
1 Content

| | | |
|----------|--|-----------|
| 1 | Content | 1 |
| 2 | Introduction | 2 |
| 2.1 | Where to start? | 2 |
| 2.2 | What is needed?..... | 2 |
| 2.2.1 | Software Setup..... | 3 |
| 2.3 | What are the steps?..... | 4 |
| 3 | Interpolate Missing Rainfall Data Values | 5 |
| 3.1 | From Raw data to Gap-Free Data..... | 5 |
| 3.1.1 | Step 1: Identify Suitable Rain Gauges..... | 6 |
| 3.1.2 | Step 2: Collecting Precipitation for a Desired Time Resolution..... | 6 |
| 3.1.3 | Step 3: Identify Neighbouring Stations and Group Them | 8 |
| 3.1.4 | Step 4: The Regression Analysis | 11 |
| 3.1.5 | Step 5: Export to ACCESS and GRASS | 15 |
| 4 | Spatial Interpolation | 17 |
| 4.1 | Importing Site Data in GRASS..... | 18 |
| 4.2 | Surface Interpolation of Site Data | 18 |
| 4.2.1 | Prepare and Run the Interpolation Process | 20 |
| 4.2.2 | Export and Contour Vector Extraction..... | 22 |
| 5 | Literature | 25 |

2 Introduction

The main goal of this paper is to form a short introduction into the production of rainfall interpolation maps. There are many cases and situations which make it desirable to have rainfall maps in order to understand the spatial rainfall distribution pattern.

This manual shall give some assistance for the production of such maps. The herewith-presented method is based on the "Regular Spline with Tension (RST)" interpolation method, and was taken from the master thesis of Bernhard Sturm (STURM, 2001).

It shall be noted, that there are many other interpolation methods, which application may lead to different results. However, a comparison between the most common interpolation models showed that the RST method proved to provide most reliable results in a reasonable amount of time.

2.1 Where to start?

At the beginning of any sort of spatial interpolation task there are measured point data. In most cases these data will originate from a rain-gauge network covering a certain period of time. In order to get reliable results from such observed point data one has to ensure data quality and minimize systematic and random errors of the entire observation chain. This is a most crucial point, because erroneously measured data will also lead to non-reliable rainfall maps¹.

This paper assumes that the measured data is of acceptable quality and the gauge density of the network can be assessed as dense enough to represent rainfall according to the definitions of the World Meteorological Organisation (ARKIN, 1985: 16-17).

Furthermore this manual assumes that the observed data is electronically stored in an MS Access Database. This is not an essential need as the spatial interpolation method is based on ASCII text files, hence independent from any sort of used data management system.

2.2 What is needed?

The manual will give step-by-step instructions based on the following technical requirements:

- Microsoft Windows Operating System (Windows9x, NT, 2000, XP)
- Microsoft Access97 (Works also with Access2000)
- Microsoft Excel97 (Works also with Excel2000)
- Linux Operating System (Any current flavour of Linux will do)
- GRASS (Geographical Resource Analysis Support System)

The above list reveals that the interpolation will take place under a Linux environment using the Open Source GIS system GRASS. This makes it possible to run a professional GIS without additional costs (as Linux and GRASS are free of cost) under almost any available hardware.

Linux and Windows can be run on the same PC, but a good knowledge of the fundamentals of Linux is needed in order to successfully conduct the interpolation tasks.

This manual will not provide the basics of the Linux and GRASS system, but it is most convenient to read one of the many Linux books and to study the GRASS user manual² before starting the interpolation model.

¹ For more information about error theory, and data quality refer to HERWEG and OSTROWSKI, 1997.

² The GRASS user manual is available at: www.geog.uni-hannover.de or at <http://www.geog.uni-hannover.de/grass/>.

Please note that this manual does not give a 'point-and-click' introduction into rainfall interpolation, but an introduction in how rainfall can be interpolated by using the above listed software.

2.2.1 Software Setup

In order to complete all the necessary steps described in this manual you will need to previously install and setup the software and tools described in this manual. All non-commercial and open source tools and software can be found on the CD-ROM available for this manual. For the installation of Linux and GRASS, please refer to the respective manuals of the particular software package.

The VBA macros needed to complete the interpolation are also available on the CD-ROM, and can be found under the following paths:

| Excel Tools | Description | CD-ROM path |
|---------------------------------|--|---|
| CrossValAnalysisMacro .XLS | EXCEL macro to cross validate different interpolation methods (not described in this manual) | /tools/macro_office97/CrossValAnalysisMacro.xls |
| GapCollectMacro.XLS | EXCEL macro to fill-in missing data and complete the datasets for spatial interpolation with GRASS (described in this manual) | /tools/macro_office97/GapCollectMacro.xls |
| | In order to run the above macros you need to install the following files in the same directory as the macro is placed ³ : COMCTL32.OCX COMDLG32.OCX MSCOMCT2.OCX MSComDlg.TWD | /tools/macro_office97/ |
| Access Tools⁴ | Description | CD-ROM path |
| FrmSNHT | Assesses different rainfall stations concerning quality according to the Standardized Normal Homogeneity Test (SNHT). For more information refer to the snht-help.htm file on the CD-ROM. | /database/EvapRain.mdb |
| FrmIntervalQuery | Collects and groups daily rainfall data into specific intervals (days). | /database/EvapRain.mdb |
| Perl Scripts⁵ | Description | CD-ROM path |
| Crossval | Automated cross validation for different interpolation methods and different interpolation parameters. Output is a comparison file with estimated and observed values. | /tools/perl_grass/crossval |
| Process | Can be used to process any sort of GRASS command for a list of files | /tools/perl_grass/process |
| Rsurfst | Automatic spatial interpolation of a specific list of site data with the | /tools/perl_grass/rsurfst |

³ If you should experience troubles in running the macro, you will have to set the correct library references in the macro by pressing [ALT]+[F11] to enter the macro editor. Under [Tools]> [References...] you have to browse for the missing libraries. (They should be marked as MISSING... by Excel).

⁴ These tools are integrated in the rainfall and evaporation database (EvapoRain.mdb). This is an Access97 database, which can easily be converted into an Access2000 database. If you should have troubles running the macros, you might have to open the macro editor [ALT]+[F11] and set the references for the used libraries under [Tools]>[References].

⁵ The perl scripts are needed for the automated spatial interpolation under the GRASS GIS system. You can copy those files to any location on your Linux machine, but when starting the scripts you need to start them from their respective location e.g.: perl \home\user\grass\scripts\crossval

regularized with splines model.
Output is an interpolated raster file.

2.3 What are the steps?

Basically this manual covers two steps:

1. Interpolate missing rainfall data values
2. Spatially interpolate point rainfall data

Step 1 is a very common problem when dealing with rainfall data. Very rarely rainfall measurement series are complete and hence gap-free. During Step 1 we will have to fill empty gaps in the rainfall measurement series. Step 1 can be performed under the Windows environment and is completely independent from Step 2.

In Step 2 we will spatially interpolate the completed rainfall series from Step 1 with the application of the RST interpolation module integrated in GRASS.

3 Interpolate Missing Rainfall Data Values

3.1 From Raw data to Gap-Free Data

Starting point are the stored precipitation data in the NRM-database. We assume these data as being quality controlled. Further we will assume that 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 1: 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.

- 1 Identify suitable rain gauges, and a time-period from the NRM database
- 2 Combine the observed values in a desired time resolution (for instance decades (10 day interval))
- 3 Extract series with gap free rainfall (they will be used as reference stations for the gap infilling)
- 4 Identify for each station where gaps were found neighbouring stations (at least one of the neighbours need to be gap free)
- 5 Compute the linear regression for each station, and determine the best-fit for each time period
- 6 Use the best-fit to linearly interpolate the missing values
- 7 Re-group the gap free stations, save them for later reference in the NRM database, and export them to GRASS

Figure 0: Workflow for the interpolation of gaps

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 was used.

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 Figure 1.

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

3.1.1 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 quality⁶, 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 paper two different sets of analysis were performed:

- A long-term analysis (30 years) with decade data resolution (10 day periods)
- A short-term analysis (8 years) with decade data resolution (10 day periods)

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

Table 2: Summary of the two used data set in this paper.

Table 2 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 proved to be robust enough to interpolate even complete missing years. 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.

3.1.2 Step 2: Collecting Precipitation for a Desired Time Resoultion

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 (ARKIN, 1985: 16-17). For the Ewaso Ng'iro Basin a station density of 656.25km² and 423.39km² (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 ob-served rainfall values:

⁶ The quality of a station is also stored in the NRM database, and was determined during a 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.

Group the Precipitation in an Interval

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

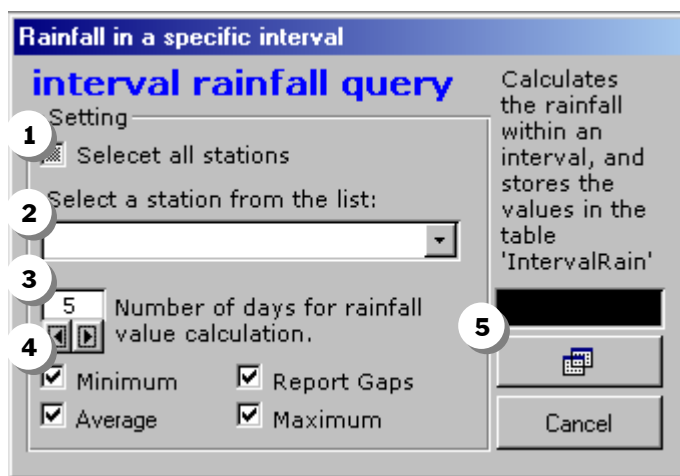


Figure 0: Interface of the Interval Rainfall Query dialog in the NRM access database.

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 tblIntevalRain 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 3: 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:

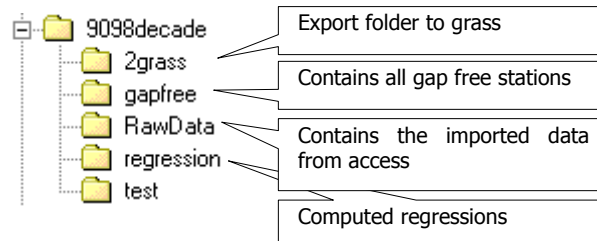


Figure 1: Suggested directory structure

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

3.1.3 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 'gap free'. These stations represent reference gauges; we will need them to compute the missing values for the other gauges.

Identification of neighbouring stations

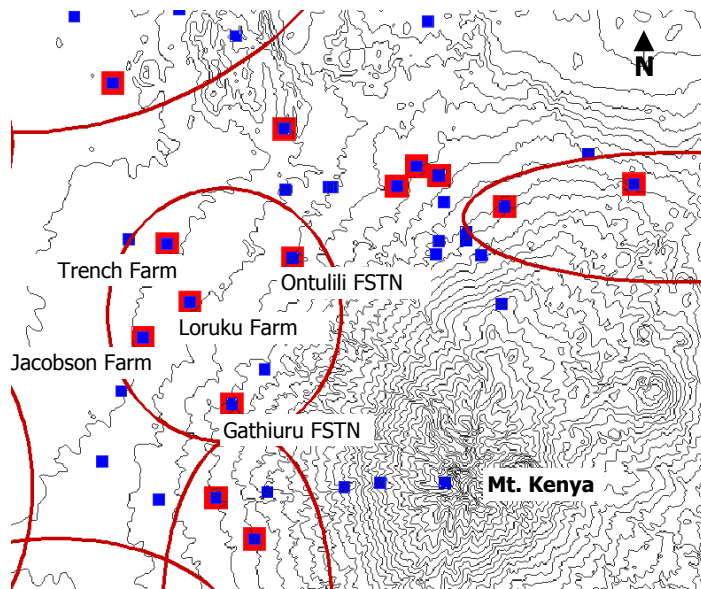
This is a crucial part of the interpolation process, as the applied method will use the values of surrounding stations in order to fill in the gaps of a particular station, a not appropriate selection will result in un-reliable results and low correlation coefficients. Stations in the neighbourhood of 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. 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 in filled 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 2: Excerpt from an example of a station control list)

| 5 | | Gaps | Neighbours | | | Performed | Regression? | Filled Gaps? | Export Access | |
|----|-------------------------|------|------------|---|---|-----------|-------------|--------------|---------------|--|
| 6 | ARCHERS POST.xls | 14 | 0 | 7 | | 16.01.01 | 24.01.01 | 13.02.01 | 13.02.01 | |
| 7 | ARDENCAPLE FARM.xls | - | C | | | | | 24.01.01 | 13.02.01 | |
| 8 | CASTLE FOREST STN.xls | 251 | 1 | 3 | 7 | 16.01.01 | 02.02.01 | 13.02.01 | 13.02.01 | |
| 9 | CHOGORIA FOREST STN.xls | - | 1 | 3 | 5 | 02.02.01 | | 24.01.01 | 13.02.01 | |
| 10 | CHUKA FOREST STN.xls | 360 | 1 | 3 | 7 | 02.02.01 | 02.02.01 | 13.02.01 | 13.02.01 | |
| 11 | COLCHECCIO.xls | - | 0 | 4 | 7 | | | 24.01.01 | 13.02.01 | |

Figure 2: 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 analysis.

Figure 3: 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 intervals⁷. 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 4). 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.

| 1 | Station | FromDate | ToDate | Rainfall | Mean | 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 4: 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.

⁷ 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.

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 4: 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 1 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 mac-ros 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 3⁸).
3. After you have opened all the necessary files, press the [CollectData] button and the following window will open:

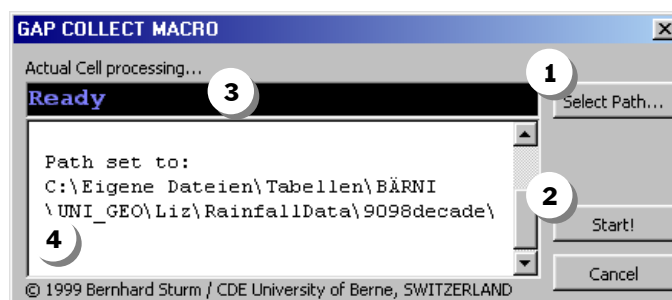


Figure 5: Collect Data dialog

- 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.

⁸ 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.

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.

3.1.4 Step 4: The Regression Analysis

To run the regression analysis, step 3 (chapter 3.1.3 Step 3: Identify Neighbouring Stations and Group Them, page 8) must be completed, and all collected stations must be organised in EXCEL spreadsheets, and saved in the folder 'regression':

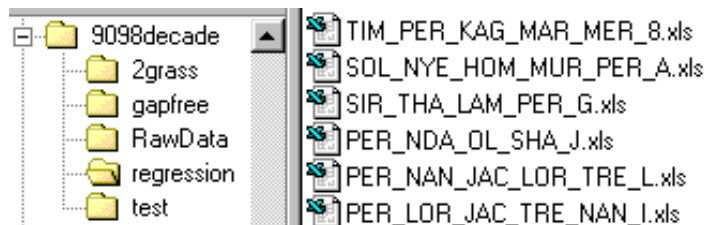


Figure 6: Content of the folder regression after the Collect Data macro was run.

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 co-efficient for each interval⁹.

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 proved to be a very reliable mode, but it is strongly recommended 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.

⁹ 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:

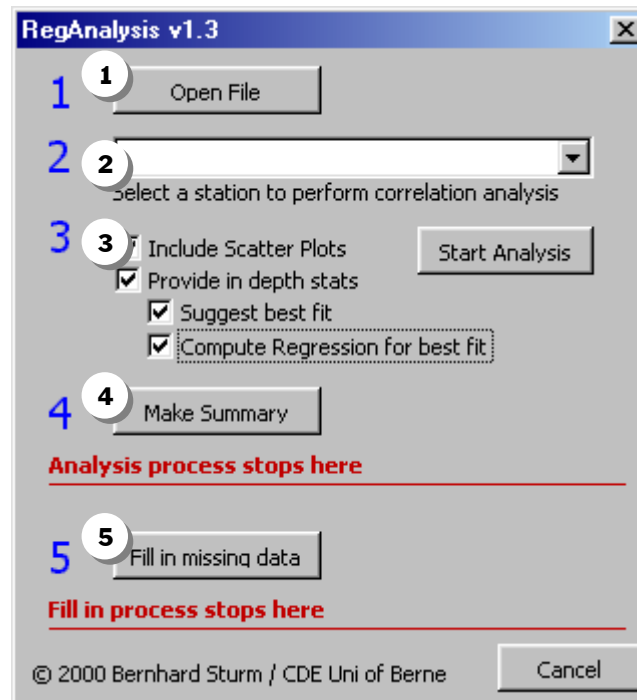


Figure 7: Regression Analysis dialog

- 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 option¹⁰. 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 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. Save 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.

¹⁰ 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 Enough Memory" Error Adding Chart to Workbook, 29.09.1999, www.microsoft.com)

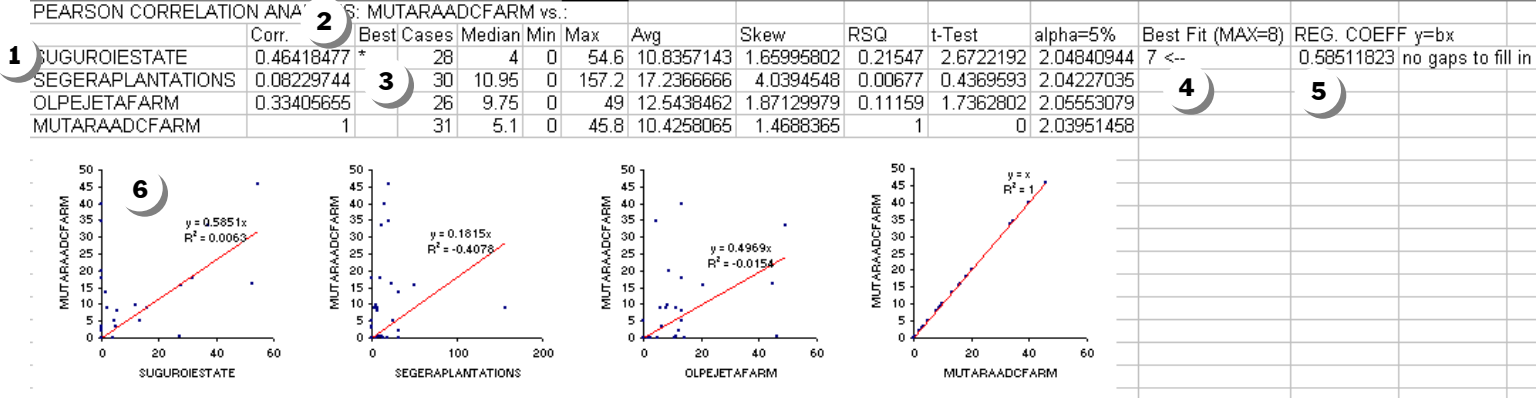
5. Press the [Fill in missing Data] button. A file open dialog will appear, select the first file you would like to complete. After pressing the [Open] button the selected file is loaded and all gaps are eliminated according to the pre-calculated best-fit Regression Coefficient. After completion any inserted value is marked with an asterisk and a comment how many values were inserted (Figure 8) is added.

| 1 | Time | PERIODREF | CHUKAFORE | IRANGIFORE | CASTLEFOR | CHOGORIAF | HOMBFORE | FORESTSTN |
|----|----------|-----------|-----------|------------|------------|-----------|----------|-----------|
| 2 | 21.12.90 | | 60.9 | 2.5 | 10.3 | 67.8 | 2.8 | |
| 3 | 21.12.91 | | 96.99999 | 10.5 | 24.2 | 46 | 0 | |
| 4 | 21.12.92 | | 20.2 | 10 | 10.8 | 20.2 | 0.5 | |
| 5 | 21.12.93 | | 10.5 | 0 | 0 | 11.6 | 31.2 | |
| 6 | 21.12.94 | | 17 | 0 | 13.3 | 0 | 13.8 | |
| 7 | 21.12.95 | | 1.3 | 2.5 | 2.49785344 | 14.8 | 21.8 | * |
| 8 | 21.12.96 | | | 0 | 0 | 21.5 | 30.5 | |
| 9 | 21.12.97 | | 134.2 | 80.5 | 78.3 | 210 | 82 | |
| 10 | | | | | | | | |
| 11 | | | | | | | | |

- 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 8: 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 in-fill procedure with all other stations.



- 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 9: The result worksheet of a regression analysis for a single interval.

3.1.5 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:

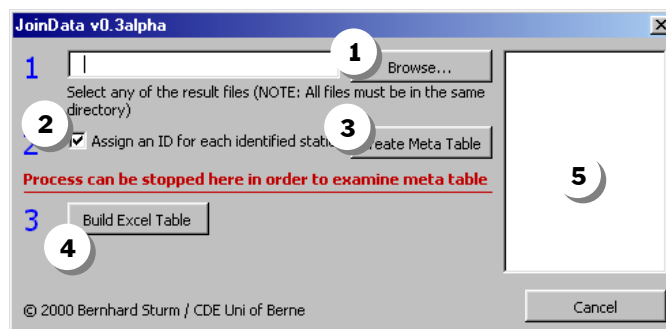


Figure 10: Dialog of the Join Data macro

- 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 station¹¹. 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 5: The first four entries of the joined rainfall values. Note the as-signed 'StationID' that will be used to identify the gauges in the NRM data-base.

9. Select the three columns [StationID], [Date], and [Rainfall] then copy and paste them to an empty table¹² 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: 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.

¹¹ In practice this proved 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 be 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.

¹² It may be necessary that you have to create a new table in access with the properties given in Table 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.

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.

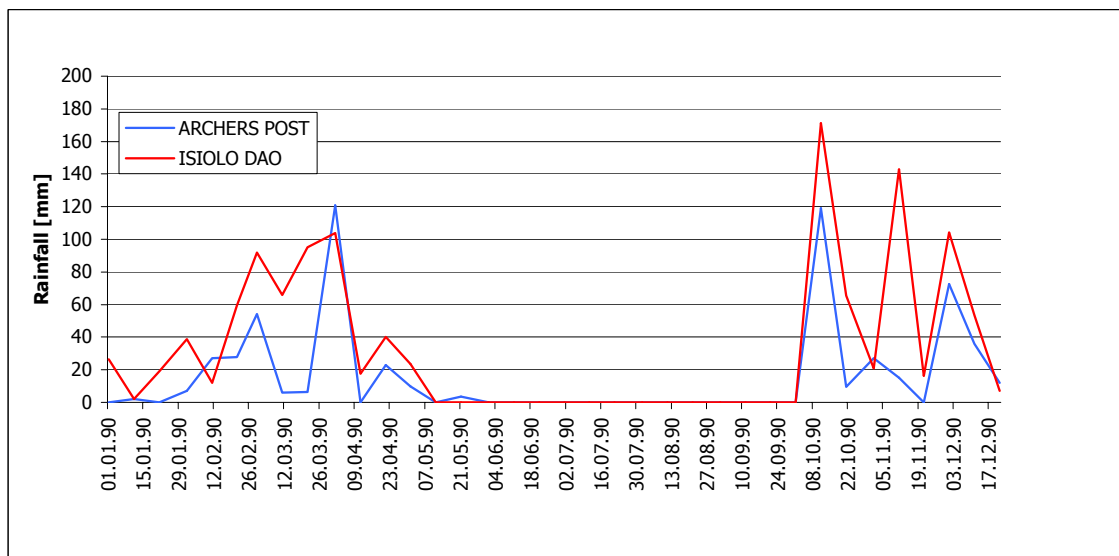


Figure 11: Example of a time domain analysis for the rainfall gauges ARCHERS POST and ISIOLO DAO, showing the rainfall of the year 1990 (1.1.1990 – 31.12.1990, 10-day interval data)

Assuming that the gap-free rainfall data was interpolated in a 10-day interval, one might produce for each single interval a precipitation distribution map. However, for long time series, this could easily end in a very time consuming task. For most applications it is therefore convenient to compute rainfall maps showing the mean spatial distribution pattern in a 10-day interval covering the available period of time.

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 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.

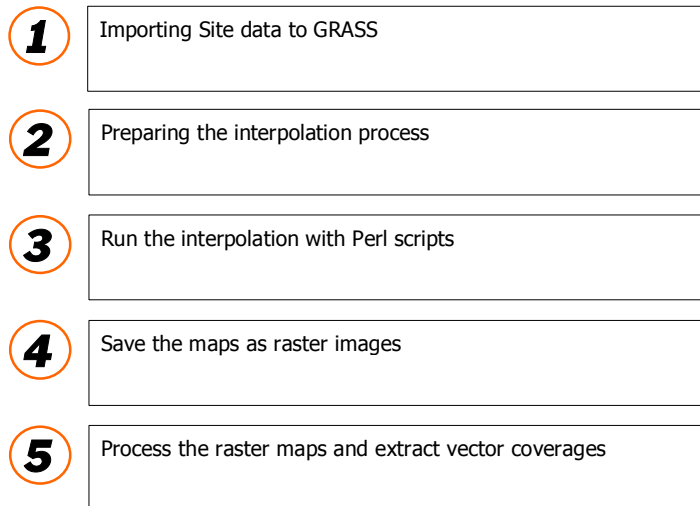


Figure 12: Work-Flow process of the spatial interpolation with GRASS.

Figure 12 illustrates each step of the workflow process of the spatial interpolation.

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. GRASS uses the following ASCII structure:

```

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

```

Figure 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.

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)¹³. It strongly depends upon the application and quality of the site data which model produces the best interpolation. A cross validation of the used rainfall data helps to determine which model represents best the observed values. In this manual we will use the RST model as this proved to produce most reliable results.

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. 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. 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 color map of the interpolated raster files:

¹³ 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 7: Color rules used for raster maps interpolated with 10-day inter-val 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 de-sired 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 15(page 20).

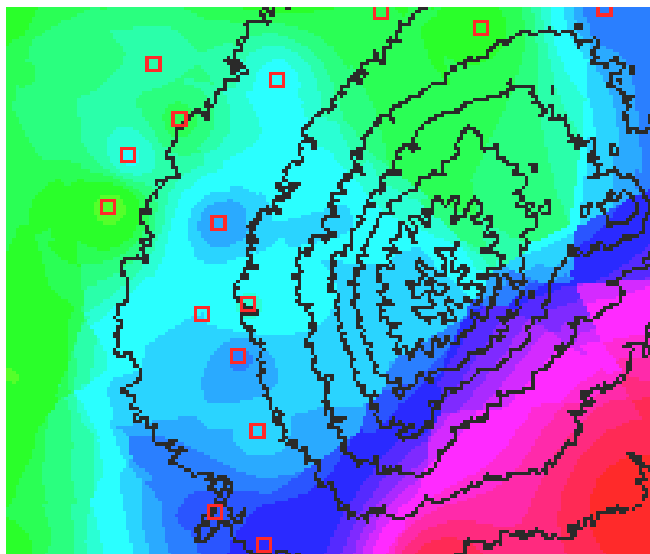


Figure 14: Example of an interpolated raster map with GRASS default color coding (IDW-interpolation, mean values for the month of October from 1961-1970, Mt. Kenya, green=low rainfall, red=high rainfall)

Additional analysis and map extraction based on these maps is possible. GRASS offers many tools to further analyse raster maps¹⁴. Export from GRASS to any graphic tool (such as GIMP) is possible through various ways:

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

¹⁴ 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>)

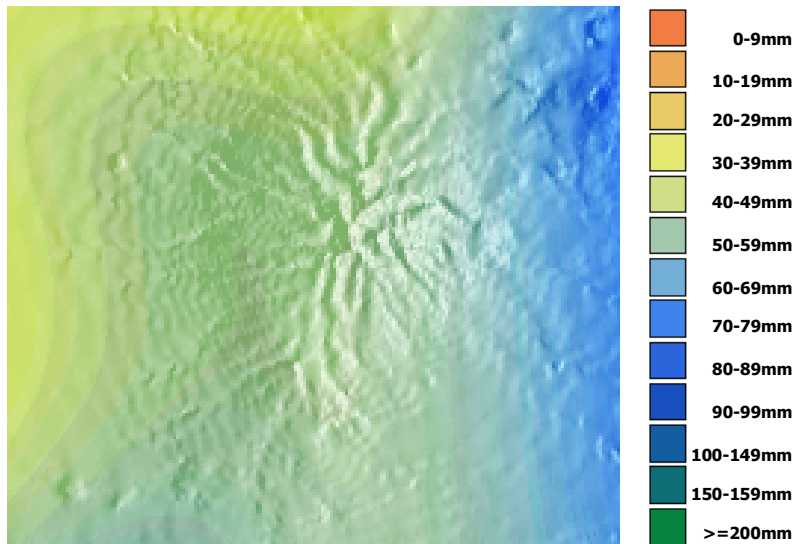


Figure 15: Example of an interpolated raster map with manually added color indexing rules (RST-interpolation, 10-day mean value for the decade 21.3. - 31.3. from 1990 to 1997, Mt. Kenya)

As explained in chapter 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.

4.2.1 Prepare and Run the Interpolation Process

As described in Figure 12 (page 18) before the above-mentioned perl scripts can be run a few things have to be prepared:

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

Surface interpolation can be done by using the script `rsurfst`.

`rsurfst` – The Surface Interpolation

`rsurfst` is designed to automatically import and interpolate rainfall data with the Regularised Splines with Tension model.

Figure 16 describes the working procedures of the script. The site list and the color-map definition files are stored in the same directory as the rainfall ASCII files, whereas the user can enter the parameters for RST interpolation at the beginning of the script.

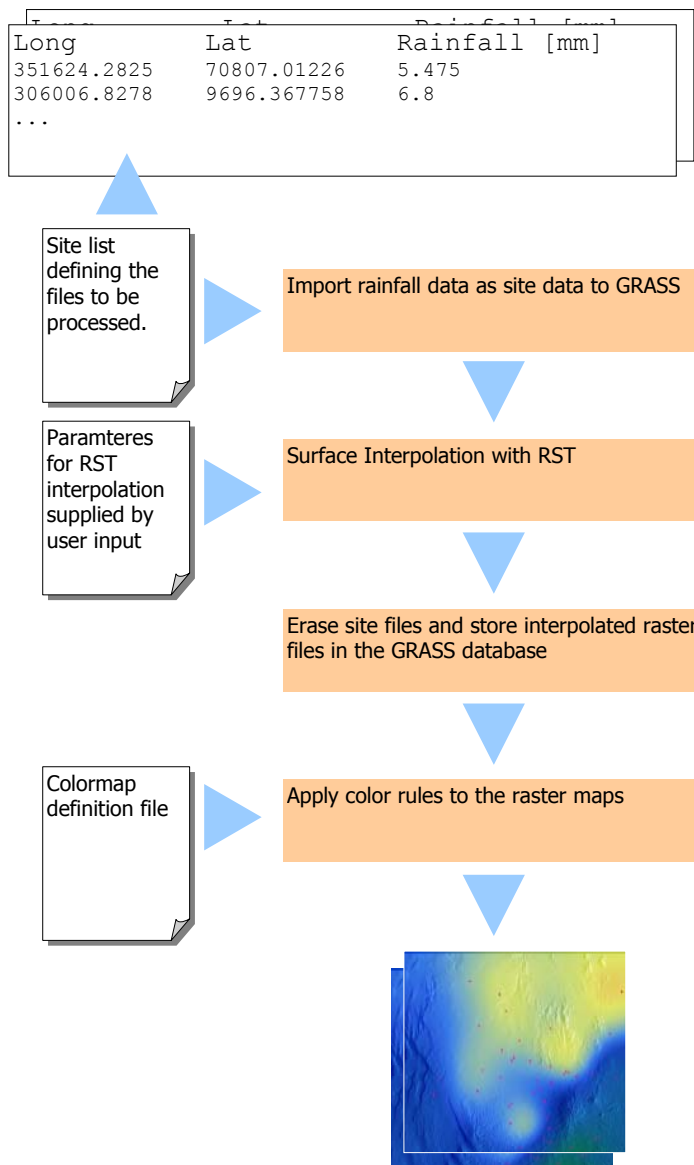


Figure 16: Working procedures of `rsurfst`

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 and 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 color map 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
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 color map file is missing, `rsurfprst` will generate its own default color map 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 purposes¹⁵.

If an input site ASCII file name was "0101.txt", then the interpolated raster map will be named by `rsurfprst` 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 rsurfprst
```

You will then be asked to enter the site list file and the color map file.

While the script is processing the raster maps you can still work with GRASS as only the shell is occupied by `rsurfprst`.

4.2.2 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 Program¹⁶) 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

¹⁵ '9097' refers to the time horizon of the decadal precipitation data sets: 1990-1997

¹⁶ GIMP: www.gimp.org

the map production process (Figure 17). The GRASS command for raster export to TIFF is `r.out.tiff`¹⁷.

When producing maps with rainfall isolines, vector coverages (Figure 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]
```

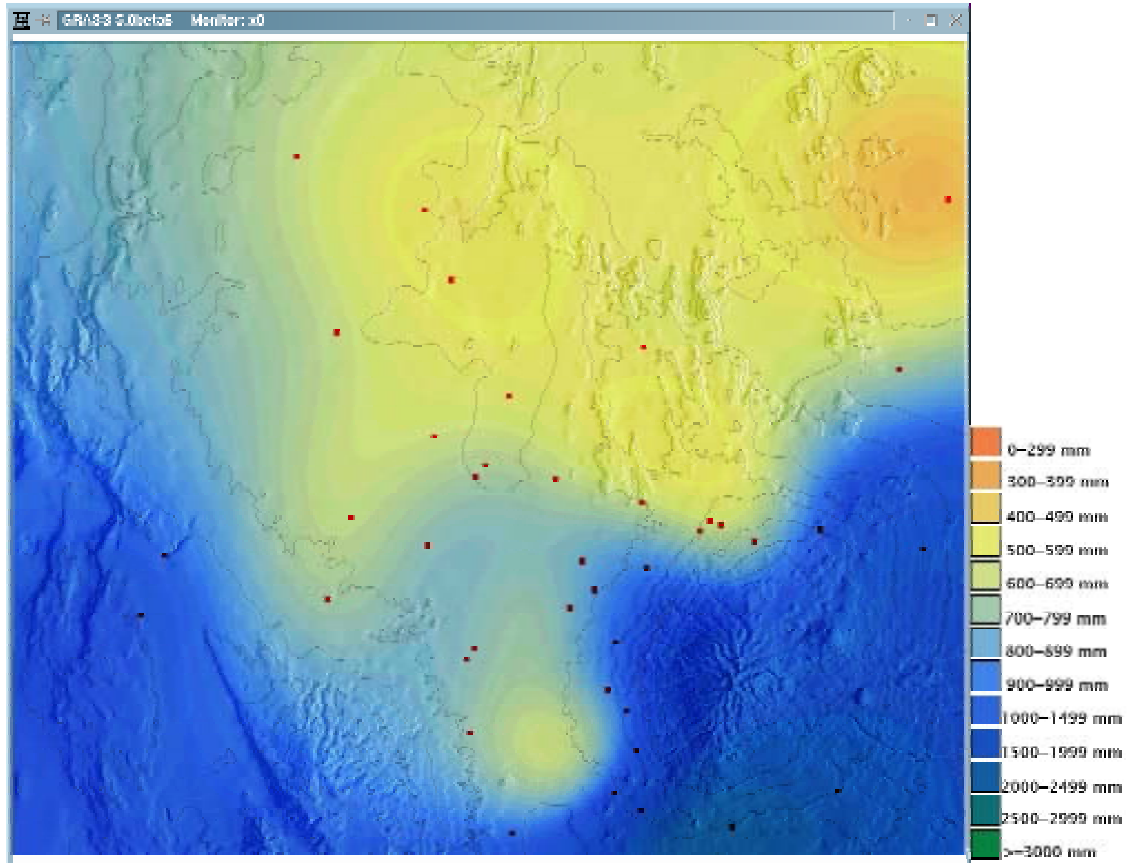


Figure 17: Superimposed raster, site and vector coverages with GIMP. Mean annual rainfall distribution 1967-1997

[Output name] in `r.contour` is the resulting vector coverage. This vector file can either be 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 AS-CII vector file with `v.out.ascii` (this is because export to the DXF format works only with ASCII vector files).
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.

¹⁷ 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.

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.

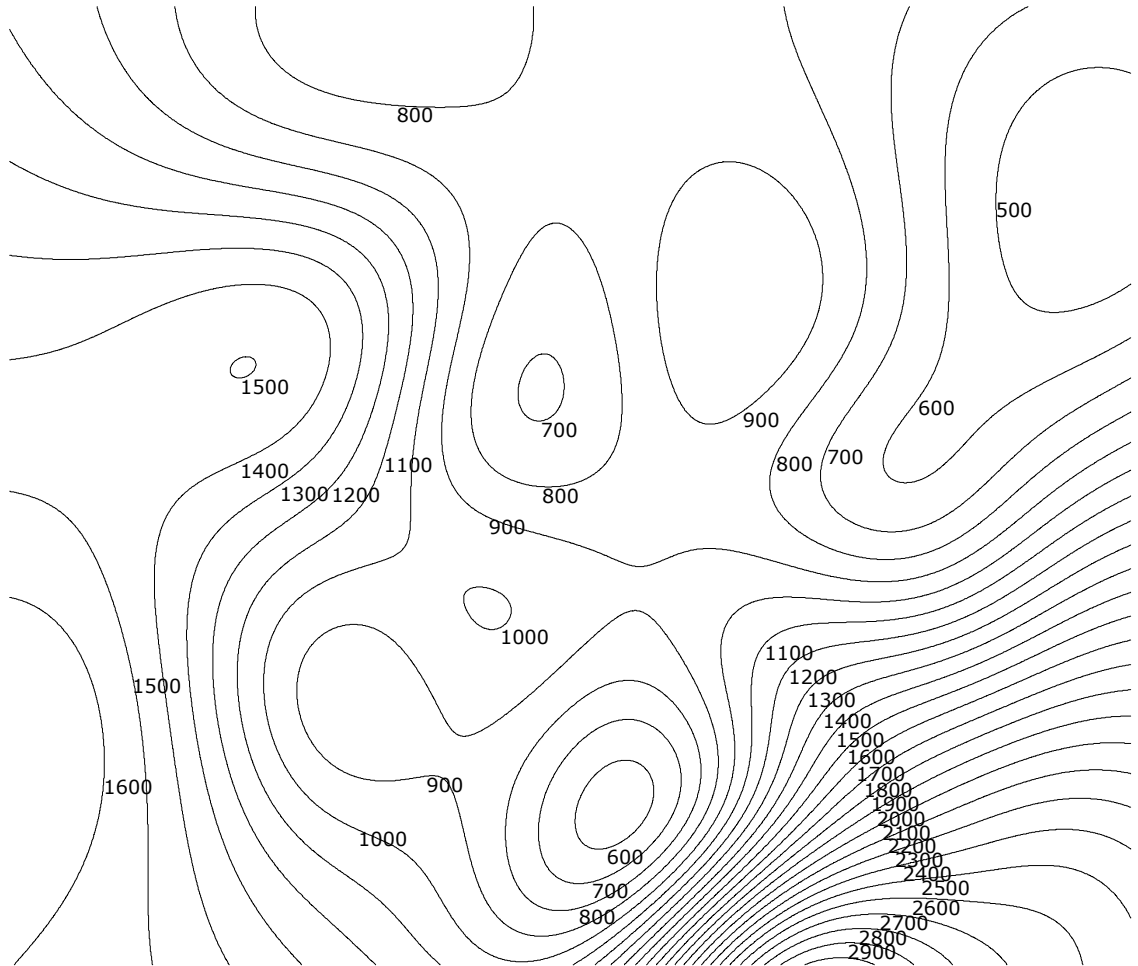


Figure 18: Extracted rainfall isoline vector coverage from a raster file with r.contour. Mean annual values in mm for the year 1977.

5 Literature

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.

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

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

Sturm, Bernhard, 2001. 'Development and Use of Long and Short-Term Precipitation Interpolation Models in the Ewaso Ng'iro Basin', Bern, *University of Berne, Switzerland*, Master Thesis
