalen Niederschlags'. Offenbach am Main, *Deutscher Wetterdienst / Zentralamt,* 196.

Saveliev, Anatoly A., Svetlana S., Mucharamova; Gennady, A., Piliugin 1998. 'Modeling of the Daily Rainfall Values Using Surface Under Tension and Kriging'. Internet, *Journal of Geographic Information and Decision Analysis,* Vol. 2.

Schwartz, Randal, L., Christiansen, Tom, 1997. 'Learning Perl, Second Edition'. Sebastopol, *O' Reilly & Associates, Inc.*

Sevruk, Boris, 1986. 'Corrections of Precipitation Measurements'. Zürich, *Verlag Geographisches Institut ETH Zürich,* 23.

Sharon, David, 1974. 'On the Modelling of Correlation Functions for Rainfall Studies'. Jerusalem, *Journal of Hydrology,* Vol. 22, 219-224.

Sturm, Bernhard, 1997. 'IMPALA - NRM3 Database Query Tool (Rainfall and Discharge Analysis Application)'. Bern, *University of Berne, Switzerland.*

Summer, Graham, 1988. 'Precipitation, Process and Analysis'. Chichester, *John Wiley & Sons Ltd.*

Tabios, III G. Q., Salas J.D., 1985. 'A comparative analysis of techniques for spatial interpolation of precipitation', *Water Resources Bulletin,* 21, 3.

Traufeter, Gerald, 2001. 'Die Launen der Sonne'. Hamburg, *Spiegel - Verlag,* 23/2.6.01.

Trenberth, Kevin E., et al., 1995. 'Climate System Modeling'. Cambridge, *Cambridge University Press.*

Wilhelm, Friedrich, 1993. 'Das Geographische Seminar: Hydrogeographie'. Braunschweig, *Westermann Schulbuchverlag GmbH.*

Wilson, David L., 1998. 'Global Positioning System Accuracy'. The World Wide Web, *www.erols.com/dlwilson/gps.html.*

Winiger, M., U. Wiesmann, J.R. Rheker, 1989. 'Mount Kenya Area: Differentiation and Dynamics of Tropical Mountain Ecosystem'. Bern, *Geographica Bernensia,* A8.

9

9.2 Index

9

1 Appendix: A - Tables

10 Rainfall Gauges

All rainfall recording stations stored in the NRM[3] database as of 1.1.2000.

ID	Station	KMD-No	Longitude[83]	Latitude	Alt	UTM	Spheroid	Protection?[84]	Remarks[85]
1	ARCHERS POST	8937035	351624.2825	70807.01226	860	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	NRM station, well maintained
2	ARDENCAPLE FARM	8937034	306006.8278	9696.367758	2162	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	
4	BORANA	n/a	308863.8988	25159.66275	2035	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Manager was not in, therefore no interview could be conducted. Gauge is only for estimates because of the open plastic construc-tion...
5	CASTLE FOREST STN	9037115	311885.8796	9955135.885	1906	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Observer was not around at visit, therefore no interview.
6	CHEHE FOREST STN	9037075							
7	CHOGORIA FOREST STN	9037123	343375.8847	9973859.809	1713	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	
16	CHOGORIA GATE	9037219							
8	CHUKA FOREST STN	n/a	344923.3853	9965172.171	1474	37M	ARC1960	Better protected (d = 5 x ObstHeight)	1994 was a wet year. Observer seems to be a reliable person.
9	COLCHECCIO	8936060	255409.1604	68804.30743	1831	37N	ARC1960	Better protected (d = 5 x ObstHeight)	Station seems to be reliable
10	DOL DOL DAO	8937033	295573.8604	43649.45	1820	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Rain gauge looks completely abandoned, terrible condition, when visited the gauge was filled with water...
95	EL KARAMA	n/a	266605.4076	21860.41486	1797	37N	ARC1960	Better protected (d = 5 x ObstHeight)	Rain gauge consists of a bottle and a simple funnel of an old Casella gauge -> Problems may occur due to too much loss caused by evaporation.
13	EMBORI (NRM)	n/a							
64	EMBORI B	n/a	312492.1808	5386.576939	2670	37N	Unknown		
76	EMBORI C	n/a	312492.1629	4612.469623	2740	37N	Unknown		

[83] Geographical position measured by means of averaged GPS measurements (minimum 60 samples) during the fieldtrip in 1998.
[84] Description of protection. Observed and analysed during the visit in 1998.
[85] Remarks made during the visit in 1998.

ID	Station	KMD-No	Longitude[83]	Latitude	Alt	UTM	Spheroid	Protection?[84]	Remarks[85]
12	EMBORI FARM	n/a	316071.1419	7766.903296	2691	37N	ARC1960	Undisturbed exposition (d = 10-20 x ObstHeight)	
14	EMBU MET STN	9037202	328296.5892	9944482.134	1540	37M	ARC1960		Not very cooperative, at visit no data collection was possible, and no interview was allowed...
15	ENASOIT FARM	n/a	285614.1558	26288.02208	1912	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Rain gauge is just for estimates as it is an open plastic model
17	GATHIURU FOREST STN	9037097	290496.9784	9989256.094	2330	37M	ARC1960	Better protected (d = 5 x ObstHeight)	Very good condition of RG, well maintained.
19	GITUNDAGA FOREST STN	n/a	200919.7318	17787.20451	2346	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	No, this is not a reliable station
20	HOMBE FOREST STN	9037069	290267.591	9961541.78	1991	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	According to the observer they do every day readings, but the data shows some consecutive dates with the same amount of measured rain! The rain gauge is located in a field where maize grows (or used to grow)... (Imagine what sort of impact this may have on the measured rain...) If this station gets a quality 'one' assessment: Please be careful with this! Observer (MIRIAM KAMAU) is very well informed in what way the wheather seemes to have changed.
21	IRANGI FOREST STN	9037077	331263.8319	9961897.335	2009	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Measuring glass is leaking! Readings are done using the container!
22	ISIOLO DAO	9037003	342591.8084	39500.94954	1163	37N	ARC1960	Good protected (d = 2-5 x ObstHeight)	Gauge is good protected against any wind influence because of the surounding buildings. Observer was not around, so no interview was conducted.
18	ITUURI (U.M.-LAND)	n/a	315831.542	9998641.474	3575	37M	Unknown		
23	JACOBSON FARM	n/a	282214.5145	9995527.142	1905	37M	ARC1960	Better protected (d = 5 x ObstHeight)	No glass for the gauge, the rain is directly collected by the cylinder... When grass is high there maybe some additional rain drops splashed into the gauge
24	JUNCTION (EWASO NAROK)	8936065	260384.3345	56118.81891	1541	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	No interview, because observer was not around... Station is in a very bad condition, looks as if abandoned...
25	KABARU FOREST STN	9037120	294393.1083	9969216.133	2237	37M	ARC1960	Undisturbed exposition (d = 10-20 x ObstHeight)	Station seems to be well maintained, although no weekend readings are being performed.
26	KAGURU	9037214	351132.788	9990738.74	1528	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	No interview, because observer was not in. Grass protects the rain gauge, the station does not seem to be very well maintained.

①

ID	Station	KMD-No	Longitude[83]	Latitude	Alt	UTM	Spheroid	Protection?[84]	Remarks[85]
28	KALALU (NRM)	8937111	295548.9549	9332.621825	2080	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	NRM3 station
29	KAMWAKI FARM	n/a	295459.1225	15008.10236	1980	37N	Unknown		Difficult to get any data about the rain gauge, as there is a new management since August 1998. Since this date no rainfall records had been kept. At visit it was impossible to see the rain gauge ot to conduct an interview with the former manager. New manager is very co-operative: Hartmut Rottcher - Kamwaki - 22688 - hartmut@africaonline.co.ke
31	KARURI (NRM)	n/a	313829.8243	3225.581535	2976	37N	ARC1960	Undisturbed exposition (d = 10-20 x ObstHeight)	
78	KARURI B	n/a					Unknown		
27	KINAMBA MOW	n/a	201815.8679	46867.18517	2097	37N	ARC1960	Undisturbed exposition (d = 10-20 x ObstHeight)	A very reliable and good maintained station.
32	KISIMA FARM	n/a	323860.7513	12663.72408		37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Not a very reliable station! Readings over weekends are noted on mondays!
33	LAMURIA MET STN	9036260	263186.0693	9986005.94	1876	37M	ARC1960	Undisturbed exposition (d = 10-20 x ObstHeight)	During visit, there was a constant wind from E (21.10.98)
11	LARIAK FOREST STN	8936092	200582.1682	36102.93245	2020	37N	ARC1960	Undisturbed exposition (d = 10-20 x ObstHeight)	They had a problem measuring rainfall in 1990.
35	LOGILADO (NRM)	n/a	309931.4231	4502.190954	2640	37N	Unknown		
81	LOGILADO B	n/a							
34	LOLDAIGA FARM	n/a	290891.7226	23732.1734	2160	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Gauge is in good condition
36	LOLDOTO FARM	8937048	279438.733	19340.93875	1790	37N	ARC1960	Better protected (d = 5 x ObstHeight)	
37	LOLMARIK FARM	n/a	307781.7926	11497.23706	2304	37N	ARC1960	Better protected (d = 5 x ObstHeight)	Gauge is for estimates only...
38	LORUKU FARM	n/a	286629.4629	9998806.799	2040	37M	ARC1960	Better protected (d = 5 x ObstHeight)	This is not a very reliable station, although the readings are done well, but the gauge is only for estimates (open plastic gauge)
30	LOWER KAMWETI	9037129							
39	MARALAL DC	8836000	244619.8745	122312.8899		37N	Unknown		
42	MARIENE CRS	8937124	349303.7469	9999065.132	1670	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Station is very well maintained
44	MATANYA (NRM)	9036356	272189.0596	9993442.602	1840	37M	Unknown		

ID	Station	KMD-No	Longitude[83]	Latitude	Alt	UTM	Spheroid	Protection?[84]	Remarks[85]
96	MATANYA B	n/a					Unknown		
97	MATANYA C	n/a							
98	MATANYA D	n/a							
99	MATANYA E	n/a							
45	MERU FOREST STN	8937038	346982.919	6439.138754	1756	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Measuring glass was broken. Station is not in a good condition. Rain gauge was contaminated with pigeon dung.
47	MOGWONI RANCH	n/a	275763.2074	25543.06045	1760	37N	ARC1960	Better protected (d = 5 x ObstHeight)	
48	MPALA FARM	n/a	266641.2286	36040.93455	1690	37N	ARC1960	Better protected (d = 5 x ObstHeight)	
50	MUGIE RANCH	n/a	231969.4108	78649.87705	1860	37N	ARC1960	Better protected (d = 5 x ObstHeight)	No interview as farmer was not in...
51	MUKENYA FARM	8936059	257178.9895	27211.96024	1802	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Rain gauge is just for estimates (open plastic model)
52	MUKOGODO (NRM)	8937109	284539.7291	42749.59131	1740	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Gauge is in a good condition, regularly checked by NRM3 staff members
100	MUKOGODO B	n/a							
101	MUKOGODO C	n/a							
102	MUKOGODO D	n/a							
103	MUKOGODO E	n/a							
53	MUNYAKA (NRM)	n/a	283657.9129	9980282.478	2070	37M	Unknown		
54	MURINGATO FOREST STN	9036251	271022.5898	9955655.064	1720	37M	ARC1960	Better protected (d = 5 x ObstHeight)	This is not a very good station. During El Niño it was not possible to read the gauge, as the whole area was flooded. Sometimes kids put water into the gauge...
55	MUTARA ADC FARM	8936014	241943.2597	12128.69671	1941	37N	ARC1960	Better protected (d = 5 x ObstHeight)	Good station, gauge is in a good condition. There is even a sort of a wind fence...
56	MWEA IRRIGATION SCHEME	9037112	315496.8625	9923172.03	1175	37M	ARC1960	Undisturbed exposition (d = 10-20 x ObstHeight)	
40	NANYUKI FOREST STN	9037156	293596.5156	9992492.392	2337	37M	ARC1960	Better protected (d = 5 x ObstHeight)	RG is in a very good condition, reliable observer
57	NANYUKI KAF	8937022	280873.6357	4724.146922	1860	37N	Unknown		
59	NARO MORU FG POST	9037064	289002.1567	9980504.32	2195	37M	Unknown		

①

ID	Station	KMD-No	Longitude[83]	Latitude	Alt	UTM	Spheroid	Protection?[84]	Remarks[85]
60	NARO MORU FOREST STN	9037138	292602.1782	9976596.395	2364	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Rain gauge is leaking! The whole rain gauge is in a bad condition!
61	NARO MORU GATE STN	9037149	293829.0187	9981023.89	2420	37M	ARC1960	Better protected (d = 5 x ObstHeight)	Maintained by two KWS park rangers, there is only a manual rain gauge installed
62	NARO MORU MET STN	9037217	301090.0318	9981442.206	3050	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	This station is being maintained by two park rangers of the KWS. The station is in a very good condition.
58	NARO MORU MOOR-LAND (R2)	n/a	304420.0306	9981854.195	3771	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	There are two stations: 1 Hellman and 1 Belfort. The Hellman is currently not running, readings are therefore only done manually when charts of the Belfort are being changed (BS/23.9.98)
63	NDARAGWA FOREST STN	n/a	224981.3816	9993021.676	2319	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	
43	NGELESHA	8936102	198358.1549	48103.06699	2060	37N	Unknown		
65	NGENIA (NRM)	n/a	299578.4499	9589.298733	2110	37N	Unknown		
104	NGENIA B	n/a	299952.3108	9622.501362	2225	37N	ARC1960	Better protected (d = 5 x ObstHeight)	Now the station seems to be reliable, but from 1992-1994 there was a period when station was not reliable. From 1994-1995 station was closed.
105	NGENIA C	n/a							
106	NGENIA D	n/a							
107	NGENIA E	n/a							
66	NICOLSON FARM	n/a	280194.1152	9990523.35	1950	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	
3	NORTH MARMANET FOREST STN	n/a	205425.1759	15715.07532	2310	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	It seems as station is more or less reliable...
67	NYERI MOW	9036017	271638.1602	9954256.784	1811	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Well maintained station, observer is reliable, station is in a good condition.
68	OL ARABEL FOREST STN	8936086	194353.309	30686.70537	2100	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	
74	OL BOLOSAT FOREST STN	9036138	203375.3302	9994276.018	2446	37M	ARC1960	Undisturbed exposition (d = 10-20 x ObstHeight)	Station does not seem to be very reliable...
69	OL DONYO FARM	n/a	309910.7418	10674.95601	2380	37N	ARC1960	Better protected (d = 5 x ObstHeight)	Measuring glass is broken.
70	OL JOGI FARM	n/a	270990.5395	34572.77553	1790	37N	ARC1960	Better protected (d = 5 x ObstHeight)	They have a somebody at Ol Jogi who is concerned with environmental and naturla resource questions.

ID	Station	KMD-No	Longitude[83]	Latitude	Alt	UTM	Spheroid	Protection?[84]	Remarks[85]
71	OL JORO OROK FTC	9036326	208445.5339	9999168.282	2506	37M	ARC1960	Undisturbed exposition (d = 10-20 x ObstHeight)	This is not a reliable station: rain gauge is in a bad condition as the cylinder is leaking...
72	OL MYSOR FARM	8936049	239206.3103	46166.69162	1820	37N	ARC1960	Undisturbed exposition (d = 10-20 x ObstHeight)	
73	OL PEJETA FARM	8936066	255941.5709	7052.59237	1877	37N	ARC1960	Better protected (d = 5 x ObstHeight)	
75	ONTULILI FOREST STN	8937040	296237.9524	2954.15493	2130	37N	Unknown		
46	PYRAMID OL JOGI	n/a	285764.0548	33716.85968	1852	37N	ARC1960	Good protected (d = 2-5 x ObstHeight)	Rain gauge is just for estimates. Exposition may cause some problems as there maybe some additional rain drops splashed into the gauge by leaves from the bush. They had problems measuring the rain 1996 as there was too much rain.
77	RAGATI FOREST STN	9037015	295190.6305	9958256.224	2030	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	At visit observer was not around, therefore no interview could be conducted. In 1995 an elephant smashed the gauge, since then no rainfall had been measured.
79	RUMURUTI (NRM)	8936100	228763.0824	37140.25304	1820	37N	Unknown		
80	RUMURUTI MOW	8936064	227029.3994	29875.36976	1882	37N	ARC1960	Undisturbed exposition (d = 10-20 x ObstHeight)	A very reliable station!
82	SATIMA FARM	n/a	278424.8276	9983931.828	1950	37M	Unknown		
83	SEGERA PLANTATIONS	8936045	264858.5682	19719.88908	1715	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	
84	SHAMATA	9036312	224926.7752	9977502.088	2923	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	
85	SIRAJI (NRM)	n/a	310380.0244	8238.694892	2490	37N	ARC1960	Undisturbed exposition (d = 10-20 x ObstHeight)	Station is in a very bad condition... Is evaporation even measured??
86	SIRIMA (NRM)	9036355	256711.1238	9990013.563	1948	37M	Unknown		
108	SIRIMA B	n/a							
109	SIRIMA C	n/a							
110	SIRIMA D	n/a							
111	SIRIMA E	n/a							
116	SIRIMON GATE	9037155					Unknown		
87	SOLIO RANCH	n/a	263894.931	9972629.215	2007	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Gauge is in a good condition, and it is a very reliable station.

①

ID	Station	KMD-No	Longitude[83]	Latitude	Alt	UTM	Spheroid	Protection?[84]	Remarks[85]
88	SOUTH MARMANET FOREST STN	8936023	207597.701	5215.428003	2291	37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Observer was not in at time of visit.
89	SUGUROI ESTATE	9036099	237764.4334	9997181.044	2083	37M	ARC1960	Better protected (d = 5 x ObstHeight)	This is quite a reliable station, although they do no longer keep the records as they are being sent directly to KMD...
90	TELEKI (MT KENYA)	9037218	310445.7251	9981924.713	4262	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	The station is automatically operated. Every month someone from the NRM3 team has to change the chart rolls and to check the batteries
91	TELESWANI (NRM)	n/a	309631.0134	3267.168198	2724	37N	ARC1960	Undisturbed exposition (d = 10-20 x ObstHeight)	No observer was around, but rain gauge is in a very good condition. Note the extreme upwind orientation of the station.
112	TELESWANI B	n/a							
113	TELESWANI C	n/a							
114	TELESWANI D	n/a							
92	THARUA FARM	n/a	264679.8511	9987996.105	1845	37M	ARC1960	Better protected (d = 5 x ObstHeight)	
115	THE ARK GATE	n/a	260334.9398	9962562.466	2170	37M	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	This is a good station, operated by KWS. A brand new Hellmann gauge is in operation.
41	TIMAU MARANIA	8937002	328143.9456	9831.22607		37N	ARC1960	Slightly protected (d = 5-10 x ObstHeight)	Seems to be a very well maintained station.
49	TREETOPS	9036274							
93	TRENCH FARM	n/a	284461.4415	4286.223513	1903	37N	ARC1960	Better protected (d = 5 x	Station is well maintained
94	WAMBA DO	8937018	314300.1569	108886.5551	1500	37N	Unknown		

11 Station Control Table used for Gap Infilling (Decadal, 1.1.1990 – 31.12.1997)

This table was used to control the progress of eliminating gaps in the listed stations for the period from 1.1.1990 – 31.12.1997 with a 10-day interval.

Rainfall Gauge	Gaps[86]	Neighbours[87]	Performed[88]	Regression?[89]	Filled Gaps?[90]	Export to Access[91]
ARCHERS POST.xls	14	0 7	16.01.01	24.01.01	13.02.01	13.02.01
ARDENCAPLE FARM.xls	-	C			24.01.01	13.02.01
CASTLE FOREST STN.xls	251	1 3 K	16.01.01	02.02.01	13.02.01	13.02.01
CHOGORIA FOREST STN.xls	-	1 3 5 K	02.02.01		24.01.01	13.02.01
CHUKA FOREST STN.xls	360	1 3 K	02.02.01	02.02.01	13.02.01	13.02.01
COLCHECCIO.xls	-	0 4 7			24.01.01	13.02.01
EL KARAMA.xls	-				24.01.01	13.02.01
EMBORI FARM.xls	-				24.01.01	13.02.01
ENASOIT FARM.xls	-				24.01.01	13.02.01
GATHIURU FOREST STN.xls	31	2 9 B	16.01.01	24.01.01	13.02.01	13.02.01
HOMBE FOREST STN.xls	-	1 A			24.01.01	13.02.01
IRANGI FOREST STN.xls	271	1 3 K	02.02.01	02.02.01	13.02.01	13.02.01
ISIOLO DAO.xls	-	0			24.01.01	13.02.01
JACOBSON FARM.xls	-	9 D I L			24.01.01	13.02.01
JUNCTION (EWASO NAROK).xls	14	0 4	16.01.01	24.01.01	13.02.01	13.02.01
KAGURU.xls	188	5 8	16.01.01	24.01.01	13.02.01	13.02.01
KALALU (NRM).xls	-				24.01.01	13.02.01
KAMWAKI FARM.xls	-	C			24.01.01	13.02.01
KINAMBA MOW.xls	92	6 7 M	16.01.01	24.01.01	13.02.01	13.02.01
KISIMA FARM.xls	-				24.01.01	13.02.01
LAMURIA MET STN.xls	-	G	16.01.01		24.01.01	13.02.01
LARIAK FOREST STN.xls	-	6 F M	16.01.01		24.01.01	13.02.01
LOLDAIGA FARM.xls	-				24.01.01	13.02.01
LOLDOTO FARM.xls	-				24.01.01	13.02.01
LOLMARIK FARM.xls	-				24.01.01	13.02.01

[86] Amount of gaps found in the incomplete data set, retrieved from the NRM³ database
[87] Which neighbouring stations will be used for the regression? Note that the same character denote the same set of neighbours. The gaps found in the rainfall data observed by the gauge located at ARCHERS POST are completed with the aid of the neighbouring stations in area '0' (COLCHECCIO, ISIOLO DAO [reference gauge with no gap], JUNCTION (EWASO NAROK), MARALAL DC)
[88] Date at which the best fit correlation analysis was computed
[89] Date at which the regression analysis was performed
[90] Date at which the gaps were completed.
[91] Date at which the completed time series was exported as a complete table to the NRM³ database

①

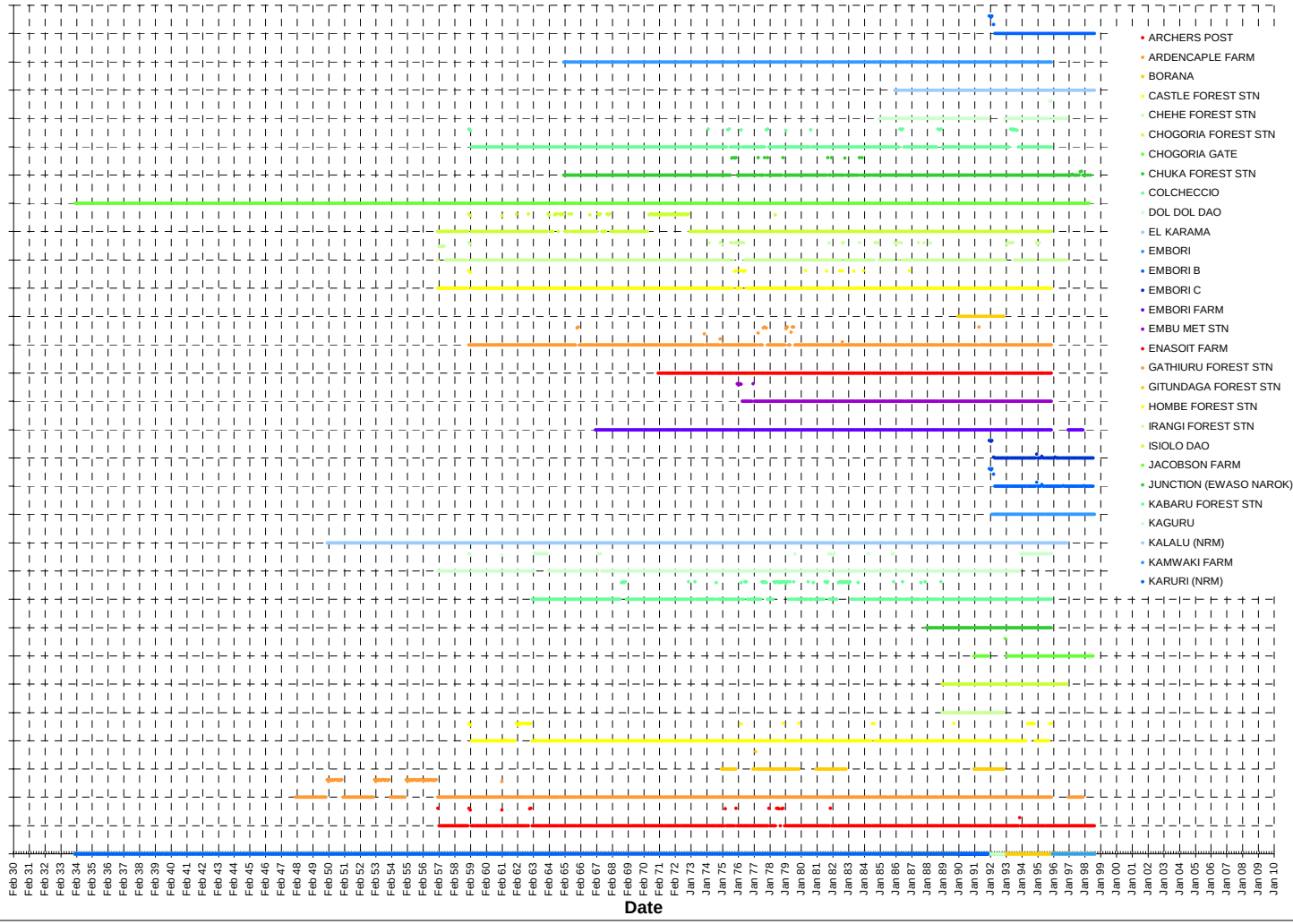
Rainfall Gauge	Gaps[86]	Neighbours[87]						Performed[88]	Regres-sion?[89]	Filled Gaps?[90]	Export to Access[91]	
LORUKU FARM.xls	-	I	L					16.01.01		24.01.01	13.02.01	
MARALAL DC.xls	427	0	7					16.01.01	24.01.01	13.02.01	13.02.01	
MARIENE CRS.xls	42	5	8					16.01.01	24.01.01	13.02.01	13.02.01	
MATANYA (NRM).xls	1	9	D					16.01.01	13.02.01	13.02.01	13.02.01	
MERU FOREST STN.xls	-	5	8							24.01.01	13.02.01	
MOGWONI RANCH.xls	-									24.01.01	13.02.01	
MPALA FARM.xls	-	4								24.01.01	13.02.01	
MUGIE RANCH	-										allready in db	this station is already in the NRM db included!
MUKOGODO (NRM).xls	-	4								24.01.01	13.02.01	
MURINGATO FOREST STN.xls	181	A						16.01.01	24.01.01	13.02.01	13.02.01	
MUTARA ADC FARM.xls	-	F								24.01.01	13.02.01	
MWEA IRRIGATION SCHEME.xls	-									24.01.01	13.02.01	
NANYUKI FOREST STN.xls	-	2								24.01.01	13.02.01	
NANYUKI KAF.xls	360	I	L					13.02.01	13.02.01	13.02.01	13.02.01	
NARO MORU FOREST STN.xls	29	2	B					16.01.01	24.01.01	13.02.01	13.02.01	
NARO MORU GATE STN.xls	-	2	B							24.01.01	13.02.01	
NARO MORU MET STN.xls	-	H								24.01.01	13.02.01	
NDARAGWA FOREST STN.xls	-	E	J							24.01.01	13.02.01	
NGENIA (NRM).xls	426	C						16.01.01	24.01.01	13.02.01	13.02.01	
NICOLSON FARM.xls	608	9	D					16.01.01	13.02.01	13.02.01	13.02.01	
NYERI MOW.xls	-	A								24.01.01	13.02.01	
OL ARABEL FOREST STN.xls	2	6	M	X				16.01.01	13.02.01	13.02.01	13.02.01	
OL BOLOSAT FOREST STN.xls	-	E	J	X						24.01.01	13.02.01	
OL DONYO FARM.xls	-	C								24.01.01	13.02.01	
OL JOGI FARM.xls	-	4								24.01.01	13.02.01	
OL JORO OROK FTC.xls	122	E	X					16.01.01	24.01.01	13.02.01	13.02.01	
OL MYSOR FARM.xls	-	4	F							24.01.01	13.02.01	
RUMURUTI MOW.xls	8	F						16.01.01	24.01.01	13.02.01	13.02.01	
SATIMA FARM.xls	360	D						-	-	-		missing years, not used for interpolation!
SEGERA PLANTATIONS.xls	-									24.01.01	13.02.01	
SHAMATA.xls	360	J						13.02.01	13.02.01	13.02.01	13.02.01	
SIRIMA (NRM).xls	2	G						13.02.01	13.02.01	13.02.01	13.02.01	
SOLIO RANCH.xls	-	A								24.01.01	13.02.01	
SUGUROI ESTATE.xls	-	G									already in db	this station is already in the NRM db included!
TELEKI (MT KENYA).xls	63	H						16.01.01	13.02.01	13.02.01	13.02.01	
THARUA FARM.xls	-	G								24.01.01	13.02.01	
TIMAU MARANIA.xls	-	8								24.01.01	13.02.01	
TRENCH FARM.xls	30	I	L					16.01.01	13.02.01	13.02.01	13.02.01	
NORTH MARMANET FORST STN.xls	730	X						04.04.01	04.04.01	04.04.01		was erroneously calculated in the first round, and is therefore recalcu-lated again.

1

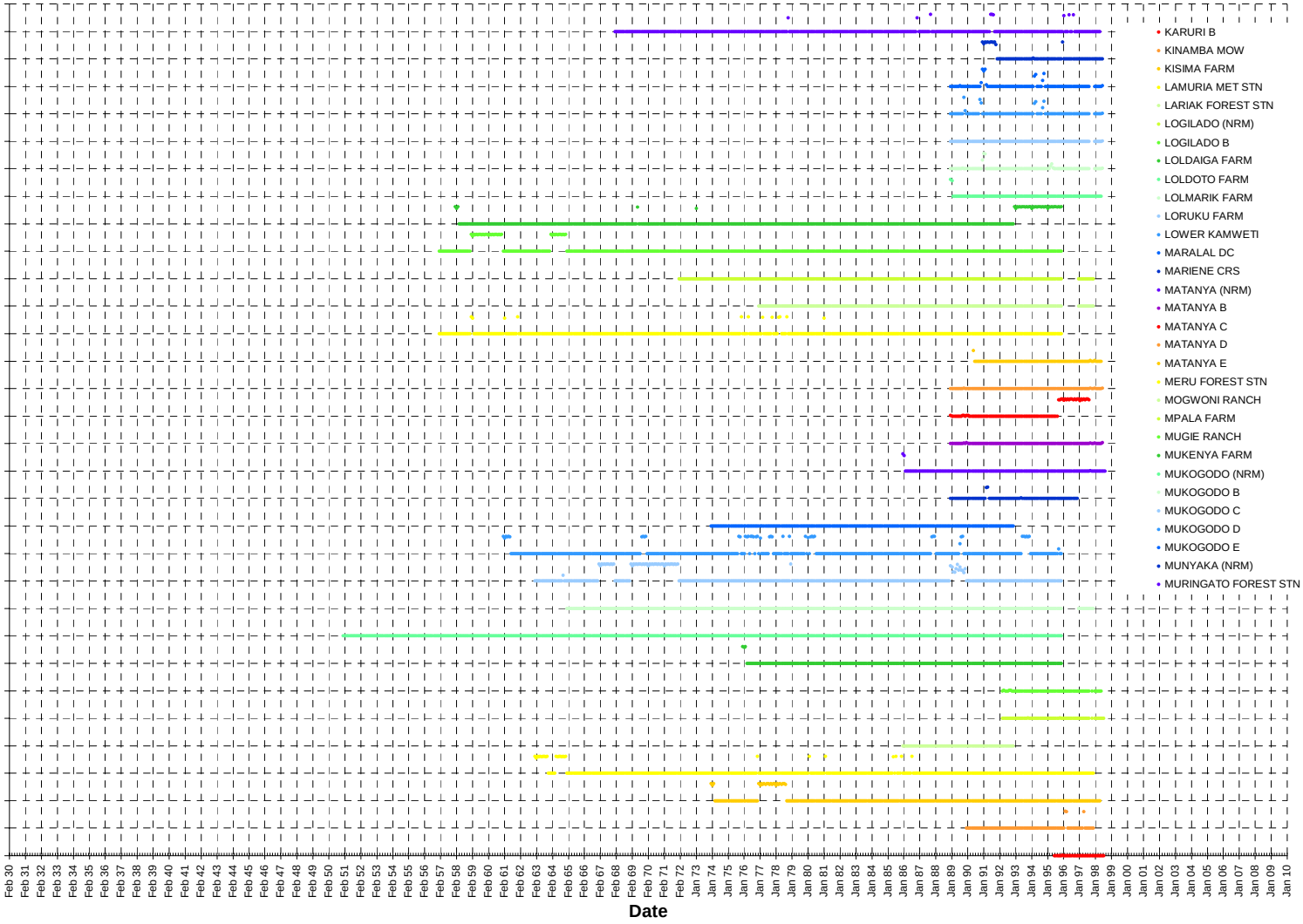
2 Appendix: B - Diagrams

12 Overview of all Available Daily Observed Rainfall Values

Please note, that the first label in the legend corresponds to the last line on the diagram.

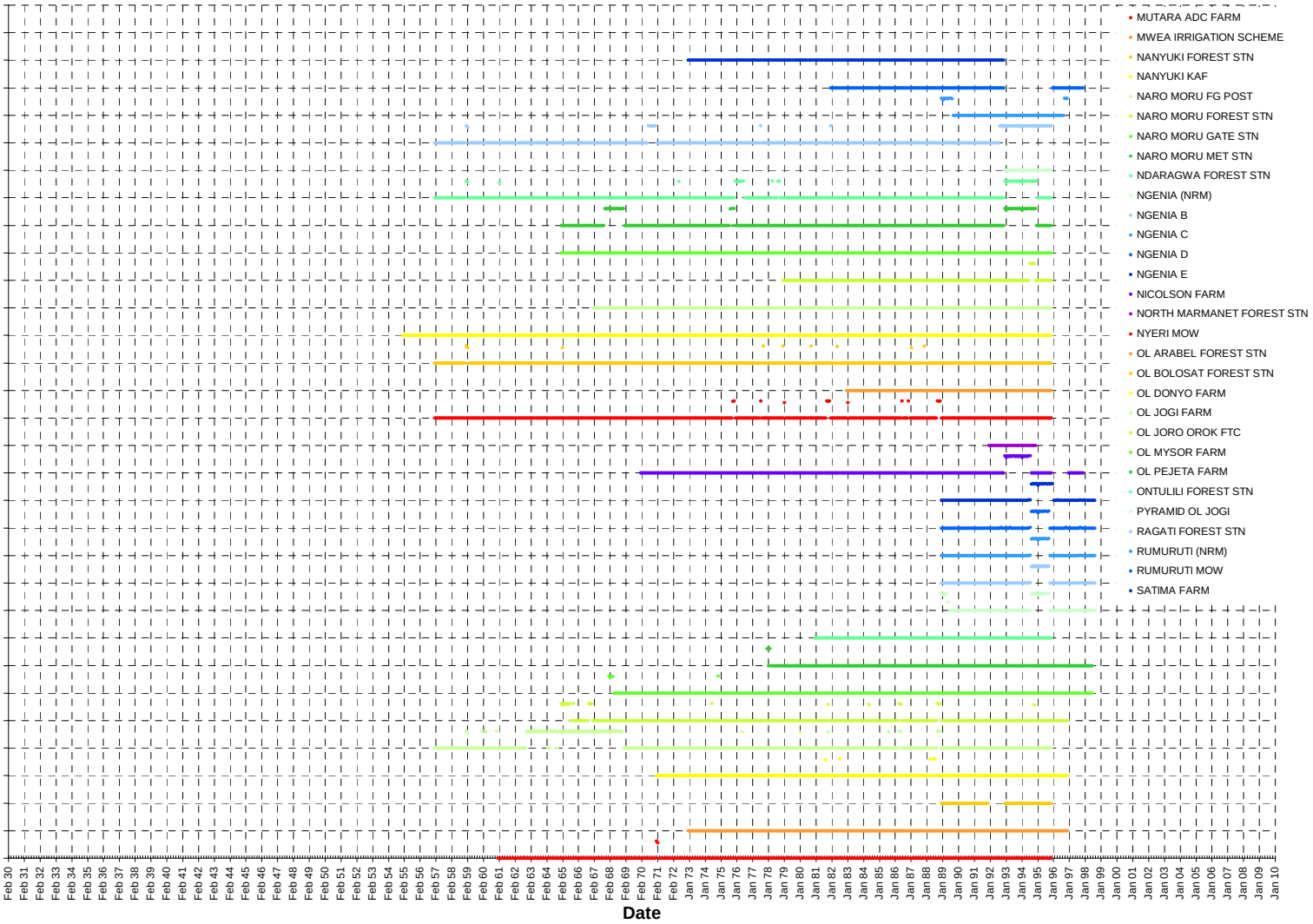


Part 2/4: Overview of all available daily observed values (a broken line indicates missing data)

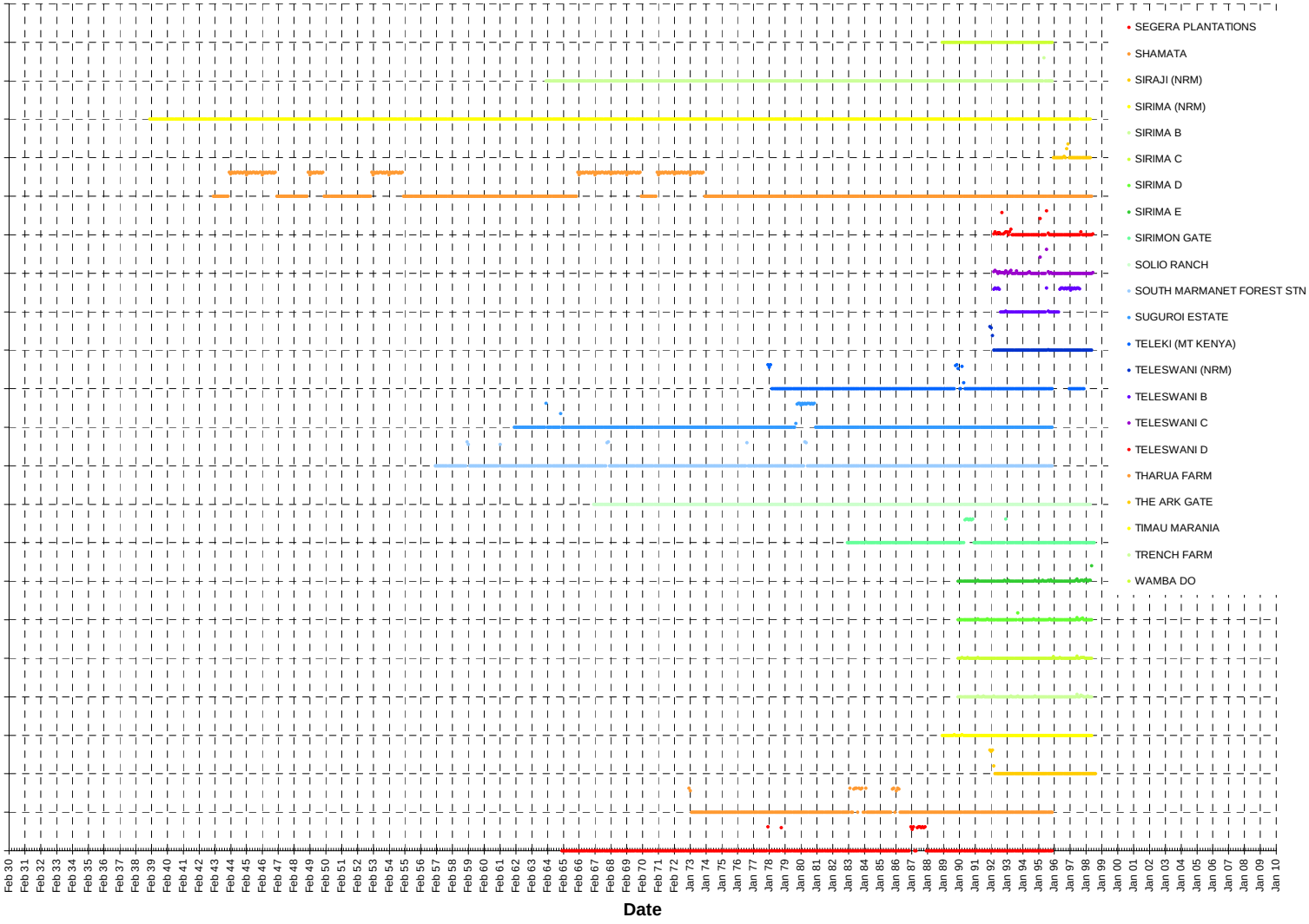


②

Part 3/4: Overview of all available daily observed values (a broken line indicates missing data)



Part 4/4: Overview of all available daily observed values (a broken line indicates missing data)



3 Appendix: C - Scripts

13 Perl Scripts

This is a collection of the scripts introduced and explained in this work. The scripts are commented, and they are not copyrighted, anybody how would like to adapt them to his or her own need shall feel free and do so. However, the author should be stated somewhere in the source code or any other reference as: Bernhard Sturm, sturm@datacomm.ch

13.1 Crossval v0.09b

```
#!usr/bin/perl -w
#
# crossval
#
# reads a one or more site definition files and cross validates different
# interpolation methods.
#
# You will have to run crossval within GRASS with: perl crossval
#
# set $path variable to the desired directory where you have stored your rainfall
# sites.
# crossval will cycle through one set of rainfall data by canceling always
# one known station. With the rest of the stations crossval computes
# the precipitation for the previously canceled gauge. This process is repeated until
# precipitation for every station in the initial set was estimated.
# The estimated and the observed values for each station are saved in a
# comparison file for later analysis.
# This file is named as:
#
# [NameOfRainfallFile]comp[InterpolationMethod][Parameter[Parameter]]_[NumberOfStationsUsed]
#
# Example:     Initial File: 6797DEC.txt Inverse Distance Weighting, 10 Nearest Stations, 40 Stations:
#                            6797DEC.compIDW10_40
#
# It is also possible to estimate a group of rainfall sets. This group has
# to be defined in a separate file which must have the follwing format:
#
# File with list of rainfall stations:
# / Set of rainfall data files for crossvalidation 30.05.00 Bernhard Sturm
# set1
# set2
```

```
# set3
#
# (note that the first line of the file must contain a '/' slash.)
#
#
#
# 30/05/2000 Bernhard Sturm sturm@datacomm.ch
#
# Version history
# Date    Version     Comments
# ----------------------------------------
# 23/05/00     0              read_stations() implemented
# 24/05/00     0.01           write_definition(), makeclean(), grass_cross_idw(npoints) implemented
# 24/05/00     0.02           first release as 'crossval'
# 25/05/00     0.03           the result is now stored in only one comparison file
# 25/05/00     0.04           list of different sites can now be processed
# 25/05/00     0.04b          crossval.1 release
# 30/05/00     0.05b          s.surf.rst implemented (not linked into program flow)
# 30/05/00     0.06b          bugfix in @comparison array re-initialisation; major bugfix
# 31/05/00     0.07b          s.surf.rst implemented (and linked to program flow)
#                             naming now: crossval (stable)
#                             crossval.x (experimental versions)
# 05/06/00     0.08b          minor bugfix (wrong parameter hand-over)
# 13/06/00     0.09b          bugfix: wrong parameter handling in system call command fixed (for s.surf.rst)


#initialise variables
#
$version="05/06/00 v0.08b Bernhard Sturm";          # Version line
$line="";                                           # reads one line out of a file
$site_file="dummy";                                 # name of the site file to be read
$count_sites=0;                                     # counts the number of sites found in the file
@sitedata="";                                       # array containing the site data
@comparison=();                                     # array containing the comparison values

$path="/home/ego/grass/Data/6797month/";            #sets the current data path

# start and show version
print "\n";
print "-- corssval -- Version: $version\n";
print "\n";

# call subs
```

```
open_site();

print "Found ".($list-1)." files in list definition file:\n";
for ($i=1; $i<$list; $i++) {
       print "$site_list[$i]\n";
}

do {

       if ($list>=1) {
              $list--;
              $site_file=$site_list[$list];        # selects the next file in the list

       }

       read_stations($site_file);
       print "processing $site_file\n";


       write_definition();

       # grass_cross_rst(dnorm,tension,smooth,npmin,dmin)
       # default values:
       # dnorm = -t
       # dmin=0.5
       # smooth=0.1
       # tension=40   (distance behaviour, high=membrane, low=stiff steel plate)
       # npmin=200

       grass_cross_rst("",40,0.1,700,0.5);  # interpolate with default values

       grass_cross_rst("-t",60,0.1,700,1);

       grass_cross_rst("",44,0.05,700,1);

       grass_cross_rst("",36,0.1,700,0.5);

       grass_cross_rst("",45,0,700,50);

       # inverse distance weighting interpolation below:

       # grass_cross_idw(5); # interpolate with 5 nearest neighbours
```

```
        # grass_cross_idw(10);# interpolate with 10 nearest neighbours

        # grass_cross_idw(12); # interpolate with 12 nearest neighbours

        # grass_cross_idw(15);# interpolate with 15 nearest neighbours

        # and delete all temporary files
        makeclean();

} while $list>1;

print "\n";
print "Finished!\n";

#**********************************************************************
#
#       grass_cross_rst(dnorm,tension,smooth,npmin,dmin)
#
#       default values:
#                       dnorm = -t
#                       dmin=0.5
#                       smooth=0.1
#                       tension=40     (distance behaviour, high=membrane, low=stiff steel plate)
#                       npmin=200
#
#                       Interpolation performed using regularised splines with
#                       tension
#                       Maximum segmentation is set to 700 in order to avoid
#                       segmentation on small data sets.
#
#       GRASS commands wll be executed in the following order:
#                       s.in.ascii     -> import sites date
#                       s.surf.rst  -> interpolates using inverse distance weighting
#                       r.what  ->     compares definition file with raster map
#                       g.remove       -> removal of site file and raster map
#
#**********************************************************************
sub grass_cross_rst {

        for ($i=0; $i<=$count_sites; $i++) {

                # first we will have to import site data
                $input=$path."inter".$i;
```

```
                $sites="inter".$i;
                $cmd="s.in.ascii input=".$input." sites=".$sites;
                system $cmd;

                # now we interpolate with regularised splines with tension
                $input="inter".$i;
                $elev="crossRST_".$i;
                $segmax=700; # avoid segmentation as there are not enough data points
                $dnorm=$_[0];
                if ($_[0] ne "-t") { # if dnorm is not -t, use default (no dnorm)
                        $dnorm="";
                }
                $tension=$_[1];
                $smooth=$_[2];
                $npmin=$_[3];
                $dmin=$_[4];

                $cmd='s.surf.rst '.$dnorm.' input='.$input.' elev='.$elev.' dmin='.$dmin.' tension='.$tension.'
smooth='.$smooth.' segmax='.$segmax.' npmin='.$npmin;
                print "interpolation with s.surf.rst, file: $site_file\n";
                print "dnorm: $dnorm\n";
                print "output elev.: $elev\n";
                print "dmin: $dmin\n";
                print "tension: $tension\n";
                print "smooth: $smooth\n";
                print "segmax: $segmax\n";
                print "npmin: $npmin\n";
                print "\n";
                print "$cmd\n";
                system ($cmd);

                # compare interpolated station with observed values at the position of the station
                $input="crossRST_".$i;
                $comp_file=" < ".$path."definition".$i;
                $t=index($site_file,".");
                if ($t == -1) {
                        $to_file =
                        $path.$site_file."compRST".$dnorm."t".$tension."s".$smooth."d".$dmin."_".($i+1);
                } else {
                        $to_file =
$path.substr($site_file,0,$t+1)."compRST".$dnorm."t".$tension."s".$smooth."d".$dmin."_".($i+1);
                }
                $cmd=("r.what input=".$input.$comp_file." > ".$to_file);   # and write it to a file
```

```
                system $cmd;

                # now remove site and raster data
                $sites="inter".$i;
                $raster="crossRST_".$i;
                $cmd="g.remove rast=".$raster." sites=".$sites;
                system ($cmd);

                # read the comparison file and store it in an array
                open (COMPARISON, $to_file) ||
                                die "I am sorry, but I couldn't read the comparison file: $!
                                        ($to_file)";
                        while ( defined ($line = <COMPARISON>)) {    # and read one line
                                chomp ($line);
                                @comparison = (@comparison, $line);
                                print "$line\n";
                        }
                close (COMPARISON) || die "Sorry, but it wasn't possible to close the
                file: $!";

                unlink ($to_file); # kill the file

        }
        # Join all comparison files to one single file
        open (COMPARISON, ">".$to_file);
        for ($i=0; $i<=$count_sites; $i++) {
                print COMPARISON "$comparison[$i]\n";

        }
        close (COMPARISON);

        # reset the comparison array
        @comparison=();       # no elements in @comparison


}



#*********************************************************************
#  grass_cross_idw($_[0] = nearest neighbours)
#
#               GRASS commands wll be executed in the following order:
```

```
#              s.in.ascii     -> import sites date
#              s.surf.idw  -> interpolates using inverse distance weighting
#              r.what         ->     compares definition file with raster map
#              g.remove              -> removal of site file and raster map
#
#*********************************************************************
sub grass_cross_idw {

       for ($i=0; $i<=$count_sites; $i++) {

              # first we will have to import site data
              $input=$path."inter".$i;
              $sites="inter".$i;
              $cmd="s.in.ascii input=".$input." sites=".$sites;
              system $cmd;

              # now we interpolate with inverse distance weighting
              $input="inter".$i;
              $output="crossIDW_".$i;
              $npoints=$_[0];
              $cmd="s.surf.idw input=".$input." output=".$output." npoints=".$npoints;
              print "interpolation with s.surf.idw, file: $site_file, nearest neighbours: $npoints\n";
              system $cmd;

              # compare interpolated station with observed values at the position of the station
              $input="crossIDW_".$i;
              $comp_file=" < ".$path."definition".$i;
              $t=index($site_file,".");
              if ($t == -1) {
                     $to_file = $path.$site_file."compIDW".$_[0]."_".($i+1);
              } else {
                     $to_file = $path.substr($site_file,0,$t+1)."compIDW".$_[0]."_".($i+1);
              }
              $cmd="r.what input=".$input.$comp_file." > ".$to_file;     # and write it to a file
              system $cmd;

              # now remove site and raster data
              $sites="inter".$i;
              $raster="crossIDW_".$i;
              $cmd="g.remove rast=".$raster." sites=".$sites;
              system $cmd;

              # read the comparison file and store it in an array
```

```
                open (COMPARISON, $to_file) ||
                                die "I am sorry, but I couldn't read the comparison file: $!
                                        ($to_file)";
                        while ( defined ($line = <COMPARISON>)) {    # and read one line
                                chomp ($line);
                                @comparison = (@comparison, $line);
                                print "$line\n";
                        }
                close (COMPARISON) || die "Sorry, but it wasn't possible to close the
                file: $!";

                unlink ($to_file); # kill the file

        }
        # Join all comparison files to one single file
        open (COMPARISON, ">".$to_file);
        for ($i=0; $i<=$count_sites; $i++) {
                print COMPARISON "$comparison[$i]\n";

        }
        close (COMPARISON);

        # reset the comparison array
        @comparison=();       # no elements in @comparison


}

#**********************************************************************
#  open_site()
#
#               Asks the user for a site file (or a list of site files)
#               that shall be cross validated
#               If there is a list of files to be processed the sub
#               evaluates $list = 1
#
#**********************************************************************

sub open_site {

        #initialise variables
        $list=0;
```

```
        # first enter the file name

        while (! -e $path.$site_file) {
                print "Enter a site file name:\n";
                $site_file=<STDIN>;
                chomp $site_file;

                # check if file exists
                if (!-e $path.$site_file) {
                        print "OOPS... The file $site_file does not exist, try again.\n";
                }
                # check if file contains a list
                open (INFILE, $path.$site_file);
                $line = <INFILE>;
                chomp ($line);
                if (index($line,"/")==-1) { #this is not a list of rainfall site files...
                        $list=0;
                } else {
                        $list=1;
                        $site_list = $site_file;
                        open (LIST, $path.$site_list);
                                while ( defined ($line = <LIST>)) {  # and read one line
                                        chomp ($line);
                                        @site_list = (@site_list, $line);
                                        $list++;       # increment
                                }
                        close (LIST);
                        $list--;
                }

                close (INFILE);
        }

}


#*************************************************************************
#  read_stations()
#
#              reads an ASCII file containing rainfall values and
#              stores each gauge in an array
#
#*************************************************************************
```

```
sub read_stations {

        # reset values
        @sitedata=();
        $count_sites=0;

        # now we have to open the site file
        open (SITEDATA, "$path$site_file") ||
                                    die "I am sorry, but this doesn't seem to be a
                                    valid file: $! ($path$site_file)";
                while ( defined ($line = <SITEDATA>)) {      # and read one line
                        chomp ($line);

                        @sitedata = (@sitedata, $line);
                        $count_sites++;       # increment
                }

        close (SITEDATA) || die "Sorry, but it wasn't possible to close the file: $!";

        # finished
        print "Found $count_sites rain gauges at file $site_file.\n";
        $count_sites--; # decrement -1

}

#*********************************************************************
#  write_definition()
#
#       writes the definition and the files needed for the
#       interpolation.
#       A definition file contains only the location of a single
#       rainfall station, and will be queried from GRASS with
#       r.what
#       The interpolation file will contain n-1 rainfall gauge
#       information, and will be imported as a site file into
#       GRASS with the s.in.ascii command
#
#*********************************************************************
sub write_definition {

        for ($def_count=0; $def_count<=$count_sites; $def_count++) {
                open(DEFINITION, ">".$path."definition".$def_count) ||
```

```
                                                                    die "Sorry, I wasn't able to create
$path.definition.$def_count\n";
                open(INTER, ">".$path."inter".$def_count) ||
                                                                    die "Sorry, I wasn't able to create
$path.inter.$def_count\n";

                print STDOUT "Writting definition and site data for station: $def_count\n";

                for ($i=0; $i<=$count_sites; $i++) {
                        if ($def_count != $i) {       # not a definition station
                                print INTER "$sitedata[$i]\n";

                        } else { # then this must be site that is going to be a definition site
                                print DEFINITION "$sitedata[$i]\n";

                        }

                }
        close(DEFINITION);
        close(INTER);
        }
}

#*********************************************************************
#  makeclean()
#
#       This will clean (delete) all files which were used during
#      interpolation process.
#
#*********************************************************************
sub makeclean {

        print "cleaning up...\n";

        # clean definition and interpolation file list
        for ($i=0; $i<=$count_sites; $i++) {
                unlink ($path."definition".$i);      # definition file is gone...
                unlink ($path."inter".$i);                  # interpolation file is gone...

        }

}
```

13.2 Rsurfrst v0.0b

```
#!usr/bin/perl -w
#
# rsurfrst
#
# reads a one or more site definition files and
# interpolates it with the regularised splines with tension interpolatation
# model, using parameters
#
#
# You will have to run rsurfrst within GRASS with: perl rsurfrst
#
# set $path variable to the desired directory where you have stored your
# rainfall sites.
#
#
#
#
# 25/07/2000 Bernhard Sturm sturm@datacomm.ch
#
# Version history
# Date    Version     Comments
# ----------------------------------------
# 25/07/00     0                    Written the whole thing ;-)



#initialise variables
#
$version="25/07/00 v0.0b Bernhard Sturm";          # Version line
$line="";                                  # reads one line out of a file
$site_file="dummy";          # name of the site file to be read
$coldef_file="dummy"; # dummy name for the color definition file


$path="/home/ego/grass/Data/9097decade/";   #sets the current data path

# start and show version
print "\n";
print "-- rsurfrst -- Version: $version\n";
print "\n";

# call subs
```

```
read_input();         # here we will open the file or a list of files, the interpolation parameters, and the color
definition files

print "There were ".($list-1)." periods defined in the entered list of files:\n";
for ($i=1; $i<$list; $i++) {
       print "$site_list[$i]\n";
}

do {

       if ($list>=1) {
              $list--;
              $site_file=$site_list[$list];        # selects the next file in the list

       }

       print "processing $site_file\n";

       surf_rst();           # interpolate it!


} while $list>1;

print "\n";
print "Finished!\n";

#**********************************************************************
#
#       surf_rst();
#
#
#              Interpolation performed using regularised splines with
#              tension
#              Maximum segmentation is set to 700 in order to avoid
#              segmentation on small data sets.
#
#              GRASS commands wll be executed in the following order:
#              s.in.ascii     -> import sites date
#              s.surf.rst  -> interpolates using spline with tension
#
#**********************************************************************
sub surf_rst {
```

```
# first we will have to import site data
$input=$path.$site_file;
# extract an eventually occurence of an extension
$i = index($site_file, ".");
if ($i!=-1) {
       $sites=substr($site_file, 0, $i);
} else {
       $sites=$site_file;
}

$cmd="s.in.ascii input=".$input." sites=".$sites;
system $cmd;

# now we interpolate with regularised splines with tension
$input=$sites;
$elev="9097".$sites."_RST";

$cmd='s.surf.rst '.$dnorm.' input='.$input.' elev='.$elev.' dmin='.$dmin.' tension='.$tension.'
smooth='.$smooth.' segmax='.$segmax.' npmin='.$npmin;
print "interpolation with s.surf.rst, file: $site_file\n";
print "dnorm: $dnorm\n";
print "output elev.: $elev\n";
print "dmin: $dmin\n";
print "tension: $tension\n";
print "smooth: $smooth\n";
print "segmax: $segmax\n";
print "npmin: $npmin\n";
print "\n";
print "$cmd\n";
system ($cmd);

# remove sites
$cmd='g.remove sites='.$sites;
system ($cmd);

# apply colormap
$cmd='cat '.$path.$coldef_file.' | r.colors map='.$elev.' color=rules';
system ($cmd);

# show something on monitor x0
$cmd='d.rast map='.$elev;
system ($cmd);
```

```
}



#**************************************************************************
#  read_input()
#
#       Asks the user for a site file (or a list of site files)
#      that shall be interpolated
#       If there is a list of files to be processed the sub
#       evaluates $list = 1
#       Any existing parameter file or color definition
#       file is read, too
#
#**************************************************************************

sub read_input {

        #initialise variables
        $list=0;

        # first enter the file name

        READ: while (! -e $path.$site_file) {
                # reads the pre defined sites
                print "Enter a site file name:\n";
                $site_file=<STDIN>;
                chomp $site_file;

                # check if file exists
                if (!-e $path.$site_file || $site_file eq "") {
                        print "OOPS... The file $site_file does not exist, try again.\n";
                        redo READ;
                } else {
                        # check if file contains a list
                        open (INFILE, $path.$site_file);
                        $line = <INFILE>;
                        chomp ($line);
                        if (index($line,"/")==-1) { #this is not a list of rainfall site files...
                                $list=0;
                        } else {
                                $list=1;
```

```
                        $site_list = $site_file;
                        open (LIST, $path.$site_list);
                                while ( defined ($line = <LIST>)) {  # and read one line
                                        chomp ($line);
                                        @site_list = (@site_list, $line);
                                        $list++;       # increment
                                }
                        close (LIST);
                        $list--;
                }

                close (INFILE);
        }
}

# enter the color definition file
READCOL: while (! -e $path.$coldef_file) {
        # reads the color definition file for the interpolation map
        print "Enter a color definition file or use the default colors [ENTER]\n";
        $coldef_file=<STDIN>;
        chomp $coldef_file;

        # check if file exists
        if (! -e $path.$coldef_file && $coldef_file ne "") {
                print "Sorry, but there is no such color definition file.\n Try again!\n";
                redo READCOL;

        } elsif ($coldef_file eq "") {
                # we will therefore use the standard color table
                # which has to be written in case it doesn't exist yet ;-)
                if (! -e $path."standard.col") {
                        # doesn't exist
                        open (OUT, ">".$path."standard.col") || die "OOPS! Cannot create standard.col!\n";
                                print OUT "0 241 125 68\n";
                                print OUT "15 236 169 86\n";
                                print OUT "30 232 203 101\n";
                                print OUT "45 229 232 113\n";
                                print OUT "60 207 223 135\n";
                                print OUT "75 162 200 174\n";
                                print OUT "90 115 175 215\n";
                                print OUT "105 64 132 235\n";
                                print OUT "120 44 100 220\n";
                                print OUT "135 25 80 193\n";
```

```
                                print OUT "150 20 93 161\n";
                                print OUT "165 14 110 118\n";
                                print OUT "180 6 131 64\n";
                                print OUT "end\n";

                        close (OUT) || die "There was a problem creating the standard color table...!\n";
                        }
                        $coldef_file="standard.col";
                }

        }

        # enter the parameters for the regularised splines with tension model
                print "Enter the paramteres for the regularised with splines model:\n";
                print "Tension: [44]\n";
                $tension=<STDIN>;
                chomp $tension;
                if ($tension eq "") {$tension=44}
                print "Smooth: [0.05]\n";
                $smooth=<STDIN>;
                chomp $smooth;
                if ($smooth eq "") {$smooth=0.05}
                print "Dnorm: [disabled]\n";
                $dnorm=<STDIN>;
                chomp $dnorm;
                if ($dnorm eq "") {$dnorm=""}
                print "Dmin: [1]\n";
                $dmin=<STDIN>;
                chomp $dmin;
                if ($dmin eq "") {$dmin="1"}
                print "Segmax: [700]\n";
                $segmax=<STDIN>;
                chomp $segmax;
                if ($segmax eq "") {$segmax="700"}
                print "Npmin: [700]\n";
                $npmin=<STDIN>;
                chomp $npmin;
                if ($npmin eq "") {$npmin="700"}


}
```

4 Appendix: D - Results

14 Long-Term Interpolation

The results for the long-term spatial interpolation are presented as decadal and mean annual maps.

14.1 Decadal Analysis

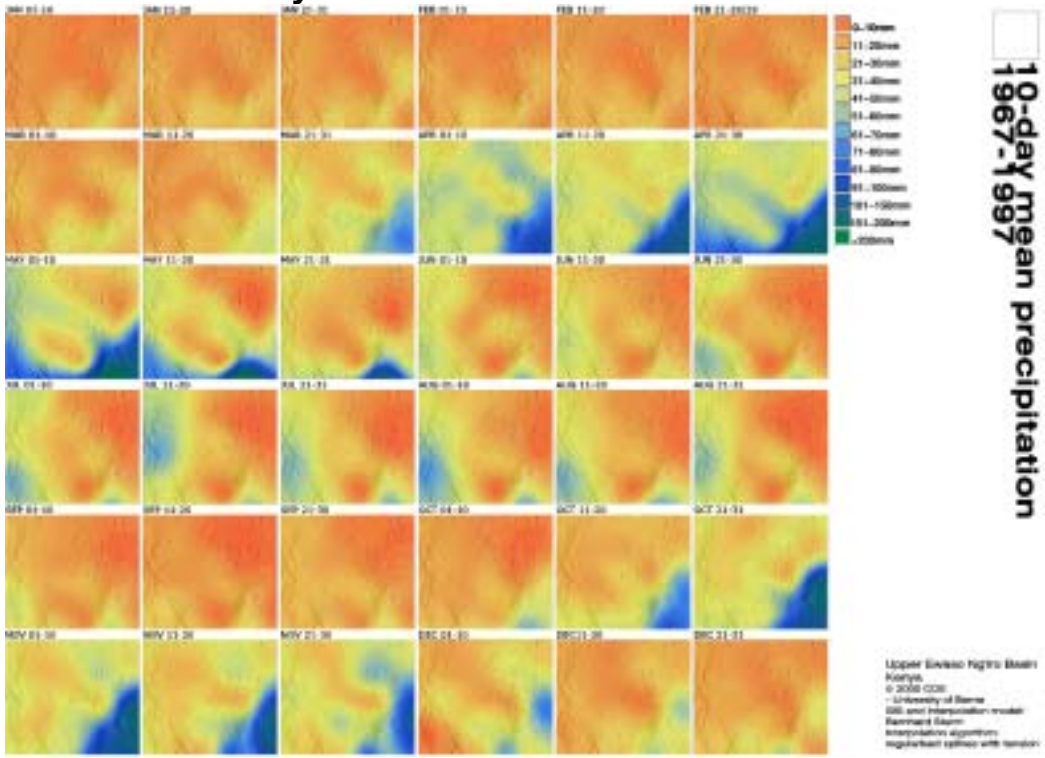


Figure 14.1: Long-term mean decadal analysis (1.1.1967-31.12.1997), 40 gauges used for the interpolation with regularised splines with tension (map by B. Sturm)

14.2 Mean Annual Analysis

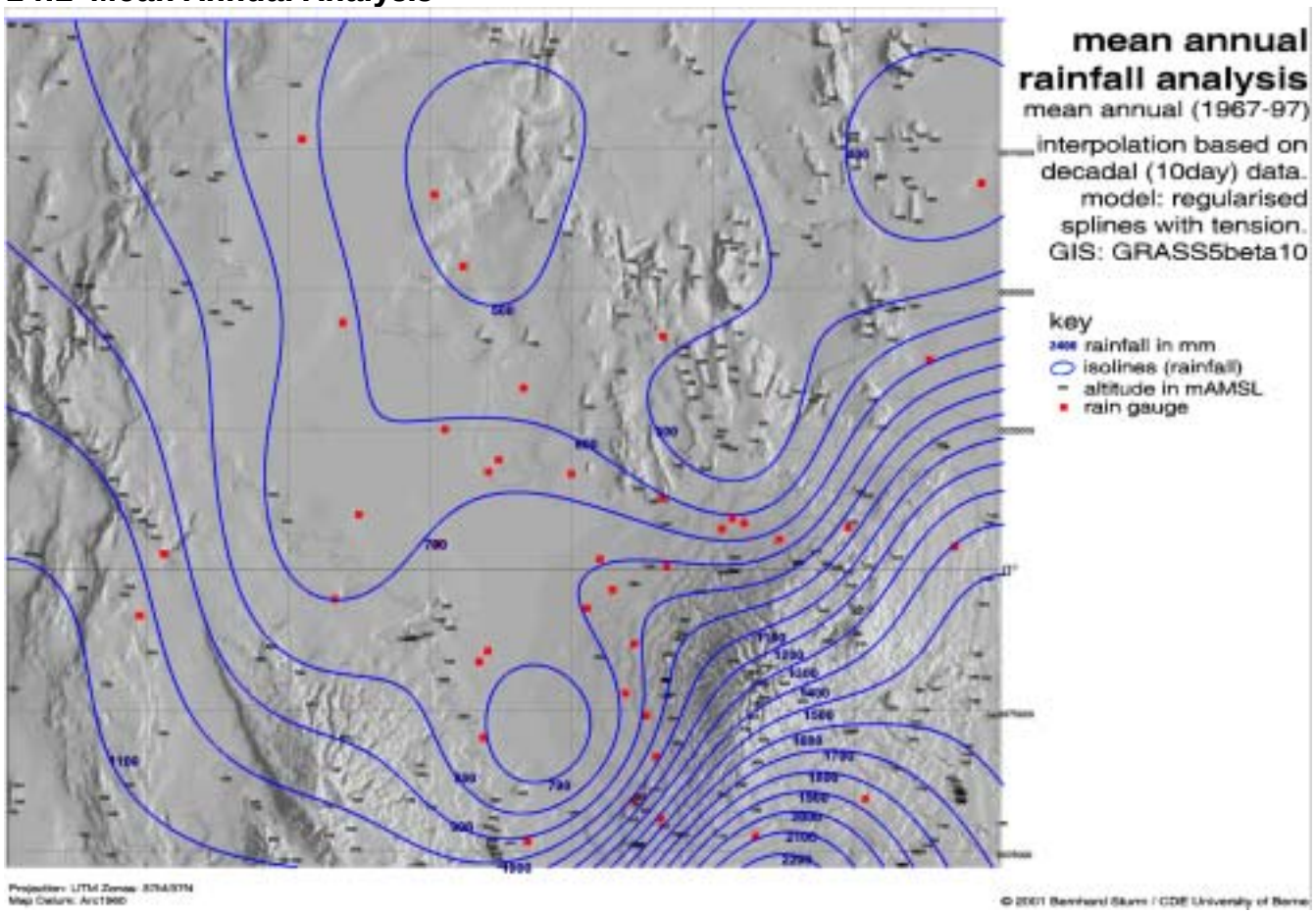


Figure 14.2: Mean annual analysis from 1.1.1967-31.12.1997. (Map by B. Sturm)

15 Short-Term Interpolation

The results for the short-term spatial interpolation are presented as decadal and mean annual maps.

15.1 Decadal Analysis

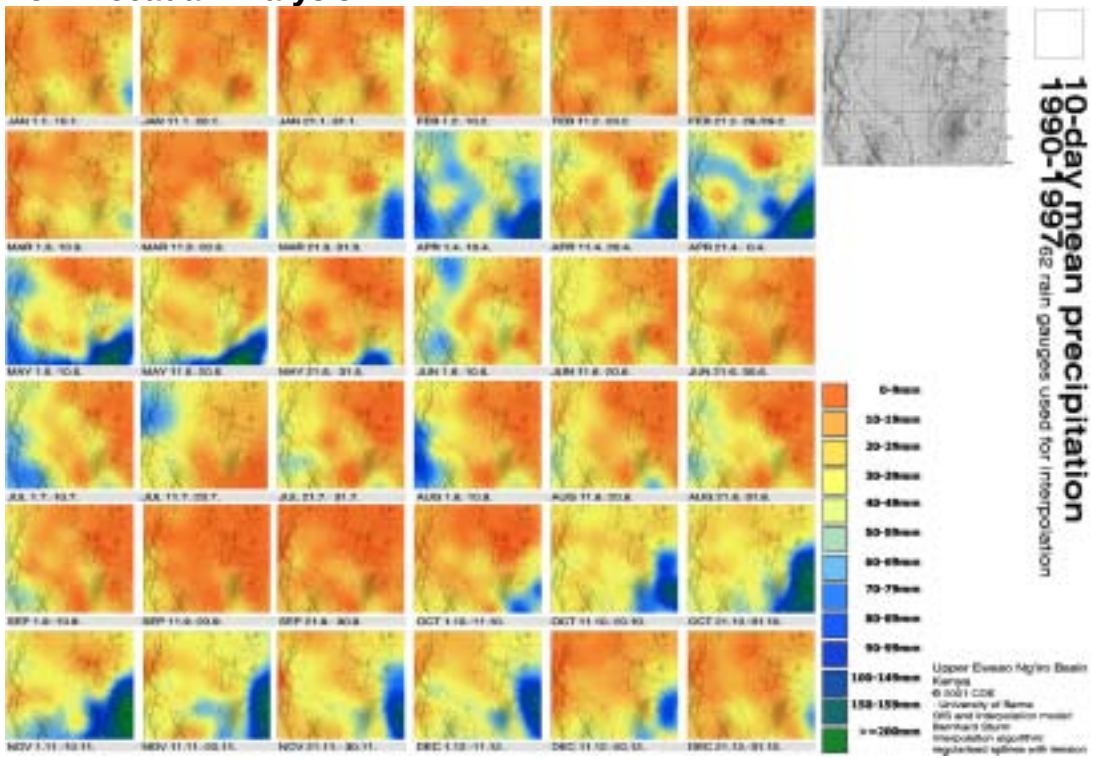


Figure 15.1: short-term mean decadal analysis (1.1.1990-31.12.1997), 62 gauges used for the interpolation with regularised splines with tension (map by B. Sturm)

15.2 Mean Annual Analysis

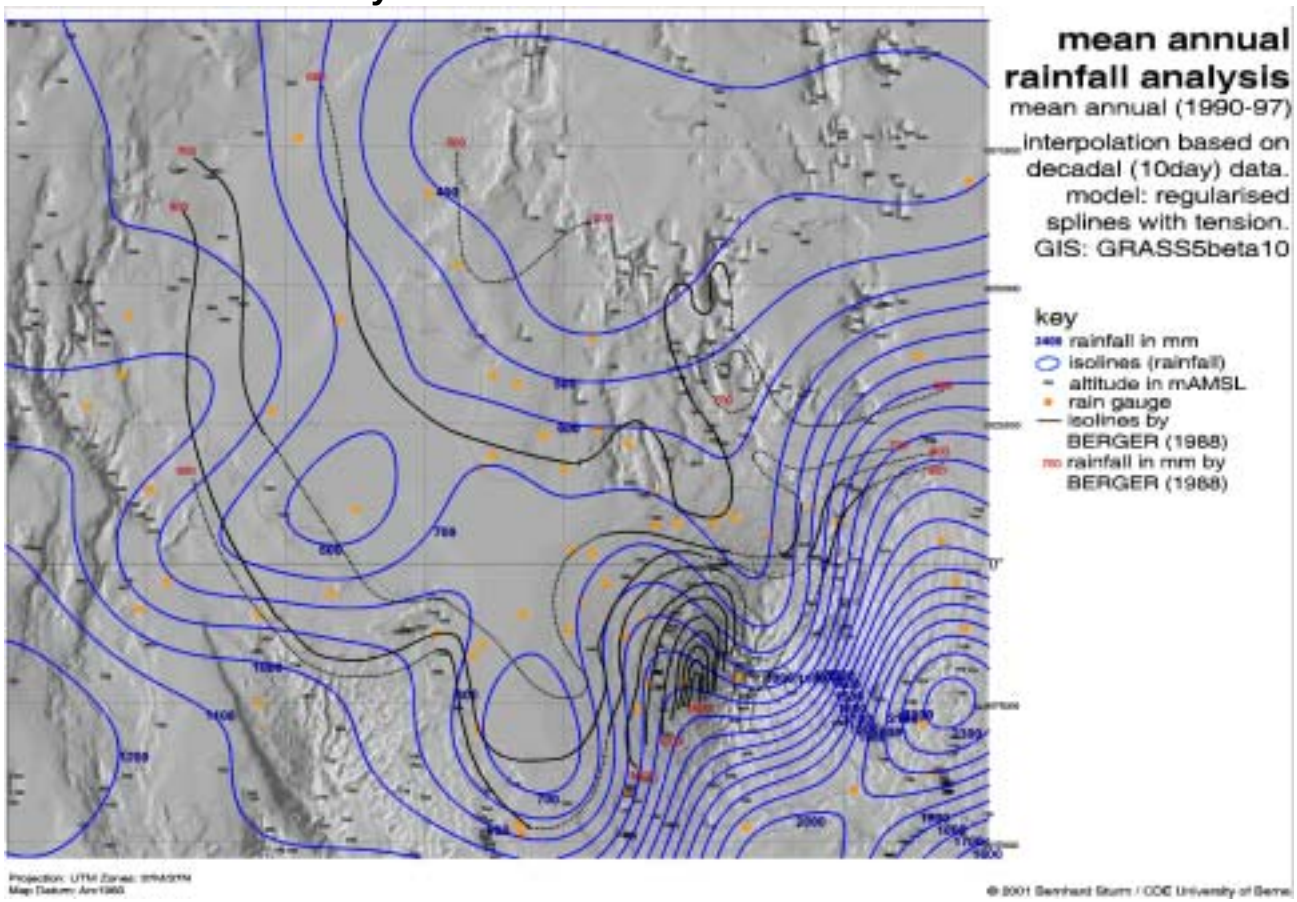


Figure 15.2: Mean annual analysis from 1.1.1990-31.12.1997. Black lines are the estimated annual rainfall isolines by Peter Berger (BERGER: 1988). (Map by B. Sturm and P. Berger)

④

16 Questionnaire

The result of the interviews with the farmers will be presented in this section.

16.1 Group: Perception of Rainfall in General

Question 3.2: *Would you assess this year as a good or a bad year? (in terms of rainfall and harvesting)*

	Answers
A good year:	11
A normal year:	2
A bad year:	3

Question 3.3: *Is a bad year characterised by the total amount of less rain per year or by un-even distributed rainfalls?*

	Answers
Less rain per year	11
Uneven distribution	9

Note: it was possible to tick both possible answers

Question 3.4: *In terms of rainfall, what do you expect for next year (1999)?*

	Answers
It will be a better year:	1
It will be an average year:	5
It will be a worse year:	4
I am not sure about it:	3

Why?

- Too many good years in the past
- There was no drought yet
- Experience

Question 3.5: *In general, how do you think has the weather changed?*

	Answers	**Remarks**
Not:	1	
Yes:	11	Its unpredictable (5)
Now:	4	
For this time period :	Changed since the last few years (6 years)	

Question 3.6: *Do you remember years of even high rainfalls such as the El Niño year in 1997/98?*

	Answers
El Niño years:	1961 (7); 1977/78 (3); 1988 (2)
Drought years:	1964/65 (1); 1980 (2); 1984 (11); 1991/92

Note: Numbers in brackets represents the number of namings.

Question 3.8: *What rules do you have to predict the onset and end of a dry or wet period?*

	Answers
The bloom of the Thorntree indicates the onset of a wet period. (2-3 weeks before the start of the short rains)	3
Old and wise people (an old Samburu man) could always tell the onset of a wet period.	2
When the tortoise show up the rains will set in.	1

Question 3.10: *Do you believe the long-term weather forecasts that predict the La Niña phenomenon?*

	Answers
Yes	4
No	4
I never heard about it	4

④

Question 3.11: *What measures do you take to be prepared for a bad year?*

	Answers	Remarks
None:	1	
Yes:	14	Sell cattle, store food

Question 3.12: *In terms of rainfall, do you think there is an ongoing change within the rainy season?*

	Answers
No	5
Yes	8
Maybe	3

Question 3.13: *Has anything changed concerning the beginning and length of the rain season?*

	Answers	Remarks
No	1	
I don't know	1	
Yes	9	The seasons have shifted more towards the end of the year.

Question 3.14: *Do you think there is a trend affecting the rainfall?*

	Answers
No	4
Yes	4

Question 3.15: *Which are good years, which are bad? List them:*

	Answers
Good years:	1994, 1982, 1984
Bad years:	1993

Question 3.16: *Do you believe that other climatic parameters than rain may undergo a change? (Wind, Radiation, Humidity):*

	Answers	Remarks
No	1	
Yes	5	Wind has changed (1); Temperature has changed (3)

16.2 Group: Questions about land use and harvesting

Question 4.1: *Which rain season is best for harvesting?*

The answers to this question were always related to the farm or ranch where the interview was held, and it is therefore not possible to summarise these answers.

Question 4.2: *Due to the climate change, did you change the way you use your land? In what way?*

	Answers	Remarks
No	8	
Yes, why?	3	Greater consciousness about soil degradation (erosion); Grass will be cut in order to protect it from dryness; Land use has intensified

Question 4.3: *Due to the climate change, did you change the way you use your land? In what way?*

Answers
2 Answers: It's too early to react. We don'thave enough information to react.

④

Question 4.4: *How do you use the observations on the climate, at what amount of rainfall do you start planting?*

Answers
5 Answers: Don't rely on the records.
1 Answer: Is using the data for a PhD study (OL DONYO FARM)

Question 4.5: *If you compare your area with other areas, can you see any difference for the last years?*

	Answers
No	7
This area is better	3
This area is worse	4
This area used to be better	1
This area used to be worse	1