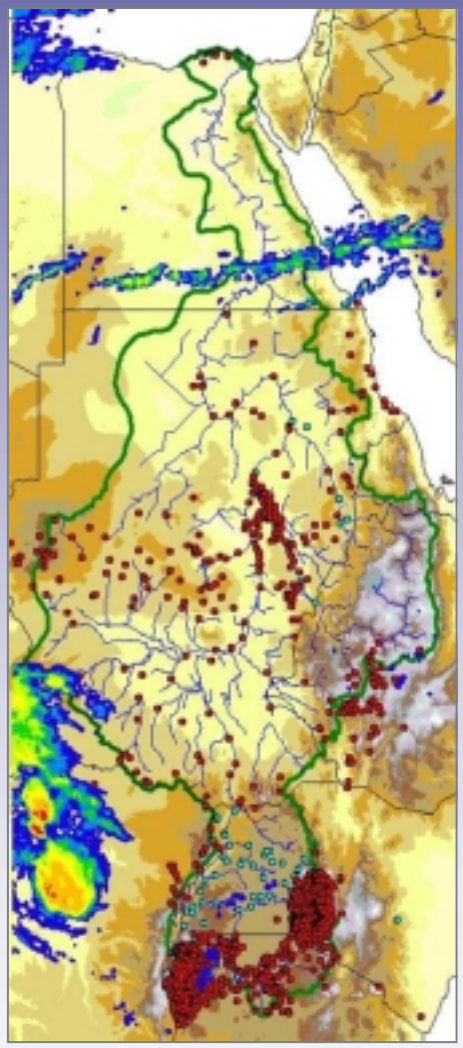
Nile Decision Support Tool
Remote Sensing

Burundi
Congo
Egypt
Eritrea
Ethiopia
Kenya
Rwanda
Sudan
Tanzania
Uganda



Developed collaboratively by

The Nile Basin Nations,

The Georgia Water Resources Institute at the Georgia Institute of Technology,

and

The Food and Agriculture Organization of the United Nations

June 2003

Nile Decision Support Tool (Nile DST)
Remote Sensing

Report developed by

Carlo De Marchi
Research Associate

Aris Georgakakos
Project Director

Georgia Water Resources Institute
School of Civil and Environmental Engineering
Georgia Institute of Technology

In collaboration with

The Nile Basin Nations

and

The Food and Agriculture Organization (FAO)
of the United Nations
Nile Basin Water Resources Project (GCP/INT/752/ITA)

June 2003

Acknowledgements

This report and associated software were developed by the Georgia Water Resources Institute (GWRI) at the Georgia Institute of Technology as part of the Nile Basin Water Resources Project (GCP/INT/752/ITA). This project was funded by the Government of Italy and was executed for the Nile Basin nations by the Food and Agriculture Organization (FAO) of the United Nations.

The GWRI Director and project staff are grateful to the Nile Basin nations (Burundi, Congo, Egypt, Eritrea, Ethiopia, Kenya, Rwanda, Sudan, Tanzania, and Uganda), their focal point institutions, their Project Steering Committee (PSC) members, and their National Modelers for entrusting us to work with them in this important basin-wide project. The development of databases, models, technical reports, software, and user manuals are key but not the only project accomplishments. Even more important are the evolving contributions relating to people and the difference the project is poised to make in data and information sharing, developing a common knowledge base for policy debates, and long term capacity building.

GWRI is also grateful to the Government of Italy and to FAO for sponsoring this project and for providing dependable logistical and technical support through the FAO offices in Rome and Entebbe.

It is our hope that the Nile DST effort will contribute in some positive way to the historic process of the Nile Basin nations to create a sustainable and peaceful future.

Aris Georgakakos
GWRI Director
Atlanta, June 2003

Disclaimer and Copyright Notice

Table of Contents

Remote Sensing of Rainfall

1. Overview

The need for satellite-based rainfall estimation arises from the scarcity of conventional ground data in large areas of the Nile Basin. On the other hand, satellite images of the Nile Basin are readily available at half-hour intervals from the Meteosat geosynchronous satellite. The premise of the remote sensing modeling effort presented herein is that useful rainfall estimation procedures can first be calibrated for areas with sufficient satellite-ground data and then be extended over the rest of the basin.

The present version of the Nile DST includes two remote sensing methods based on the Cold Cloud Duration (CCD) approach:

- A classical CCD approach that uses thermal infrared images;
- A modified CCD approach that uses both visible and infrared images.

The conceptual basis of CCD is that (**1**) precipitation over a pixel during a given period is directly proportional to the number of hours the pixel thermal infrared temperature is below a given value and (**2**) the average precipitation over a region can be estimated by averaging the precipitation over all region pixels. The CCD approach is used in this Nile DST version because it is relatively simple, robust, computationally fast even for large basins, and most likely familiar to the potential Nile DST users. These characteristics give CCD a clear advantage in the early stages of the Nile DST implementation and usage. Furthermore, CCD application provides valuable insights regarding data issues and the plausibility of developing more sophisticated, potentially more accurate, but also more data-intensive methods.

This report contains 13 chapters. Chapter 2 briefly describes the Meteosat satellite system, the spatial and temporal availability of Meteosat digital images, and the process to extract the data used for CCD calibration and validation. The chapter also describes the spatial and temporal availability of rain gage precipitation. It further summarizes the procedures adopted to ensure the quality of these data. Chapter 3 illustrates the characteristics of the adopted CCD models and the general characteristics of the calibration and verification process. The details of CCD model calibration and validation in different sub-regions of the Nile River basin, from Lake Victoria to the Nile River Delta, are given in Chapters 4 to 11. Chapter 12 discusses an application of the remote sensing module in the Lake Victoria basin. CCD model performance, unresolved issues, data availability, and further developments in satellite based precipitation estimation are summarized in Chapter 13.

2. Satellite and Rain Gage Data Used for Model Calibration and Testing

2.1 Meteosat Data

The Meteosat System

Meteosat is a family of meteorological satellites collocated on geostationary orbit 36,000 km above the Equator, at the intersection with the Greenwich meridian (i.e., at 0°E, 0°N). From this position, the Meteosat satellites are able to monitor the weather evolution over South America, the Atlantic Ocean, Europe, Africa, the western Indian Ocean, and most of the Middle East (Figure 2.1). Quantitative analyses, however, can be performed over a limited part of this view.

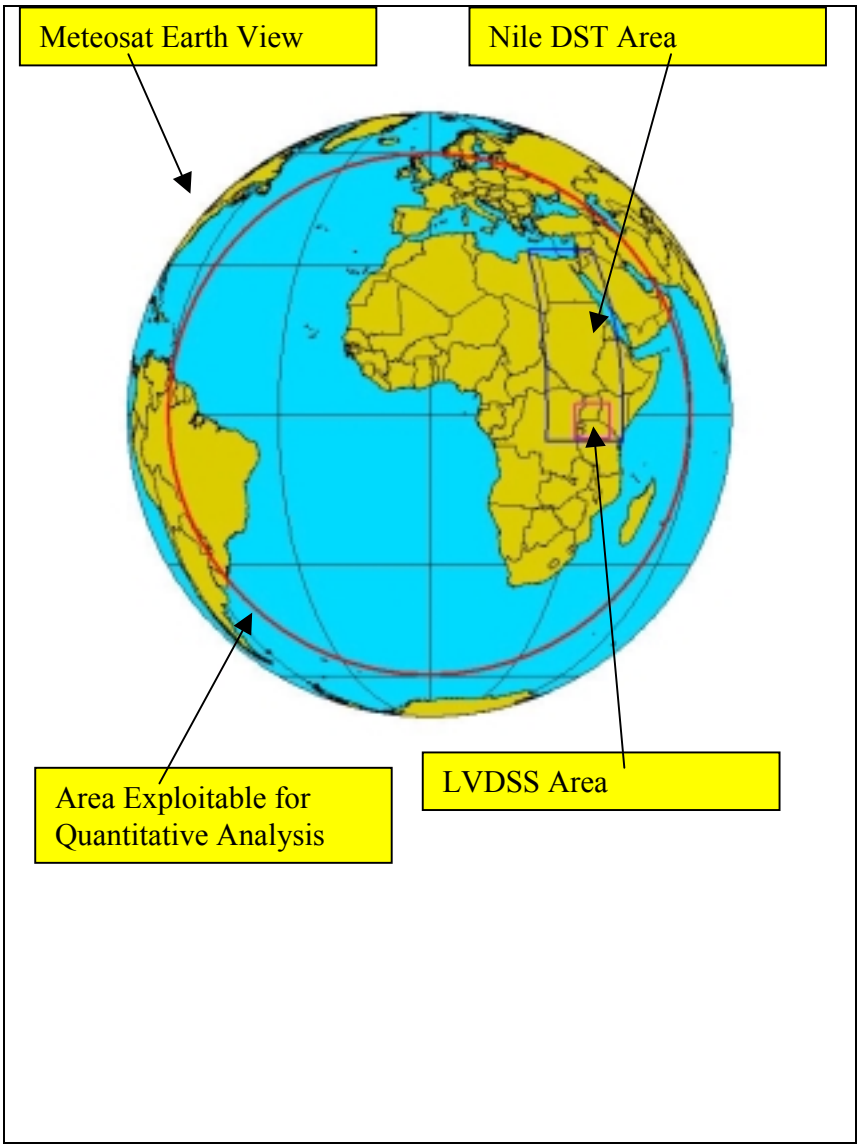


Figure 2.1: Full Disk view of the Meteosat Satellites.

The Meteosat satellites main sensor is the VISSR, an imaging radiometer able to frame a digital image (i.e., a two-dimensional array of numerical values) of the full earth disk every 30 minutes in the three channels described in Table 2.1.1. Each element (pixel) of the image corresponds to the radiance coming from a different area of the view shown in Figure 2.1. The size of this area is approximately equal to that indicated in Table 2.1.1 at the sub-satellite point, but increases moving towards the periphery of the view. The observed radiance is codified into a value between 0 and 255 (Digital Counts).

Table 2.1.1: Characteristics of the VISSR imaging radiometer (Eumetsat, 2000).

Channel	Electromagnetic band (μm)	Sub-satellite spatial resolution (km)	Direct and indirect meteorological information provided
VIS	0.5 –0.9	2.5 x 2.5 (5.0 x 5.0)	Albedo (daytime) – Thickness of clouds.
WV	5.7 – 7.1	5.0 x 5.0	Water vapor in the middle troposphere -- Thickness of clouds/Overshooting
IR	10.5 –12.5	5.0 x 5.0	Thermal infrared temperature – Surface temperature and cloud top's elevation

The digital images captured by the sensors aboard the Meteosat satellites are transmitted to the Eumetsat headquarters in Darmstadt, Germany, for rectification and other processes. From there, images are transmitted back to the Meteosat satellite, which broadcasts them to the ground receiving stations. However, due to restrictions on the receiving hardware and software, only two channels are broadcasted at a time. Normally, these two channels are the thermal infrared (IR) channel and either the visible (VIS) or the water vapor (WV) channel, according to a predefined schedule. Further, the VIS channel is transmitted at half its nominal spatial resolution, by sub-sampling the full resolution image. VIS images are available between 5:30 GMT and 22:00 GMT, while the WV images are available between 22:30 GMT and 5:00 GMT. It is noted that at 11:30 GMT high definition VIS and IR images of the Mediterranean basin are transmitted in place of the Meteosat Full Disk images.

Over the period 1992-2003, four different Meteosat satellites, Meteosat-4 through Meteosat-7, were in service and provided radiation images. While the type of image remained the same as described in Table 2.1.1, each satellite carries sensors that require specific digital count – radiance relations. Further, the performance of each satellite sensor varies during their operational life because of sensor degradation and environmental influence, requiring a continuous recalibration of the count-radiance relationships. The satellite specific sensor count-radiance relations and the calibration coefficients can be retrieved form the Eumetsat web site at www.eumetsat.de. Calibration coefficients are embedded in the satellite broadcasted digital images as auxiliary parameters. The quality of the image (i.e., the noise level) is also different from satellite to satellite.

Conversion of Digital Images into Nile DST Grid Files

The files holding the satellite digital images came in three different formats, namely B.U.R.S.L. Autosat 3.1, B.U.R.S.L. Autosat Block 5, and Eumetsat OpenMTP. A common feature of all three formats is that they hold much more information than what is needed for the Nile DST, dramatically increasing storage requirements. Furthermore, extraction of the data from the Meteosat files is a very cumbersome process that can easily fail if the satellite digital files are for some reason corrupted. It is also often necessary to merge data files from different formats for filling gaps in the satellite coverage. These reasons render the satellite digital files unsuitable for direct input to the Nile DST or any other operational software. Instead, it was decided to transform the satellite files into a nimbler format called Nile DST Grid Format. Regardless of their original format, the process for converting a satellite digital image file into Grid Format for use the Nile DST framework involves several steps:

1. Separation of digital images into IR, VIS, and WV images;
2. Extraction of the digital counts corresponding to pixels falling within the Nile Basin from each image;
3. Elimination of missing and noisy lines;
4. Extraction of ancillary information (such as calibration coefficients, satellite number, etc.);
5. Geo-referencing of the extracted pixel (i.e., transformation of the pixel Meteosat image coordinates into longitude and latitude) and re-sampling according to the Nile DST grid;
6. Conversion of digital counts to blackbody temperatures (IR and WV channels) and albedo (VIS channel);
7. Filtering of invalid values (IR temperature lower than 185 °K or higher than 333°K; WV temperature lower than 180 °K or higher than 270 °K, and VIS albedo lower than 0 or higher than 100);
8. Filtering of each channel pixel values outside the range $AVG_{5x5} - STDEV_{5x5} <$ pixel value $< AVG_{5x5} - STDEV_{5x5}$, where AVG_{5x5} is the average of the values over a five by five degree pixel square centered on the pixel of interest, and $STDEV_{5x5}$ is the standard deviation over the same area;
9. Codifying of the data into Nile DST Grid file holding the 48 digital images normally transmitted by Meteosat during one day. The day is beginning with Meteosat slot 1 image (00:25 GMT) and terminating with Meteosat slot 48 (23:55 GMT). Missing images are substituted with images composed by the *NO DATA* value (-9999.0).
10. Detection of corrupted, spatially or temporally displaced, or duplicated images through analysis of the correlation between consecutive IR, VIS, and WV images and their elimination;
11. Merging of Nile DST Grid files obtained from satellite digital images coming from different sources (optional);
12. Correction of the displacement error.

It is noted that only Steps 1, 2, and 11 depend on the original file format.

Meteosat Imagery Available for Model Calibration and Testing

The availability of satellite imagery and contemporaneous ground data varies widely in the basin since imagery has been gathered from a variety of sources and in a variety of formats. Furthermore, some of the satellite imagery could not be processed because it was availed at a very late stage of the Nile DST development. Table 2.1.2 shows the periods for which Meteosat imagery was used for the CCD calibration and verification in different parts of the basin.

Table 2.1.2: Meteosat data available for different areas of the Nile Basin and their format.

Period	Area 5.0S÷10.0N	Area 10.0N÷11.5N	Area 11.5N÷19.0N	Area 19.0N÷22.0N	Area 22.0N÷33.0N	Source
01Jan1996 26Nov1996	X	X	X	X		1
27Nov1996 31Dec1996	X	X	X	X	X	2
01Jan1997 11Jun1997	X	X	X	X	X	2, 3R
12Jun1997 16Jul1997	X	X	X	X	X	2, 3R
17Jul1997 19Dec1997	X	X	X			2
20Dec1997 31Dec1997	X	X	X	X	X	2
01Jan1998 31Jul1998	X	X	X	X	X	3
01Aug1998 31Dec1998	X					3R, 4

1 Nile Forecast Center, Cairo (Egypt) – B.U.R.S.L. Autosat 3.1
2 Nile Forecast Center, Cairo (Egypt) – B.U.R.S.L. Autosat Block 5
3 Eumetsat, Darmstadt (Germany) – OpenMTP
3R Eumetsat, Darmstadt (Germany) – OpenMTP Replacement of missing slots
4 Meteorological Office, Entebbe (Uganda) – B.U.R.S.L Autosat Block 5

Data from sources 1 and 2 are in the form of two digital images, one for the northern part of the Nile basin and one for the southern part. On two occasions, the tapes holding the images of one of these two sub-regions were unreadable. This is the reason that images coming from the same source have different spatial extensions. The following sections give a brief account of the translation of each set of data into the Nile DST Grid format.

Satellite Images 01Jan1996-26Nov1996 (Source 1)
The satellite digital images compiled for this period are in B.U.R.S.L Autosat 3.1 format. This format dedicates a different file for each image received at each time slot (i.e., if two IR and WV are received at a given slot, two different files are generated). Autosat 3.1 files can hold only images with a fixed size of 512 lines (for 512 pixels), thus limiting the area they can cover. Two files for each channel are used to hold the Meteosat data for one time slot. The first file is labeled WN window and covers the Lake Victoria basin, northern Uganda, southern Ethiopia, and

southern Sudan. The other file is called BN window and covers all eastern Sudan, Ethiopia, and Eritrea. Egypt is not covered by this block of data (Figure 2.2).

The files belonging to this block of data are very noisy: around 30% of available images are affected by one or more lines of noise like the one shown in Figure 2.2. Furthermore, these lines are not revealed by the auxiliary data contained in the satellite files and dedicated for this purpose. This indicates that the noise is probably introduced by the receiving system and not by the satellite.

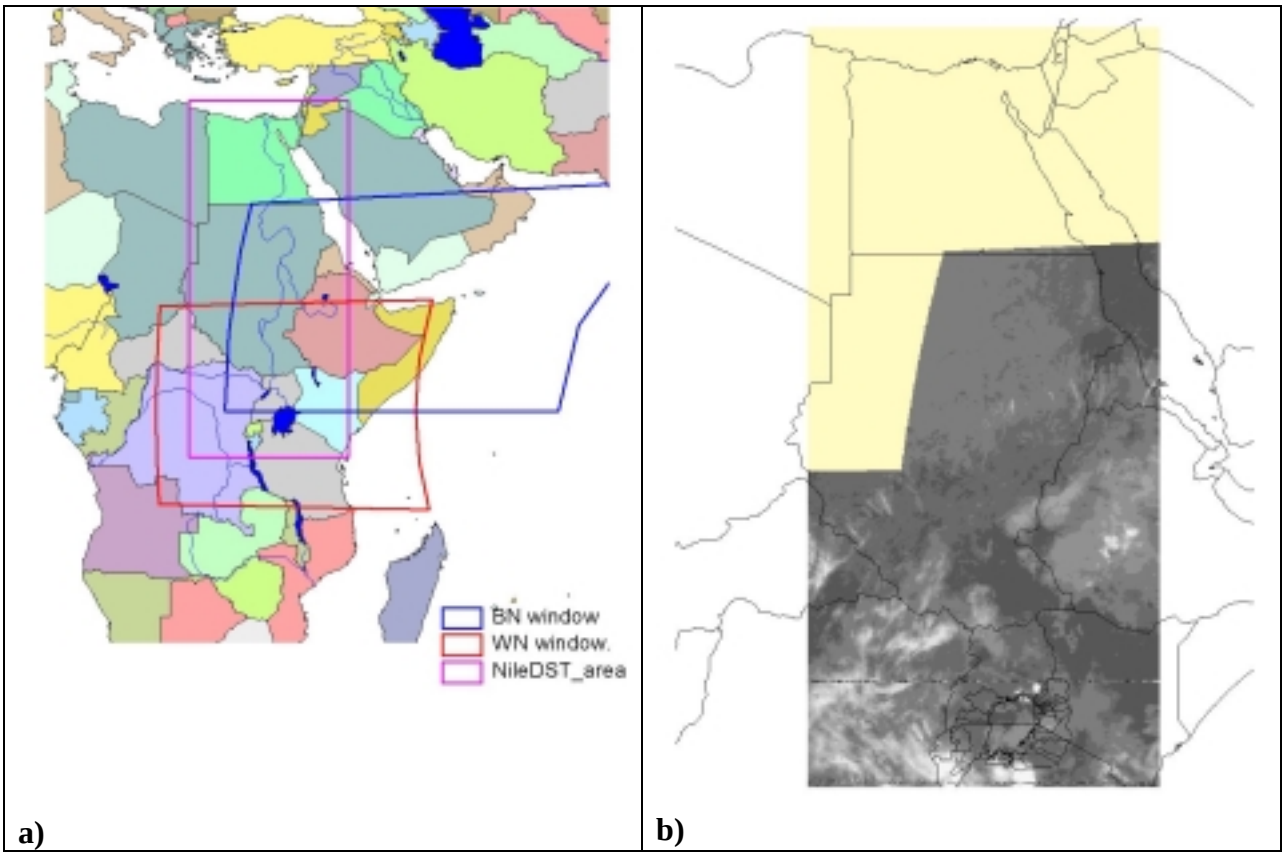


Figure 2.2: a) Areas covered by the Meteosat BN and WN windows of the Nile DST. **b)** Example of the conversion of IR BN and WN windows into the Nile DST Grid Format. A line of noise slightly north of Lake Victoria is maintained for illustrative purposes. No data are available in the part of the Nile DST area filled in yellow.

The noise lines could have an undesirable impact on precipitation estimation. Such lines are composed of values randomly varying from maximum to minimum count and introduce false cloudy pixels in otherwise "clear sky" areas or vice versa. The filters normally used to eliminate noise (such as the one used in Step 8 of the translation process) are not effective against this type of noise. Fortunately, these noisy lines coincide with lines of the Meteosat images, facilitating the process of screening them out. Along a line of Meteosat data, pixel values typically vary gradually with sudden jumps only at relatively few instances such as at the interfaces between clouds and clear sky regions or between sea and land areas. This means that the mean square

difference between the values of adjacent pixels across a data line is relatively low. In a line of noise, however, adjacent pixels take random values yielding a much larger value of this statistic. Thus, Meteosat lines can be screened out as noisy if the mean square difference between the adjacent pixel values is above a certain threshold. However, the threshold that screens out most noisy lines without mistakenly eliminating normal lines requires some experimentation, especially for the visible channel which is intrinsically noisier than the infrared and the water vapor channels.

The temporal coverage of this data block is generally good for January, February, July, August, and November. In most days, the difference with full coverage is contained to one time slot. As mentioned, slot 23 is not transmitted by Eumetsat to local PDUS receivers. WN files have only two days of corrupt data (22Jan1996 and 17Feb1996), while the BN files include a larger gap (28Jan1996 – 29Jan1996 and 26Nov1996 – 27Nov1996).

Most days in March and September, and many days in April and October have an image temporal density of less than 90%. However, the difference with the best practically obtainable slot density is of just one or two slots (slot 23 being one of them). During these months, the Sun cannot illuminate the geostationary satellites for some hours around the sub-satellite point at midnight because of Earth interference. This effect prevents the satellite to generate electrical power and its sensors malfunction. Consequently, the satellite coverage of these days cannot be significantly improved.

For some days (22Jan1996, 05Apr1996 – 08Apr1996, 19Apr1996 – 21Apr1996, etc.), the density of available images is numerically unacceptable (<60%). However, for some of these days, images are not missing on consecutive slots, but they are missing every other slot. Because of this, a correct assessment of precipitation during those periods may still be possible.

BN files are missing during the period from 13Jun1996 to 15Jun1996.

Satellite Images 27Nov1996-31Dec1997 (Source 2)

The satellite digital images of this data block are in B.U.R.S.L Autosat Block 5 format. In this format, two or sometimes three different images received during a single time slot are held within a single file. Autosat Block 5 files can hold only images with a user specified size, but two files are still used to hold the Meteosat data contained in each time slot. The first file is labeled NI and covers the part of the Nile basin south of 19°N, while the second file is labeled EG and covers the part north of 13.5 °N (Figure 2.3).

The quality of these images is similar to the 1996 data with approximately 20% of the images being affected by faulty lines. However, the faulty lines are typically identified by the auxiliary file parameters, making filtering unnecessary. Noisy areas, however, are sometimes present at the border of large bands of missing data. Furthermore, the auxiliary data embedded in the satellite files are often corrupted, and the calibration and satellite identification parameters had to be externally supplied to the translating software.

The temporal coverage of the NI files is very good in December 1996, and in January, February, July, August, November and December 1997. In most days, the difference with the best

available temporal density is contained in one or two slots. EG files in the period 17Jul97-20Dec97 are completely missing.

Most days in March, April, September, and October of 1996 have less than 90% coverage. The satellite coverage of these months cannot be significantly improved because of the solar eclipse effect.

As in 1996, there are some days in 1997 (23May1997, 16Apr1997, etc.) for which the density of available images is numerically unacceptable (<60%). However, OpenMTP digital images were ordered from Eumetsat and have already been incorporated to replace the longest stretches of missing slots in the period from 01Jan1997 to 11Jun1997 (source 3R).

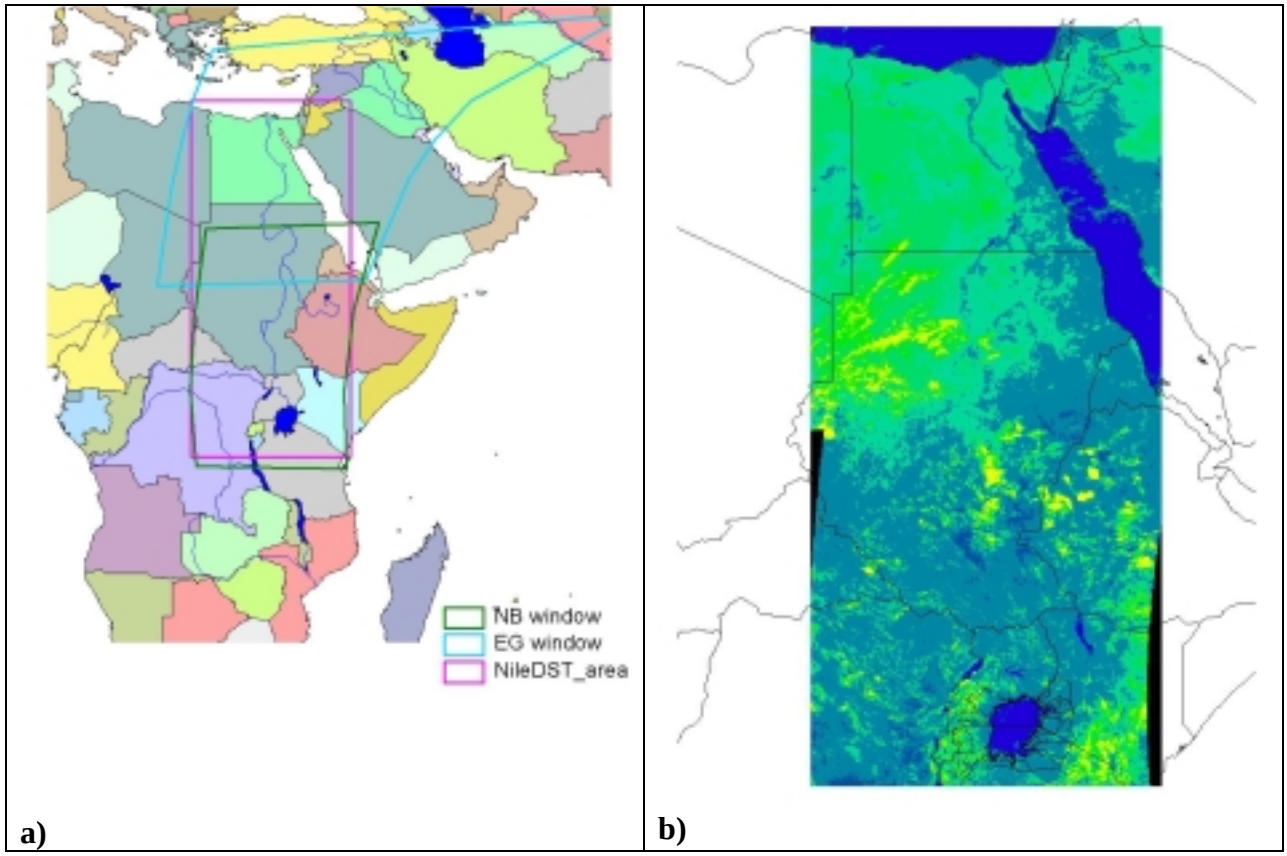


Figure 2.3: a) Areas covered by the Meteosat NI and EG windows and the Nile DST. **b)** False color VIS image derived from the NI and EG windows. No data are available for the parts of the figure filled in black.

Satellite Images 01Aug1998-31Dec998 (Source 4)

The satellite digital images of this data block are in B.U.R.S.L Autosat Block 5 format as for source 2. Two windows are used here as well. The first window, labeled UG, is used for 46 out of 48 slots and covers the part of the Nile basin south of 10.5°N (Figure 2.4). The other window, labeled OV, is used only for slot 48 and covers the entire Full Disk View (not shown).

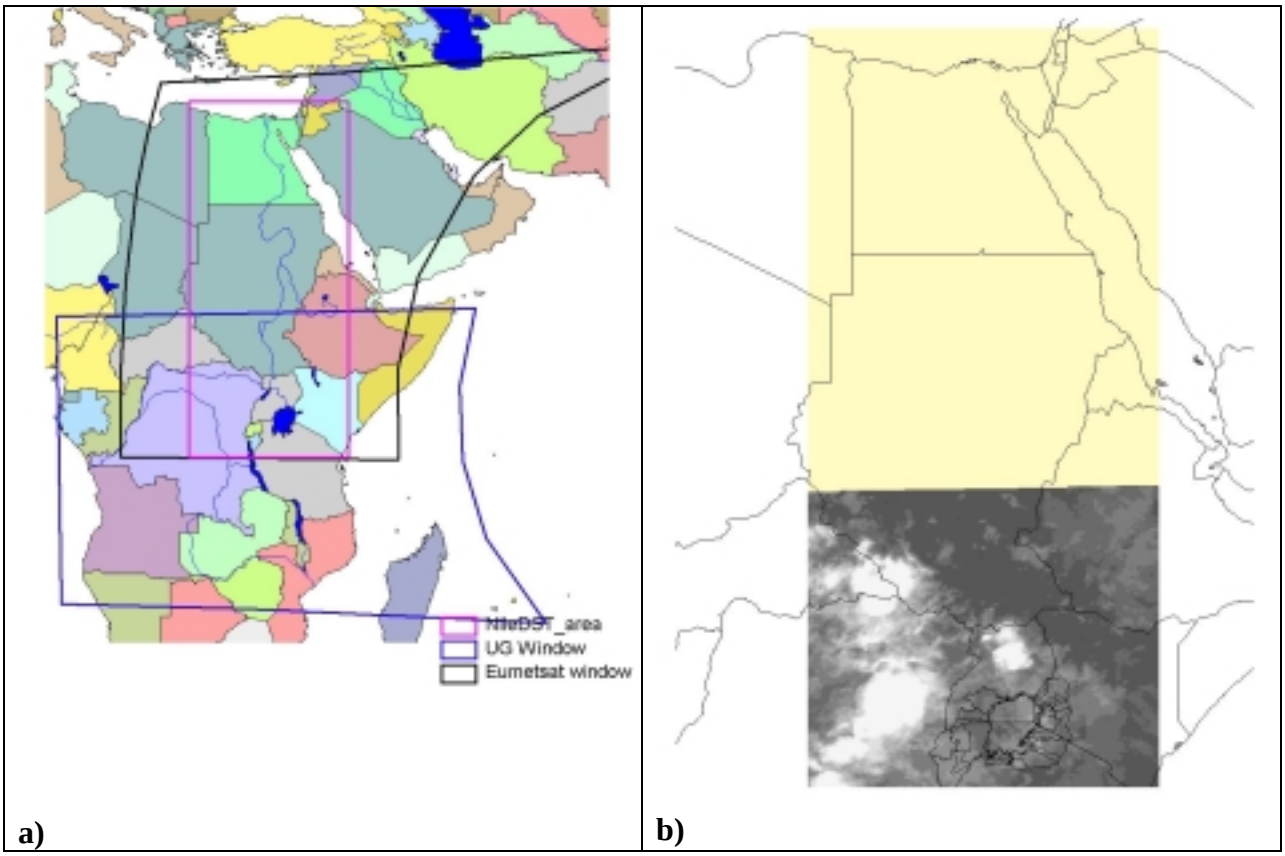


Figure 2.4: Areas covered by the Meteosat UG and Eumetsat windows in the Nile DST. **b)** IR image derived from the NI and EG windows. No data are available for the parts of the figure filled in yellow.

The quality of these images is much better than that for 1996-1997. Noisy or missing lines are practically absent, and the auxiliary data embedded in the satellite files are always well preserved.

The *temporal* coverage of the UG files, however, was not very good. Only 66% of the slots were available during the entire period. Missing images concentrated in September, November, and December. Replacements of these missing slots from Eumetsat were acquired and merged with the existing data producing a perfect coverage for the portion of the Nile basin south of 10.5°N.

Satellite Images 01Jan1997-11Jun1997, 01Jan1998-31Jul998, and 01Aug1998-31Dec1998 (Source 3 and 3R)

The Eumetsat Meteosat Archive facility delivers satellite images in the OpenMTP format. In this format, all three channels are normally available at every time slot, each contained in a different file. Further, the VIS channel is given at full resolution (i.e., at 2.5 x 2.5 km resolution). The auxiliary data coming with the digital images are many more than for the Autosat formats. For these reasons, the Eumetsat files need much more computer storage than a comparable set of Autosat files. On the other hand, since these files have undergone rigorous inspections at MARF, they are almost error free. Unfortunately, these data files could be

delivered only at the rate of one month worth of data every two to three months, placing constraints on their use in this Nile DST version.

2.2 Rain Gage Data

A satellite based rainfall estimation procedure such as the CCD method needs to be calibrated with and validated against rain gage data. Consequently, the reliability and distribution of the rain gages is very important.

The CCD methods included in this Nile DST version have been calibrated using the December 2002 Nile Data Base (NBD-Dec02-2), with the exception of Sudan where most station coordinates reported in the database appear to be in error. In this case, the NBD-Oct02 version of the database was used. The databases have been integrated with some additional data and have been extensively inspected for eliminating as many erroneous or ambiguous records as possible.

Inclusion of Additional Precipitation Records to the NBD Database

Precipitation data for a few tens of stations in Tanzania, Kenya, and Ethiopia have been added to fill the gaps in rain gage coverage in the NBD database (Table 2.2.1 and Table 2.2.2).

Table 2.2.1: Precipitation data added to the NBD-Dec02-2 Kenya and Tanzania databases.

Country	HMStationID	HMName	Lon	Lat	Period
Kenya[1]	60129	BUSIA FARMERS T. C.	34.100	0.470	1992 1993 1998
Kenya[1]	60398	KISUMU MET. STATION	34.750	-0.100	1992 1993 1998
Kenya[1]	60456	AHERO IRRIG. R. S.	34.930	-0.130	1992 1993 1998
Kenya[1]	60475	KIBOS SUGAR R. S.	34.820	-0.030	1992 1993 1998
Kenya[1]	60647	KORU COFFEE BOARD S. S.	35.280	-0.130	1998
Kenya[1]	60654	KERICHO TIMBILIL	35.350	-0.350	1992 1993
Kenya[1]	90010	BUKOBA MET. STATION	31.820	-1.330	1998 1999
Tanzania[2]	90022	KAYANGA HYDROMET	31.170	-1.530	1998
Tanzania[2]	90024	RUBAFU	31.830	-1.050	1998 1999
Tanzania[2]	90038	KYAKAKERA	31.450	-1.300	1998 1999
Tanzania[2]	90045	MUSOMA MET.	33.800	-1.500	1998 1999
Tanzania[2]	90062	TARIME HYDROMET	34.330	-1.330	1998 1999
Tanzania[2]	90070	BIHARAMULO HYDROMET	31.300	-2.630	1998
Tanzania[2]	90075	MWANZA AIRFIELD	32.920	-2.470	1998 1999
Tanzania[2]	90088	SUMVE COLL. OF EDU.	33.220	-2.770	1998 1999

1 Kenya Ministry of Water
2 Tanzania Ministry of Water

NBD 1998 precipitation records in Kenya and Tanzania are not reliable, creating a serious constraint for CCD calibration. The new precipitation records have only slightly improved the situation.

Unlike the previous cases where all new data are associated with stations already existing in the NBD database, most of the new Ethiopia precipitation records were compiled for stations not present in the NBD (Table 2.2.2). Only one new station could be used for comparing the new data with NBD-Dec02-2 data. The comparison showed that the new data are compatible with the old data, but it also revealed a possible discrepancy in the coordinates of some stations in the Ethiopian database in NBD-Dec02-2. As shown in (Table 2.2.3), the coordinates of the Jima station in NBD-Dec02-2 are slightly different from those of the same station in the NBD database and from the coordinates of the city of Jima in the Rand-McNally World Atlas. The minutes in the station coordinates of the two other sources are equal to the decimal part of the station coordinates (in decimal degrees) of the NBD-Dec02-2 database, pointing to a possible conversion error. The fact that, only seven out of 68 stations in the NBD-Dec02-2 show a decimal part of latitude or longitude higher than 0.60 may be a confirmation of this error. Up to now, it has not been possible to verify the correctness of the other station locations in the NBD-Dec02-2 Ethiopian data.

Table 2.2.2: Precipitation data added to the NBD-Dec02-2 Ethiopian database.

Country	HMStationID	HMName	Lon	Lat	Period
Ethiopia[1]	50069	Bahir Dar	37.420	11.600	1993-1997
Ethiopia[1]	50070	Debre Markos	37.670	10.330	1993-1999
Ethiopia[1]	50071	Gonder	37.420	12.550	1993-1999
Ethiopia[1]	50072	Mehal Meda	37.430	10.250	1993-1999
Ethiopia[1]	50073	Nekemte	36.450	9.080	1993-1999
Ethiopia[1]	50074	A.A.Bole	38.750	9.030	1993-1999
Ethiopia[1]	50075	Awassa	38.480	7.080	1993-1999
Ethiopia[1]	50076	Bekoji	39.300	8.430	1993-1999
Ethiopia[1]	50077	Debre Zeit	38.950	8.730	1993-1999
Ethiopia[1]	50078	Dere Dawa	41.850	9.600	1993-1999
Ethiopia[1]	50079	Desse	39.630	11.140	1993-1999
Ethiopia[1]	50080	Haik	38.930	8.750	1993-1999
Ethiopia[1]	50081	Kibre Mengist	38.970	5.870	1993-1999
Ethiopia[1]	50082	Kombolcha	39.730	11.120	1993-1999
Ethiopia[1]	50083	Mekele	39.480	13.500	1993-1999
Ethiopia[1]	50084	Methara	39.900	8.870	1993-1999
Ethiopia[1]	50085	Negelle	39.570	5.330	1993-1999
Ethiopia[1]	50086	Shola Gebeya	39.330	9.050	1993-1999
Ethiopia[1]	50087	Ziquala	38.700	8.530	1993-1999
Ethiopia[1]	50069	Bahir Dar	37.420	11.600	1993-1999

1 WMO

Table 2.2.3: Station coordinates from different sources.

Source	Place	Longitude	Latitude
NBD-Dec02-2	Jima	36.50	7.40
NFC	Jimma	36°50’	7°40’
Rand-McNally World Atlas	Jima, Kefa	36°50’	7°40’

Data in the augmented NBD-Dec02-2 database have been inspected to eliminate as much as possible erroneous values both in the station coordinates and in the precipitation records.

Verification of Rain Gage Coordinates Against National Boundaries

The rain gage coordinates reported in the *HMStations* table of each *HM<Country>z.mdb* database have been matched against the border of the corresponding country as reported in the GIS maps. Since errors in the GIS layer may also be possible, only stations located more than one kilometer outside the borders of their country have been flagged as problematic and not used in the calibration procedures. Table 2.2.4 lists the stations considered as misplaced and the reasons that led to this evaluation.

Table 2.2.4: Stations not used in the CCD calibration due to incompatibility with country borders.

Country	HMStationID	HMName	Lon	Lat	Comments
Ethiopia	50026	Fejij	36.250	4.400	In Kenya, 6 km from the border.
Kenya	60089	KITALE NOIGAM ESTATE	34.200	0.983	In Uganda, 30 km from the border.
Kenya	60201	CHEPTAIS FOREST STATION	34.333	0.850	In Uganda, 9 km from the border
Kenya	60129	BUSIA FARMERS TRAINING CENTRE	34.100	0.467	In Uganda, 3 km from the border
Kenya	60195	BUSIA CATHOLIC MISSION	34.117	0.467	In Uganda, 1.2 km from the border
Kenya	60372	UOZI DISPENSARY MFANGANO	33.467	-0.467	In Uganda portion of Lake Victoria: 50 km from the border (water)
Kenya	60622	LOLTIANI MAU SUMMIT			Coordinates missing
Rwanda	70017	COLLEGE ST ANDRE	29.067	-2.950	In DRC, 23 km from the border. Likely latitude -1.95
Rwanda	70149	CYANGUGU	28.883	-2.367	In DRC, 3 km from the border
Rwanda	70154	NYAMASHEKE	29.083	-2.033	In DRC, 7 km from the border
Rwanda	70173	PFUNDA II			Coordinates missing
Rwanda	70175	RWAMATAMU	0.000	0.000	Coordinates clearly in error

Rwanda	70176	RWAMBURA	0.000	0.000	Coordinates clearly in error
Rwanda	70177	RWERERE MARAIS	0.000	0.000	Coordinates clearly in error
Rwanda	70179	SHAGASHA GISUMA	29.000	2.383	In DRC, 420 km from the border. Likely lat. –2.383
Rwanda	70180	RWINYANA PR SCHOOL	0.000	0.000	Coordinates clearly in error
Rwanda	70181	KIBUNGO HYDR PROJ	0.000	0.000	Coordinates clearly in error
Sudan	80273	GAMBELA	34.583	8.250	In Ethiopia, 65 km from the border
Sudan	80296	FARAJOK	32.583	3.633	In Uganda, 15 km from the border
Sudan	80315				Coordinates missing
Sudan	80327				Coordinates missing
Sudan	80328				Coordinates missing
Sudan	80329				Coordinates missing
Sudan	80330				Coordinates missing
Sudan	80331				Coordinates missing
Sudan	80334				Coordinates missing
Sudan	80335				Coordinates missing
Sudan	80336				Coordinates missing
Tanzania	90157	KILULA MISSION	30.000	-2.867	In Burundi, 46 km from the border
Uganda	1000598	Arua	0.000	0.000	Coordinates clearly in error
Uganda	1000608	Entebbe	0.000	0.000	Coordinates clearly in error

Two stations, physically located in Uganda, are included both in the Tanzania database (90001 and 90002) and in the Uganda database (1000593 and 1000595). One station, physically located in Rwanda, is listed both in the Rwanda database (60131) and in the Uganda database (1000590). These three stations have not been considered misplaced, but they are assumed to be used by both countries in trans-boundary basins.

Verification of Rain Gage Coordinates Against WMO or National Identification Codes

World Meteorological Organization station identification codes contain information on the station location. More precisely:

- The code's first three characters indicate the station's longitude at the degree level. The number represented by these two characters is set to zero for latitudes between 89 and 90° N and increases by 1 for each degree southward (e.g., a station at latitude 1.15° N has the number 88 as the first three characters of its WMO code, while a station located at –2.72° N features the number 92.)
- The fourth and fifth character indicate a station's longitude at the degree level (e.g., the WMO code of a station located at 31.05° E has as fourth and fifth digits the numbers 3 and 1.)

Table 2.2.5 lists the stations that have a WMOCode field not matching the station's latitude and longitude and, consequently, have not been used in the CCD calibration process. Unfortunately, only the databases from Kenya, Tanzania, and Uganda reported the station WMO code.

In particular, Ethiopia has a national code system for meteorological stations based on a similar principle: the first two digits of the Ethiopian code indicate the station longitude, while the third and fourth digits indicate the station latitude.

Table 2.2.5: Rain gages with the WMOCode not matching their coordinates.

Country	HMStationID	WMOCode	HMName	Lon	Lat
Tanzania	90091	9233007	IKWATA	32.500	-2.917
Tanzania	90114	9330005	Kibondo District Office	35.350	-2.960
Tanzania	90115	9330007	Kibondo Maji Depot	35.360	-2.250
Tanzania	90157	9234003	KILULA MISSION	30.000	-2.867
Uganda	1000478	90290030	Kayonza	30.150	-0.850

Among the stations for which the *HMEthioz.mdb* database reports the National ID code, two show a mismatch between *HMNationalID* field and geographical coordinates that suggests their exclusion.

Table 2.2.6: Rain gages with the *HMNationalID* not matching their coordinates.

Country	HMStationID	HMNationalID	HMName	Lon	Lat
Ethiopia	50003	36070043	Asendabo	37.140	7.460
Ethiopia	50024	37080114	Dimtu	37.150	7.200

Removal of duplicate stations

As it was pointed out previously, the same station may be labeled by two or more different *HMStationID* codes. As a consequence, unnecessarily greater weight may be given to such stations when the spatial distribution of a variable is computed for the region. It is therefore recommended to use only one instance of the station records for further computations. The following table shows the rain gage instances that have been kept and those that have been eliminated where duplicate labels exist.

The instances eliminated include a smaller set of measurements than those kept.

Table 2.2.7: Removed and kept instances of the same rain gage.

***HMStationID* instances kept**					
Country	HMStationID	OtherCode	HMName	Lon	Lat
Sudan	80142	07JBN779	JEBEL EIN	32.800	12.583
Uganda	1000590	91300040	Mirama Hill Customs	30.450	-1.050
Uganda	1000593	91300090	Kamuganguzi	30.000	-1.050
Uganda	1000595	91300110	Rubale	30.167	-1.017

***HMStationID* instances eliminated**					
Country	HMStationID	OtherCode	HMName	Lon	Lat
Sudan	80141	07JBE775	JEBEL EIN	32.800	12.583
Rwanda	70131	45035600	KAGITUMBA	30.450	-1.050
Tanzania	90001	9130009	Kamuganguzi (in Uganda)	30.000	-1.050
Tanzania	90002	9130011	Lubale (in Uganda)	30.167	-1.017

Two stations (70097 and 70102), although located at different sites, include exactly the same precipitation records for the entire period 1970-1982. Only one of these two sets of records is valid, but it has not been possible to identify which one. Therefore, both records have been eliminated from further analysis.

Country	HMStationID	OtherCode	HMName	Lon	Lat
Rwanda	70097	35067500	GATUMBA	29.633	-1.900
Rwanda	70102	40132500	BULERA-LAC	29.767	-1.383

Elimination of duplicate and ambiguous precipitation records

There are instances where for a given date, time, and station there are two or more records reporting the same measurement (in this case the same precipitation value). These are called duplicate records. Only one instance of these records must be kept because the additional copies could introduce errors when counting the number of available records or when computing monthly averages and other statistics. Duplicate records are often created during the data input process.

The databases of the series *HM<Country>0.mdb* (NDB-Jun02) contained 1220 duplicate records. The focal point institutions and FAO eliminated most of the duplicate records in the *HM<Country>Z.mdb* database since only 188 duplicate records have been found. Duplicate records still exist because one instance of the duplicate pair contains precipitation values rounded to the first decimal place, while the other has not been rounded and contains several decimal digits (probably as a result of converting inches to millimeters).

On other occasions, for a given station, date, and time, there exist two or more records reporting different measurements (in this case different precipitation values). These are called ambiguous records and can cause more harm than duplicate records, because they may result in incorrect values being used. Unfortunately, to correct these errors, the history of both records should be revisited, a task that has to be performed by the national meteorological agency. It was thus decided to eliminate all instances of ambiguous records and notify the project of these errors for further investigation. A total of 845 ambiguous records were found in the *HM<Country>Z.mdb* databases.

Comparison of monthly precipitation with 15-year averages

Ideally, the stations to be used for calibration and verification of the remote sensing component should hold precipitation records for at least 90% of the days of every month in the calibration and verification periods. Unfortunately, this criterion would have yielded a very limited number

of rain gages, mostly concentrated around Lake Victoria. To enlarge the set of stations for calibration while maintaining a certain degree of consistency in the calibration and verification process, it was decided to use all of the 1996-1998 monthly datasets that have at least 90% of their records, regardless of the individual station monthly datasets. In evaluating the performances of the remote sensing component in specific areas, however, the user must pay attention to the temporal extension of the data used for calibration in those specific areas, since it may be incomplete.

Unfortunately, the *HMRwandz.mbd* database reports only the "rainy" records for the period 1995-1999. The "no rain" records are not reported and, consequently, the 90% level of daily records in a month is never reached. Given the frequency of missing records in the region, it could not be assumed that all missing records are "no rain" records.

Ideally, daily records should be checked for eliminating errors due to incorrect measurements, faulty instrumentation, and wrong transcriptions. This task would be daunting even in dense rain gage networks because of the inherent randomness of the precipitation events and the variety of the errors affecting rainfall measurements and their transcription in databases. The task becomes prohibitive if the network of rain gages is as sparse as it is in most areas of the Nile River Basin. Monthly rainfall averages, on the other hand, have a stricter relation with the climatic characteristics of the period and location. Therefore, comparing the average precipitation during a month against the historical average for the same month and station or against the same month average in neighboring stations allows to screen out major errors with a certain degree of reliability. Such errors include damaged instrumentation, precipitation records from one station wrongly attributed to another station for extensive periods, important shifts in the sequence of records for one station, and incorrect translation from one measurement system to another.

For each station (j), month (i), and year (Y), precipitation records are flagged suitable for further use if they satisfy the following criterion:

$$\mu R15_{ij} - 3*\sigma R15_{ij} < \mu RY_{ij} < \mu R15_{ij} + 3*\sigma R15_{ij}$$

where
μRY_{ij} : average precipitation at station *j* during month *i* in year Y;
$\mu R15_{ij}$: average precipitation at station *j* during month *i* in the 15 years preceding year Y;
$\sigma R15_{ij}$: standard deviation of the average precipitation at station *j* during month *i* in the 15 years preceding year Y.

If the criterion is not satisfied, the particular monthly dataset is set aside for manual examination. It was decided to use only 15 years to characterize typical monthly precipitation patterns to minimize uncertainties introduced by changes in station instrumentation and environmental conditions and by possible climatic shifts. It is noted that both the average monthly precipitation and its standard deviation during the 15 years preceding year Y are computed using only monthly datasets including at least 90% of all possible records. If the number of such datasets is below five, both the average precipitation estimate and its standard deviation are considered unreliable and the station is set aside for further examination.

Of the 9244 monthly datasets that passed the 90% density test in the period 1996-1998, only 7612 also passed the comparison with the historical monthly average rainfall for the same station. The 1632 datasets that did not pass the test were submitted to visual comparison with contemporaneous datasets from neighboring stations. If the average rainfall during a suspicious month were also above (or below) the normal range in the neighboring stations, the monthly dataset was accepted even if it did not meet the comparison with historical data from the same station. Several one-month datasets that did not meet the comparison with the normal average rainfall were accepted after comparison with neighboring stations. A likely reason of this is that 1998 is an El Niño/La Niña year, with exceptionally strong *long rains* and weak or absent *short rains*.

The general approach to the visual comparison was to place trust to the individuals that compiled the *HM<Country>Z.mdb* as much as possible. Only after no explanation could be found for failing the normal range test, a monthly dataset was dismissed. Thus, only 289 monthly datasets were actually dismissed because they appeared to be in error or because there was no neighboring station available for comparison. A few stations were dismissed completely (all records) for the reasons reported in Table 2.2.8.

Table 2.2.8: List of completely dismissed stations.

Country	HMStationID	HMName	Lon	Lat	Comments
Sudan	80318	KOBER POLICE	32.550	18.230	Records more similar to a 15N station than to its 18N neighbors.
Tanzania	90087	KAFUNZO	32.730	-2.720	Records too different from neighbors 90082, 90161, 90089, 90110
Uganda	1000525	Ruhengeri FS	30.750	-0.430	Records identical to station 1000524 records.

The set of rain-gage records contemporaneous to the available Meteosat imagery is shown in Figure 2.5. The precipitation records have a good density in the Lake Victoria basin, with the exception of Rwanda and the eastern part of the basin. Tanzania has a very good rain gage density, but records are available only for 1996-1997 only. Southern Sudan and northern Uganda have a very low gage density. Central Sudan shows a good coverage, albeit somewhat sparse in the western part. Northern Sudan and most of Egypt do not include many rain gages, but precipitation is sporadic. Central Ethiopia has good precipitation record density, but coverage of the western and northern regions is rather limited.

It is finally noted that other precipitation stations and records exist in the national databases; however, due to time limitations, it was not possible to include them in the Nile DST database.

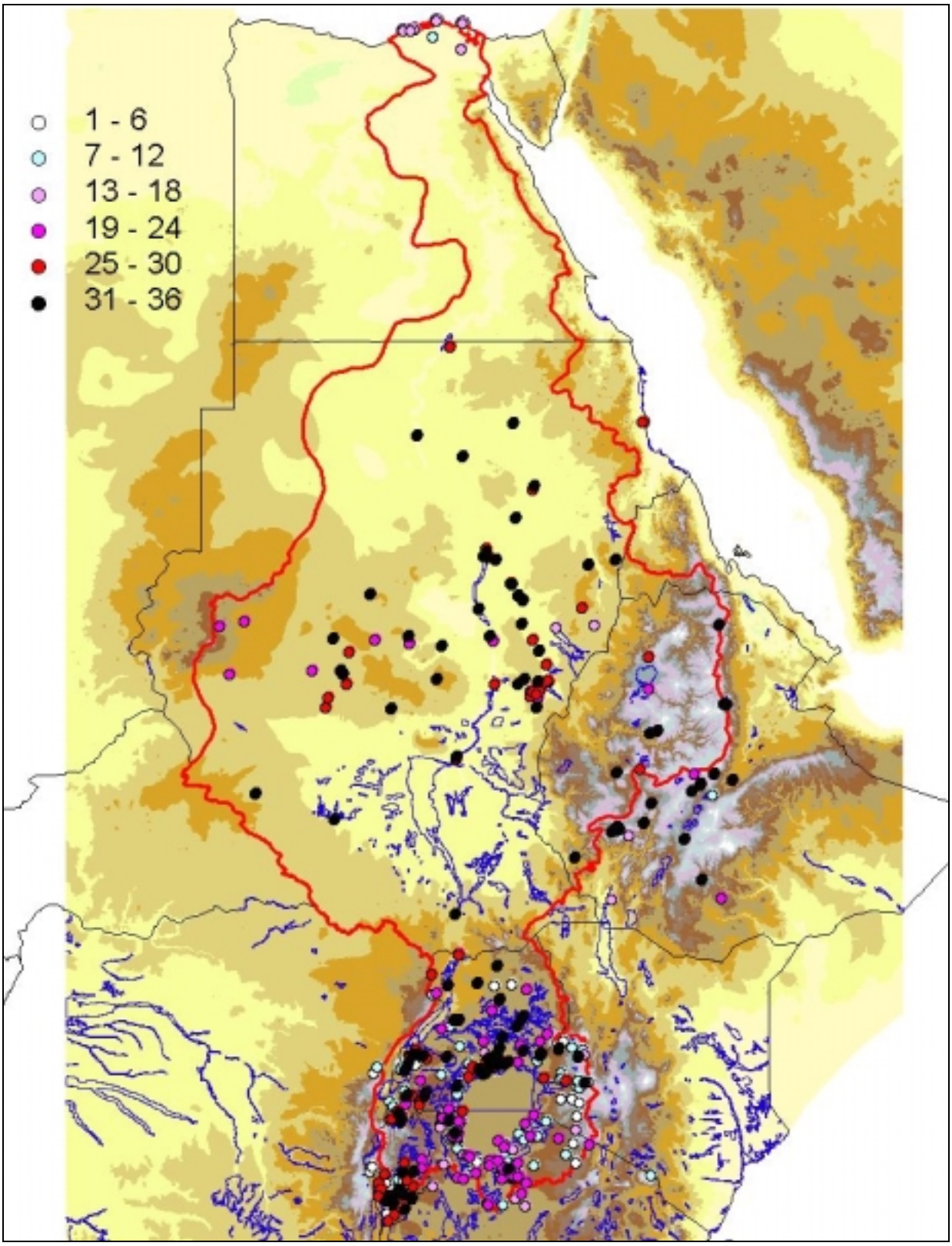


Figure 2.5: Months of rain gage records available for model calibration/verification in the period 1996-1998.

3. Model Calibration and Verification Procedure

The CCD module being implemented is calibrated for estimating the mean precipitation at a 10-day (hereafter referred to as decade) temporal resolution and 1°x1° spatial resolution. The choice of the 10-day temporal resolution is consistent with the time resolution of the hydrologic and the agricultural Nile DST modules. Furthermore, CCD estimates are most reliable at coarse spatial and temporal scales.

It is widely recognized that the optimal CCD parameters vary in space and time because the relation between cloud characteristics and generated precipitation is influenced by a variety of factors (e.g., amount of water vapor in the air and orography of the region) that have definite seasonal and spatial dynamics. A CCD module using the same set of parameters over wide areas and for the entire year is consequently less precise than a CCD module using different set of parameters for specific regions and seasons.

To capture the spatial variability of the climate, a different set of parameters has been obtained for each 1°x1° square that has a sufficient number of rain gages. This is a compromise between the reliability of the CCD model and the capacity to represent weather patterns that can substantially vary within few tens of kilometers (for example in the Lake Victoria basin or in the Ethiopian plateau).

As shown in Chapter 2, ,however, the availability of data was not uniform in space and in time. In many regions, the record density was insufficient to obtain seasonal sets of CCD coefficients. In other regions, such as most of Lake Victoria, 1996-1997 satellite images were availed to the project only recently and that prohibited their use in calibration. It is, however, possible that seasonal sets of CCD parameters could be developed in the coming months when Eumetsat delivers the Meteosat images for replacing missing slots.

An annual parameter set can be generated by calibrating the parameters using data from one or more years and verifying their validity by applying the model to one or more different years. This approach, however, suffers from two drawbacks:

1. Precipitation patterns exhibit strong inter-annual variability that discourages model calibration on a limited set of one or two years. A model calibrated during an El Niño event would not perform well during an average year and vice versa.
2. Satellite sensor performance varies from one member to another in the same satellite family. This introduces an additional source of noise in model calibration. Thus, it is better to include samples from several satellites in the calibration set.

On the other hand, the sparseness of the rain gage network and the heterogeneity of the precipitation patterns in different regions of the basin prevent partitioning the data set into calibration and verification sets.

To overcome this impasse, the first ten days of each month were used as the calibration set, while the remaining part of each month was used as the verification set. It was thus possible to

capture the variation of rainfall patterns during several years, while providing an adequate number of samples for calibration and verification. This is especially true for those squares containing several meteorological stations.

A preliminary study of the CCD calibration, carried out in 2002 (De Marchi and Georgakakos, 2003), revealed that rain gage density in the Lake Victoria basin allowed the development of a different set of CCD coefficients for each 1°x1° square. In other parts of the basin, however, the precipitation record density was too low to yield reliable sets of parameters at this spatial resolution. Thus, the present calibration was carried out over areas of different size, trying to strike a balance between the available data and the variation in rainfall patterns within the areas. The following chapters give an account of how each area was selected.

For each calibration area, the ideal set of parameters is obtained by minimizing the sum of the mean square difference between the precipitation recorded on the ground at each rain gage and the precipitation generated by the CCD method for the pixel containing the rain gage. Namely, by minimizing:

$$\Sigma_s \Sigma_p [G_{sp} - S_{sp}]^2,$$

where

G_{sp}: total rainfall recorded at rain gage *s* during the 10-day period (decade) *p* of the calibration set;

S_{sp} : total rainfall estimated by the CCD method for the pixel above rain gage *s* during the decade *p* of the calibration set.

Calibration results (i.e., the ability of a model to fit experimental data) can be very good, but they provide no warrantee that the model is actually able to estimate precipitation under conditions other than those used for calibration, especially if the calibration data is sparse. The best way to establish model skill is to apply the model to an independent data set and evaluate how well model results agree with ground data.

The following model performance measures are adopted:

3.1 Bias (BIAS)

The bias is given by the ratio:

$$|\Sigma_s\Sigma_p G_{sp} - \Sigma_s\Sigma_p S_{sp}| / \Sigma_s\Sigma_p G_{sp} ,$$

where

G_{sp} : total rainfall recorded at rain gage *s* during the decade *p* of the verification set;

S_{sp} : total rainfall estimated by the CCD method for the pixel above rain gage *s* during the decade *p* of the verification set.

Bias is a measure of how well the CCD model reproduces the average annual precipitation within the calibration area. It is noted that the absolute value of the difference between average

ground measured rainfall and satellite estimated rainfall does not indicate if the station is overestimating or underestimating. Furthermore, the absolute value avoids "hiding" cases where underestimating in some periods is compensated by overestimating in others. If the model were a perfect predictor, the bias would be zero. However, it must be also remembered that BIAS, as all relative measures, may be somewhat misleading when precipitation is low. For example, if the average rain gage precipitation in one period is 0.2 mm and the satellite estimation is 0.4 mm, the BIAS is 1.0, but it is clear that the practical effect of this error would be very low.

3.2 Correlation (CORR)

The correlation coefficient indicates how well the CCD model is able to reproduce the variation of the average rainfall in the calibration area during the year. Ideally, CORR would be close to one. A value near 0 implies that satellite estimates and ground measurements are completely uncorrelated. Strong correlation exists if CORR is above 0.6 - 0.7. Note however, that if satellite estimates were exactly 3 times the ground measurements for each decade, the correlation between the two would be equal to one.

3.3 Mean Absolute Error (MAE)

The mean absolute error is given by:

$$\Sigma_p \mid \Sigma_s G_{sp} - \Sigma_s S_{sp} \mid / \Sigma_s \Sigma_p G_{sp}$$

MAE is a measure of the average absolute difference between the average rain gage measured precipitation and the average satellite estimated precipitation for all decades available for verification. It is given as a fraction of the mean ground measured average precipitation during the verification period so that it may be used to compare verification results from squares featuring very different weather patterns and to compute regional averages. Ideally, MAE should be close to zero.

3.4 CCD Model Description

The Nile DST area has been subdivided into several sub-areas:

- The area 2°S to 4°N and 29°E to 36°E, corresponding to the Lake Victoria plateau, where the climate is rather heterogeneous and the ground monitoring network rather dense;
- The area 4°N to 10°N and 25°E to 38°E, corresponding to the part of the basin for which satellite images are available for the entire 1998;
- The area 10°N to 32°N and 24°E to 37°E, corresponding to the part of the basin for which satellite images are available only for the period January – July 1998.

In De Marchi and Georgakakos (2003) five CCD model types were explored for 1998. These models included:

Original GPI model
In this model, originally developed by Richards and Arkin in 1981, the presence of clouds generating rain over a pixel is identified by the presence of thermal infrared temperature below 235°K. The ensuing precipitation is generated at a rate of 3 mm/h.

Unconstrained CCD Calibration Based on Infra Red (IR) Radiation
In this model, clouds generating rainfall are identified as those having thermal infrared temperature below a threshold varying from area to area. The IR threshold, as well as the rate with which rain is produced, are determined by the mean square error minimization explained previously.

Constrained CCD Calibration Based on Infra Red (IR) Radiation
The only difference with the previous model is that the rain-producing IR thresholds are forced to vary over a reduced range of values. This limitation stems from the recognition that mathematical minimization, especially if based on small samples, can lead to parameters that are meteorologically implausible.

Unconstrained CCD Calibration Based on Infra Red (IR) and Visible (VIS) Radiation
The premise of the CCD model is that rainy clouds are characterized by cold top-of-the-cloud temperature. This is due to the fact that precipitation is most often generated by lifting of moist air from low levels to high altitude, either by convection or by convergence. During lifting, air parcels undergo rapid cooling that causes moisture to precipitate and the temperature of the cloud top to drop.

Normally, rainy clouds are also thick and highly reflective of the incoming solar visible radiation. High albedo values are therefore good indicators of rainy activity during daylight. In this model, daytime rainy clouds are recognized by the contemporaneous presence of IR values below an IR threshold and albedo values above a visible threshold. During nighttime, rainy clouds are identified as those having thermal infrared temperature below a threshold, as in the previous methods. The daytime and nighttime IR thresholds may differ from each other. These three thresholds, as well as the precipitation rate, presumed to be the same for day and night, are determined by the mean square error minimization explained previously.

Constrained CCD Calibration Based on Infra Red (IR) and Visible (VIS) Radiation
In this approach, mean square error minimization is performed within a reduced IR and VIS threshold range. De Marchi and Georgakakos (2003) showed that the original GPI model was useful for delineating the precipitation patterns over a region, as an aid in identifying the areas for which different sets of CCD parameters are going to be developed, and as a reference to evaluate calibration performance. They also showed that unconstrained calibration was generally not producing good results. The presence of constraints on the value of the IR and VIS thresholds gave better verification results and more homogeneous sets of parameters. This tendency extended throughout the basin even during this calibration/verification exercise. Therefore, it was decided to show and discuss only the results of the constrained IR only and IR + VIS models. In the IR only CCD model, the IR threshold was generally allowed to vary between 223 and 263 °K, with some exceptions, where it was allowed to reach 273 °K. In the IR + VIS CCD models, the daytime and nighttime thresholds are allowed to vary within the same

interval used for the IR only model. Further, the difference between these two thresholds is constrained to be less than 20 °K. This is to ensure that the type of clouds screened at night are not too different from those screened during daytime. Typically, the IR threshold at night is lower than the daytime threshold. The more permissive daytime criterion is possible because of the additional screening introduced by the daytime VIS threshold.

In De Marchi and Georgakakos (2003), the calibration and verification decades were selected from the same region for which the CCD coefficients are derived. That study, however, showed that in several areas the calibration square contained few stations, while additional stations falling just across the calibration area borders could not contribute to the calibration. For this reason, the calibration/verification of the CCD over a given area is accomplished using all of the available precipitation decades from that area and from stations within 0.5 degrees in latitude and longitude from the area borders. This "enlargement" of the calibration areas should increase the likelihood that the coefficients for one area incorporate the regional climate variation and should result in more homogeneous CCD coefficients.

4. CCD Model Calibration in the [4°S, 2°N] and [29°E, 36°E] Region

4.1 Precipitation Records Available for Calibration

As it is shown in Figure 4.1, data available for calibration/verification in Lake Victoria basin are abundant in Burundi, Uganda, and most of Tanzania, but scarce in the eastern part of the basin, and absent in Rwanda. Unfortunately, Tanzania databases stop at the end of 1997, while Rwanda stations report only rainy day values, without separating actually missing values from no-rain records.

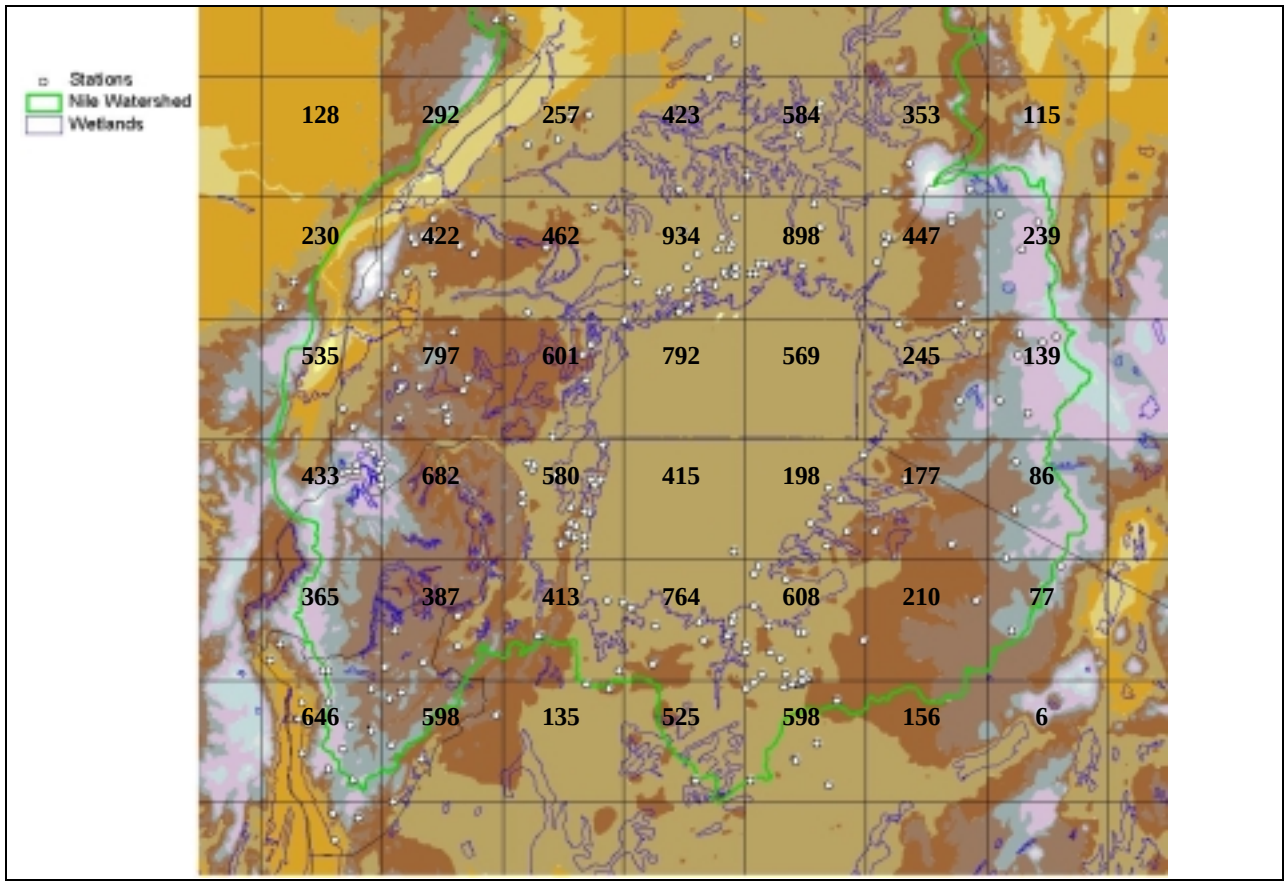


Figure 4.1: Spatial subdivision of the [4°S, 2°N] and [29°E, 36°E] region into 1°x1° squares and distribution of stations reporting at least one month of valid data in 1996-1998. The numbers represent the decades available for calibration in 1996-1998 within the enlarged 1°x1° squares. Periods available for verification are twice those available for calibration.

4.2 Original GPI Model

The original GPI model shows relatively good correlation with ground data when it is applied to the [4°S, 2°N] and [29°E, 36°E] region (Figure 4.2). Only over the western part of the basin, the correlation is low.

Table 4.2.1: Verification Statistics for the GPI Model; [4°S, 2°N] and [29°E, 36°E] region.

Calibration 10-day periods	Average BIAS	Average CORR	Average MAE	Median BIAS	Median CORR	Median MAE
>=10	0.850	0.745	0.960	0.897	0.744	0.940
>=30	0.850	0.745	0.960	0.897	0.744	0.940
>=60	0.850	0.745	0.960	0.897	0.744	0.940
>=120	0.900	0.744	0.987	0.908	0.747	1.005

GPI bias, however, is rather high reaching average and median values between 72 and 86% (Table 4.2.1). The only exception is over the eastern mountains, where overestimation is only 12%.

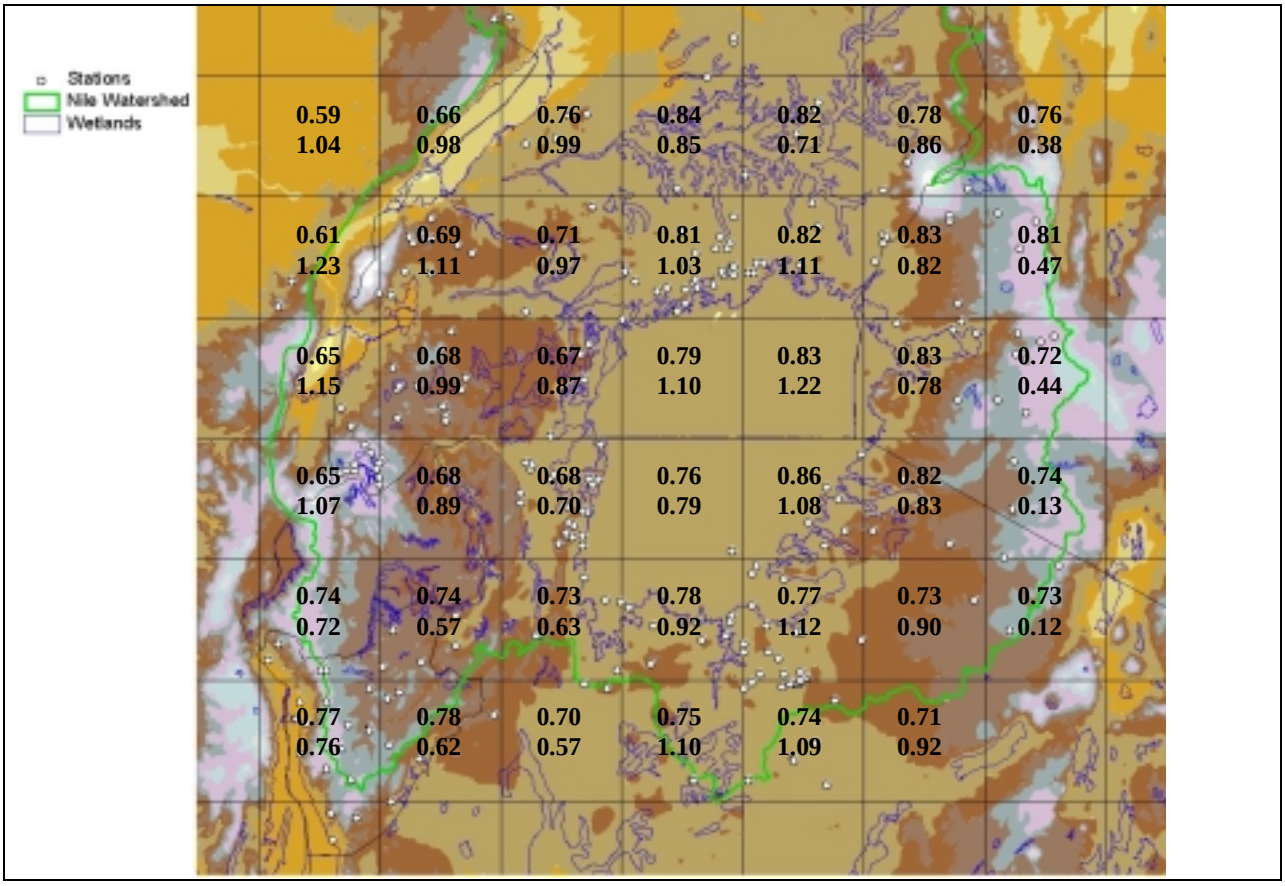


Figure 4.2: Spatial distribution of the GPI algorithm's verification CORR (top row) and BIAS (lower row); [4°S, 2°N] and [29°E, 36°E] region

4.3 Constrained CCD Calibration Based on Infra Red (IR) Radiation

After several trials, the IR thresholds were constrained to vary between 228 and 263°K producing the spatial distribution of the optimal IR thresholds and rain rates reported in Figure 4.3. The basin shows three relatively homogeneous areas: the highland in the western part of the

basin, the Lake and central plateau, and the highland in the eastern part of the basin. High IR thresholds are found over the highlands, possibly due to the fact that orographic rain is often characterized by the absence of significant convective lifting. The optimal thresholds over the western highlands are actually very similar to those over the eastern mountains. The rainfall rate over the latter region, however, is three times that over the western mountains.

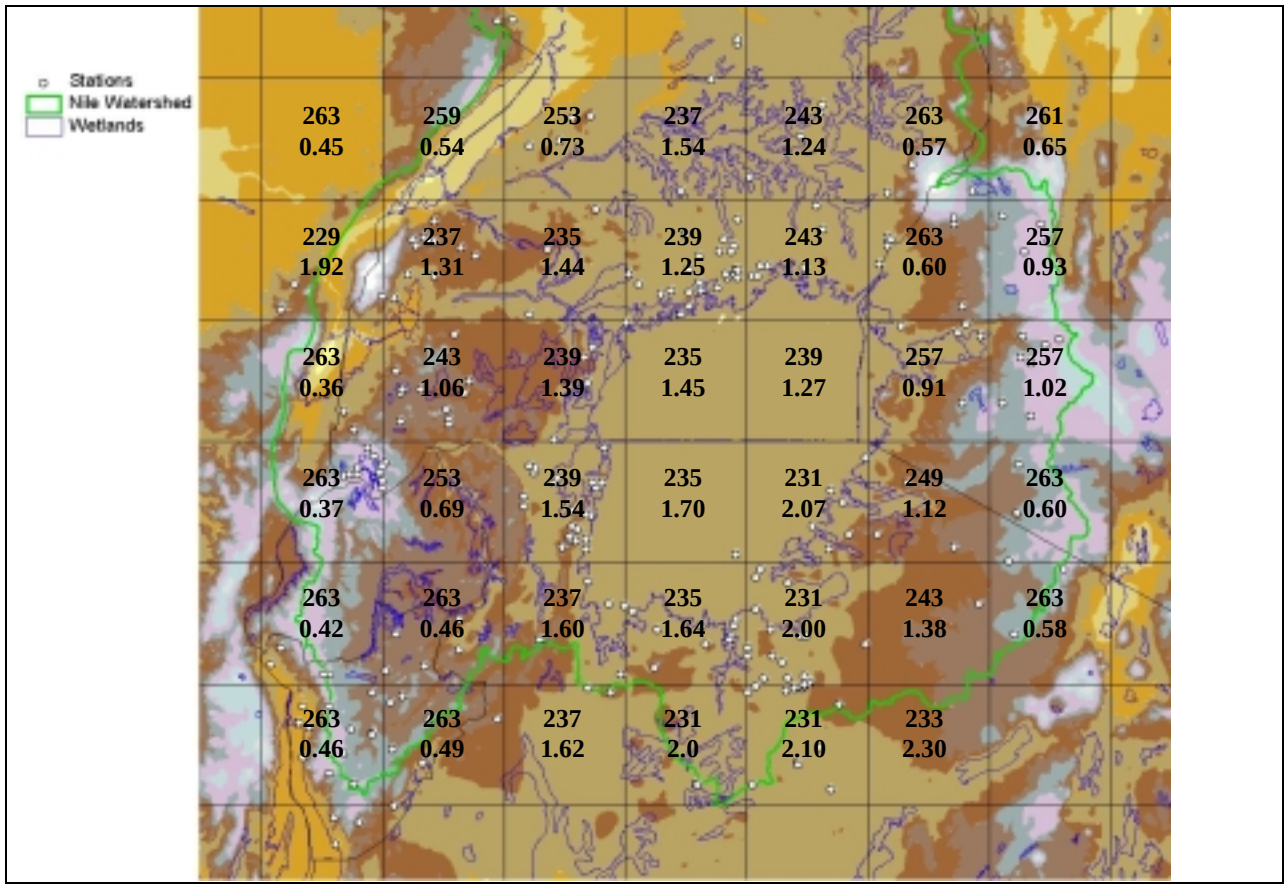


Figure 4.3: Spatial distribution of the constrained CCD/IR model's IR thresholds (°K) and rain rates (mm/h) in the [4°S, 2°N] and [29°E, 36°E] region.

The optimal parameters in the Kisumu area are more believable than those derived using the 1998 data alone over 1°x1° normal squares (De Marchi and Georgakakos, 2003). This is due to the increase in time and space of precipitation data records. On the other hand, there is the risk that they reflect the highland climate more than the shoreline climate due to the station distribution in the enlarged set.

The average BIAS shows a marked improvement over the GPI case, decreasing to less than 9%. The correlation has also slightly improved to 0.76, while the MAE has decreased to 42%.

Table 4.3.1: Verification Statistics for the Constrained CCD/IR Calibration; [4°S, 2°N] and [29°E, 36°E] region

Calibration 10-day periods	Average BIAS	Average CORR	Average MAE	Median BIAS	Median CORR	Median MAE
>=10	0.087	0.760	0.428	0.054	0.768	0.415
>=30	0.087	0.760	0.428	0.054	0.768	0.415
>=60	0.087	0.760	0.428	0.054	0.768	0.415
>=120	0.082	0.758	0.425	0.053	0.759	0.412

On the other hand, the spatial distribution of CORR and BIAS shows that model performance improves with respect to the GPI in all locations. It also shows that correlation is lower in the western side of the basin, while BIAS is higher in the eastern side (Figure 4.4). Only two squares have unsatisfactory correlation or bias, but both of them overlap very little with the Nile Basin. One reason for the poor performance could be the existence of a few rain gages in the squares, providing insufficient calibration data.

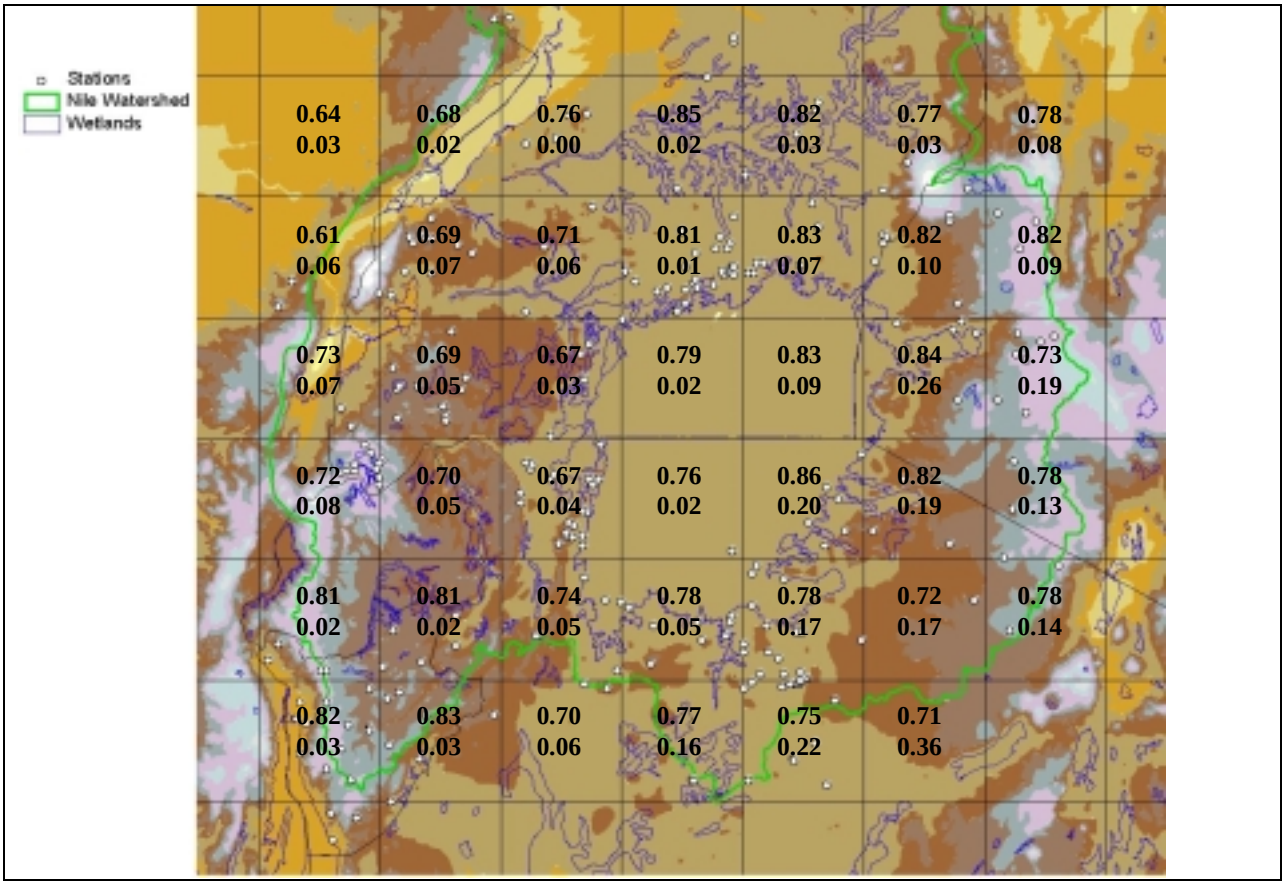


Figure 4.4: Spatial distribution of the constrained CCD/IR model's CORR (top row) and BIAS (lower row) in the [4°S, 2°N] and [29°E, 36°E] region.

4.4 Constrained CCD Calibration Based on Infra Red (IR) and Visible (VIS) Radiation

By imposing a set of constraints on the range of the parameters produced by the calibration process, it is possible to generate more physically plausible values without having to increase the calibration dataset. After several trials, the best results were obtained for the following ranges:

- 228 <= Daytime IR Threshold <= 263
- 228 <= Nighttime IR Threshold <= 263
- |Daytime IR Threshold - Nighttime IR Threshold| ≤ 20
- 20 <= VIS Threshold <= 60

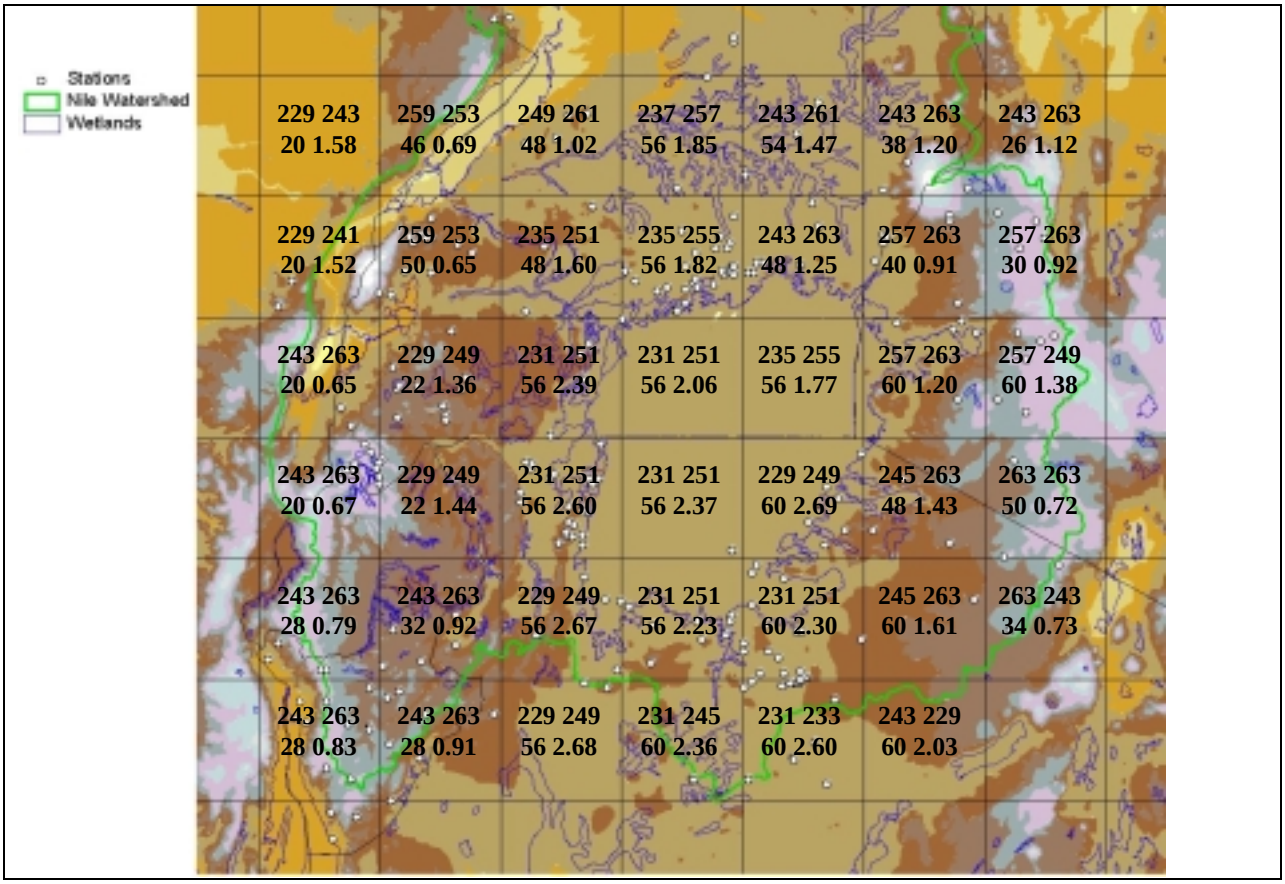


Figure 4.5: Spatial distribution of the constrained CCD IR + VIS model's parameters (top row nighttime IR and daytime IR thresholds (°K); lower row VIS threshold (albedo, %) and rain rate (mm/h)) in the [4°S, 2°N] and [29°E, 36°E] region

The distribution of optimal CCD coefficients (Figure 4.5) has substantially improved in comparison with the results obtained using only the 1998 data (De Marchi and Georgakakos, 2003) since they are now available for all squares and are much more stable. For example the standard deviation of calibrated precipitation rate is now only 0.67 against the 1.2 of 1998. Optimal thresholds in adjacent squares are often very similar, while precipitation rate is a little

more variable. This is the result of enlarging the calibration areas to 2°x2° squares, increasing the calibration period to three years and augmenting the precipitation record pool by adding data from other sources.

As with the IR only case, IR thresholds tend to be higher at higher elevations, both on the East and on the West sides. At the same time, VIS thresholds and rain rates in the former areas are often lower than in the latter. This seems to indicate that while over the lake and the areas surrounding it, storms tend to be of a convective origin, at higher elevations, rainfall is less frequently associated with deep and high clouds. This could be attributed to the occurrence of orographic rain.

A good indication that the calibration process produces credible results is the fact that nighttime IR thresholds are generally lower than daytime IR thresholds. This happens because during daytime VIS values contribute to discriminating raining clouds from non-raining clouds. There are only few instances where this does not hold, but this occurs in mountainous areas with few rain gages.

With the introduction of the visible channel and the adoption of these constraints, model calibration improved producing the best set of results so far (Table 4.4.1 and Figure 4.4.2).

- BIAS is further reduced to a basin wide average of a little less than 8%. Among relevant squares, only the one surrounding Kisumu shows an overestimation of over 20%.
- CORR increases to a basin wide average of over 0.78. The improvement is higher over the lake and the lowlands than over the highlands and only three squares maintain a correlation below 0.7;
- MAE is reduced to 0.39.

Table 4.4.1: Verification statistics for the Constrained CCD/IR-VIS Calibration; [4°S, 2°N] and [29°E, 36°E] region

Calibration 10-day periods	Average BIAS	Average CORR	Average MAE	Median BIAS	Median CORR	Median MAE
>=10	0.078	0.783	0.400	0.058	0.790	0.392
>=30	0.078	0.783	0.400	0.058	0.790	0.392
>=60	0.078	0.783	0.400	0.058	0.790	0.392
>=120	0.075	0.782	0.395	0.058	0.793	0.387

The spatial distribution of optimization parameters is much more uniform than before.

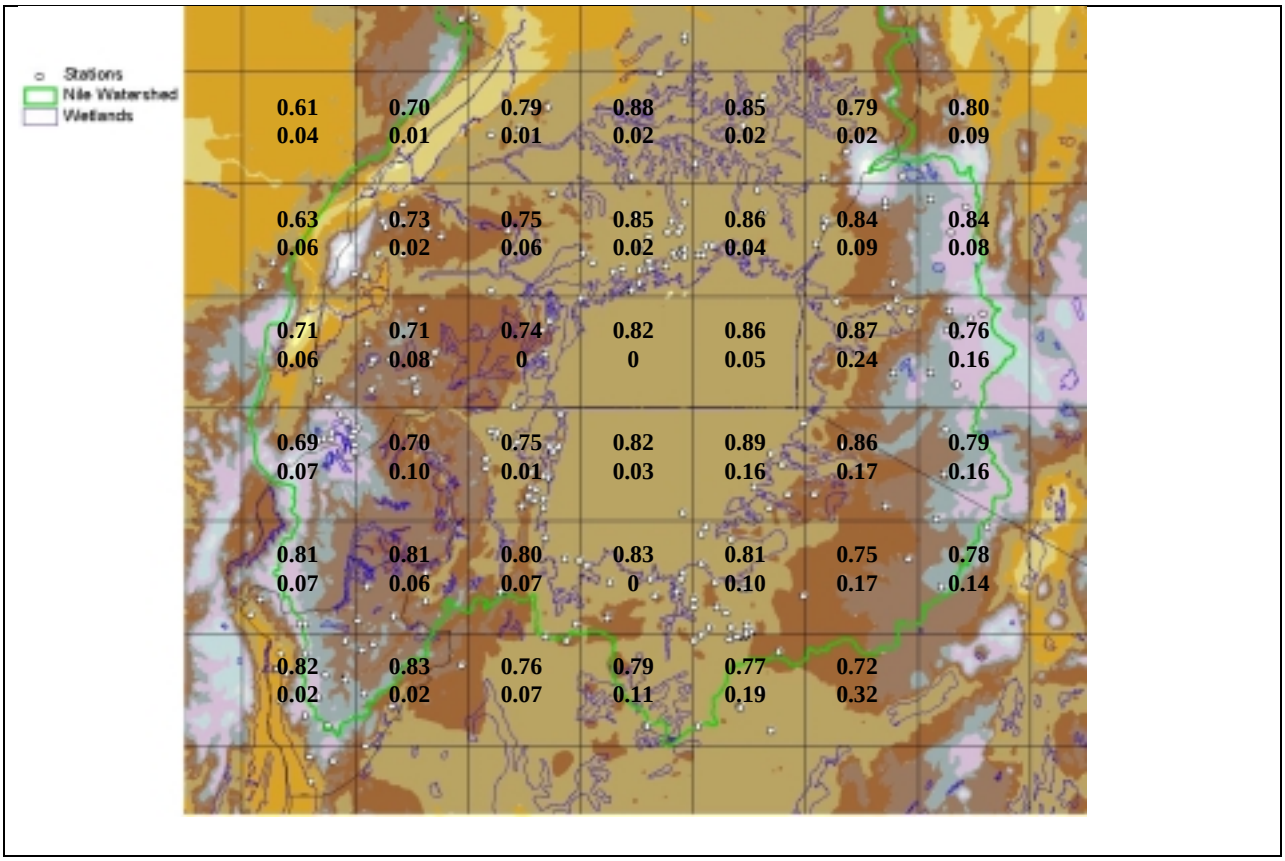


Figure 4.6: Spatial distribution of the constrained CCD IR + VIS model's CORR (top row) and BIAS (bottom row) in the [4°S, 2°N] and [29°E, 36°E] region.

The verification which is based on three years of data, including one El Niño/La Niña event, shows that the model has a very good agreement with ground stations.

5. CCD Model Calibration for the [2°N, 5°N] and [30°E, 34°E] Region

5.1 Precipitation Records Available for Calibration

Given that in northern Uganda (2.0N÷4.0N) and southern Sudan the stations are much sparser than in the Lake Victoria basin and that the climate is somewhat more uniform, calibration has been initially explored for the 2°x2° squares (turquoise and blue lines) depicted in Figure 5.1 using the same procedure as in Lake Victoria (i.e., including the stations within the square and within a 0.5 degree border around it).

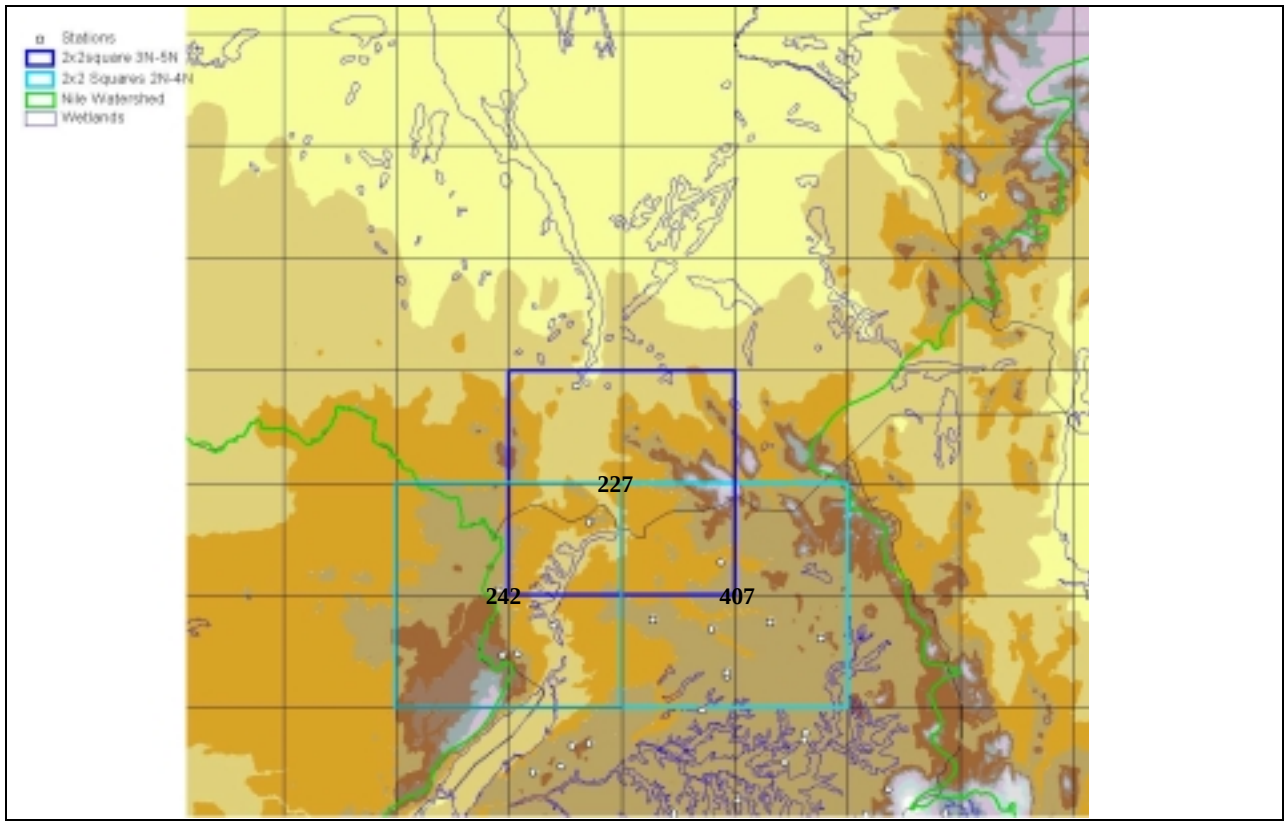


Figure 5.1: Spatial subdivision of the [2°N, 5°N] and [30°E, 34°E] region into 2°x2° squares and distribution of stations reporting at least one month of valid data in 1996-1998. The numbers represent the decades available for calibration in 1996-1998 within the enlarged 2°x2° squares. Periods available for verification are twice those available for calibration.

Calibration results for the [2°N, 4°N] area are extended to the remaining 1°x1° areas eastward. There is only one station between 4°N and 5°N. Stations north of this latitude are few and far apart. It was thus decided to calibrate the model for these latitudes using also the available stations between [3°N, 4°N] and extend the calibration results to the 1°x1° areas within the Nile Basin between 4°N and 5°N.

5.2 Original GPI Model

The spatial distribution of GPI performance results shows that rainfall patterns in the whole area should be relatively similar, thus justifying the use of larger calibration areas. As usual with this model, correlation is good (~0.8), but BIAS is around 100%.

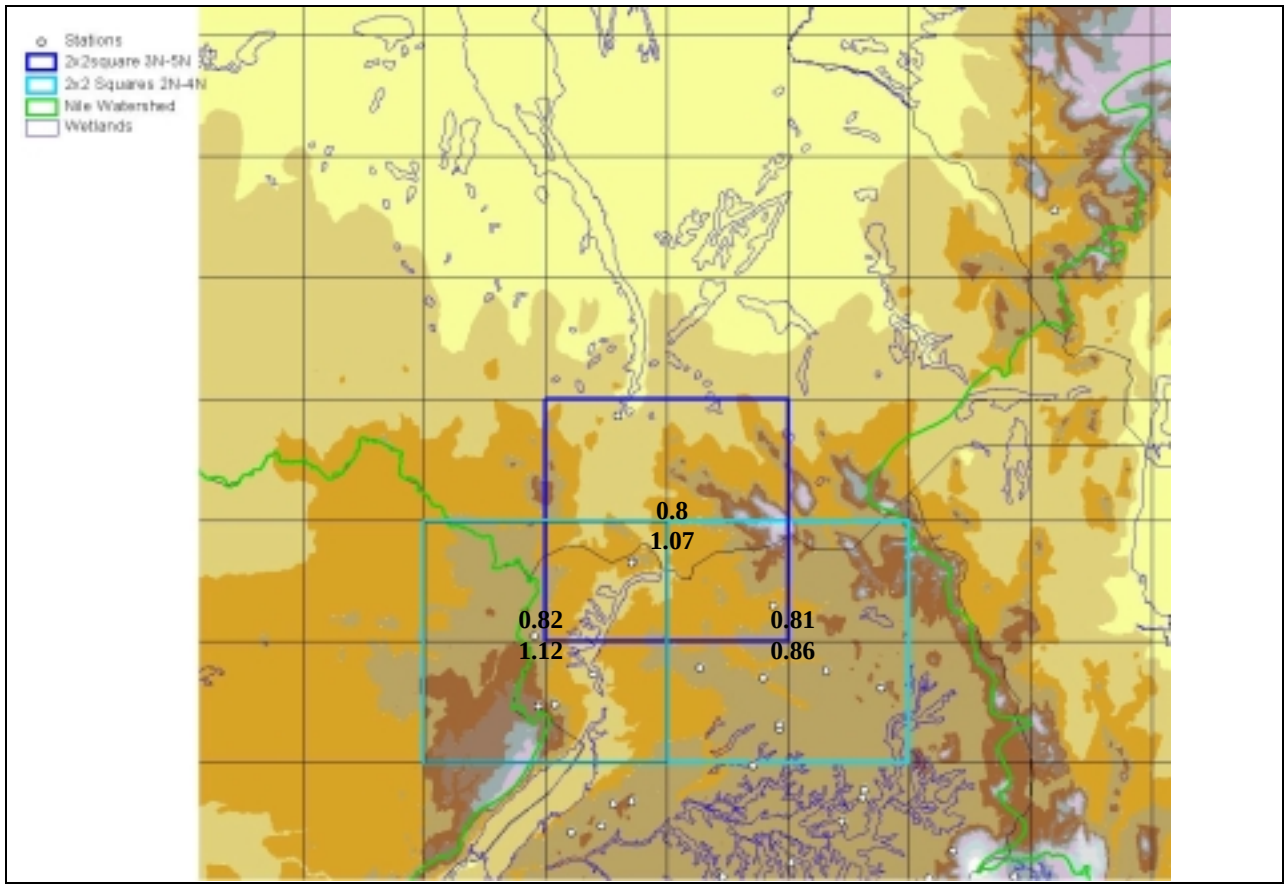


Figure 5.2: Spatial distribution of the GPI algorithm's verification CORR (top row) and BIAS (lower row); [2°N, 5°N] and [30°E, 34°E] region

5.3 Constrained CCD Calibration Based on Infra Red (IR) Radiation

The spatial distribution GPI performance results show a certain degree of uniformity in rainfall patterns. Given the small number of available rain gages in the area it was decided to perform the model calibration over only two areas: the [2°N, 4°n] and [30°E, 34°E] obtained by merging the two lower 2°x2° squares (turquoise squares) and the upper 2°x2° square (blue line). It was not deemed appropriate to unite the three squares into a single entity that would have contained very different rain patterns.

Calibration parameters for IR only show that the two areas had practically identical behavior during the calibration. Of course, this also is due to the fact that only one station in the upper calibration area does not belong to the lower area.

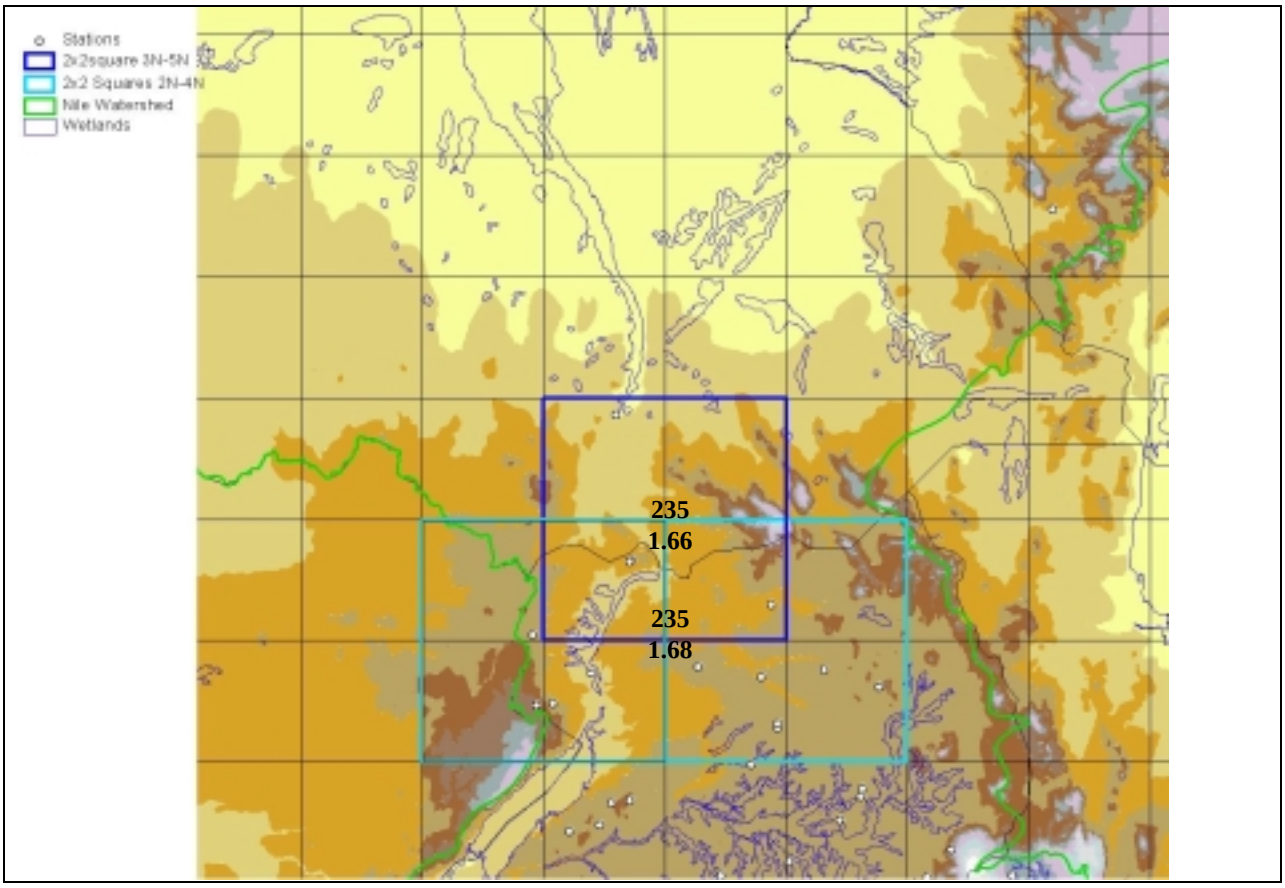


Figure 5.3: Spatial distribution of the constrained CCD/IR model's IR thresholds (°K) and rain rates (mm/h) in the [2°N, 5°N] and [30°E, 34°E] region

Verification results (Table 5.3.1) in both areas show average correlation of more than 0.8 and bias below 15%. On the other hand, the model is overestimating precipitation in the northern area more than it does in the southern area, indicating that it was correct to consider them separately.

Table 5.3.1: Verification Statistics for the Constrained CCD/IR Calibration; [2°N, 5°N] and [30°E, 34°E] region

Latitude	Longitude	Average BIAS	Average CORR	Average MAE
[2°N, 4°N]	[30°E, 34°E]	0.081	0.830	0.332
[3°N, 5°N]	[30°E, 34°E]	0.15	0.80	0.387

5.4 Constrained CCD Calibration Based on Infra Red (IR) and Visible (VIS) Radiation

The CCD model using both IR and VIS data has been calibrated using the same constraints used in Section 4.4. The optimal calibration values reported in Table 5.4.1are relatively different, thus confirming that the northern and southern calibration area had to be separated.

Table 5.4.1: Constrained CCD IR + VIS model's optimal parameters; [2°N, 5°N] and [30°E, 34°E] region

Latitude	Longitude	Nighttime IR Threshold (K)	Daytime IR Threshold (K)	VIS Threshold (Albedo, %)	Rain rate (mm/h)
[2°N, 4°N]	[30°E, 34°E]	237	257	54	1.783
[3°N, 5°N]	[30°E, 34°E]	243	261	56	1.477

Model performances (Table 5.4.2) show a slight improvement over the already good performances of the IR only CCD.

Table 5.4.2: Verification Statistics for the Constrained CCD IR + VIS Calibration; [2°N, 5°N] and [30°E, 34°E] region

Latitude	Longitude	Average BIAS	Average CORR	Average MAE
[2°N, 4°N]	[30°E, 34°E]	0.057	0.853	0.302
[3°N, 5°N]	[30°E, 34°E]	0.089	0.80	0.362

6. CCD Model Calibration for the [5°N, 9°N] and [25°E, 34°E] Region

6.1 Precipitation Records Available for Calibration

The belt between 5.0°N and 9.0°N exhibits very low rain gauge density. This makes calibration and verification of satellite based precipitation more difficult than in the regions analyzed previously. The number of stations is insufficient to compensate for measurement or conversion errors, even if calibration is applied to a very large area. Furthermore, precipitation patterns over large areas may vary significantly, undermining the calibration process for single sub-regions. The western part of this latitude belt features a relatively uniform terrain, thus suppressing the diversity in precipitation patterns introduced by orographic rain. On the other hand, the region is at the boundary between the tropical humid and tropical dry climatic zones, increasing the variability of rainfall patterns.

Only five stations are available for calibration in the western part of the belt. However, all stations include a nearly complete set of data (Figure 6.1).

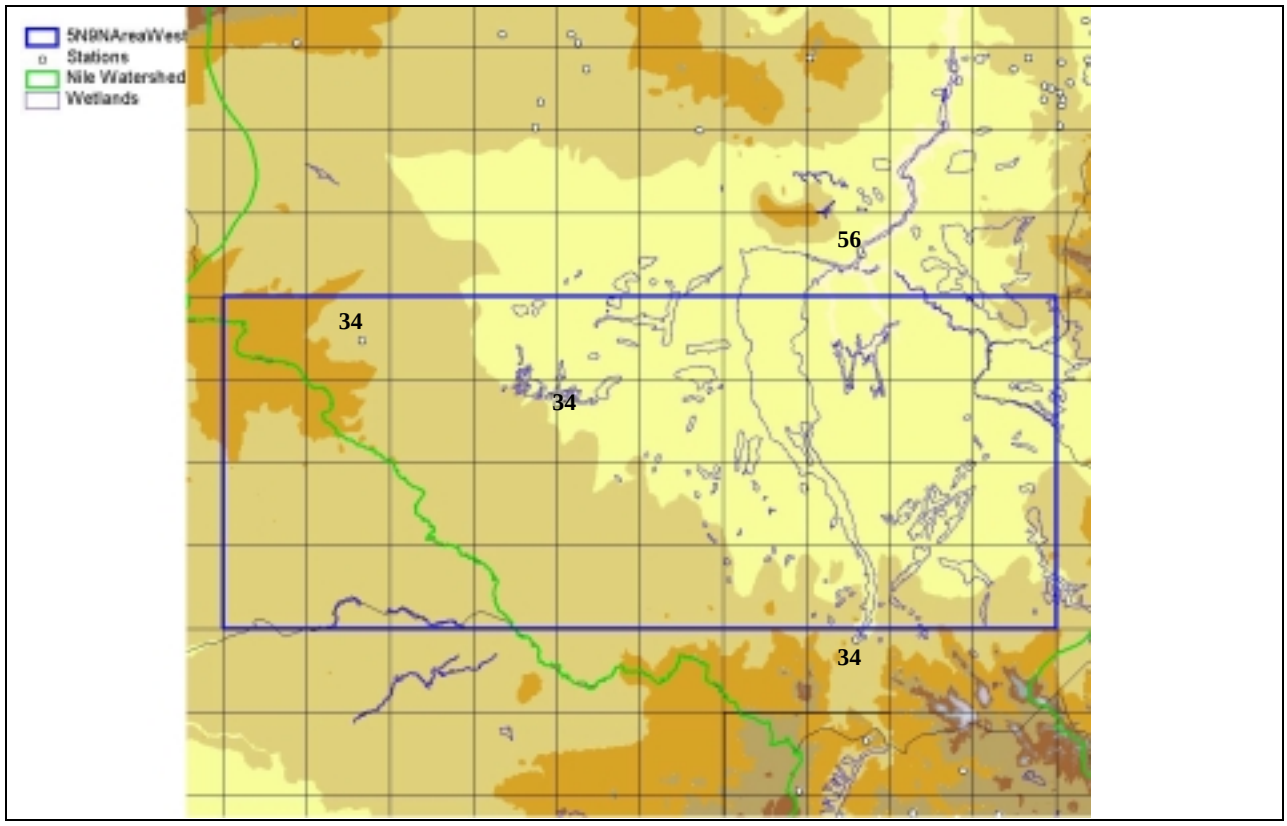


Figure 6.1: Calibration area [5°N, 9°N] and [24°E, 34°E] region and distribution of stations reporting at least one month of valid data in 1996-1998. The numbers represent the decades available for calibration in 1996-1998 within the 1°x1° squares. Periods available for verification are twice those available for calibration.

6.2 Original GPI Model

GPI model features average BIAS, CORR, and MAE of 1.059, 0.846, and 1.059 for the stations in the calibration area. On the other hand, Figure 6.2 shows that while the correlation with single stations is relatively good (>0.66) and uniform, the northeastern corner of the area is much drier than the southern and western parts of the area. This would suggest to further sub-divide the calibration area. On the other hand, the small number of rain gages cannot support this approach.

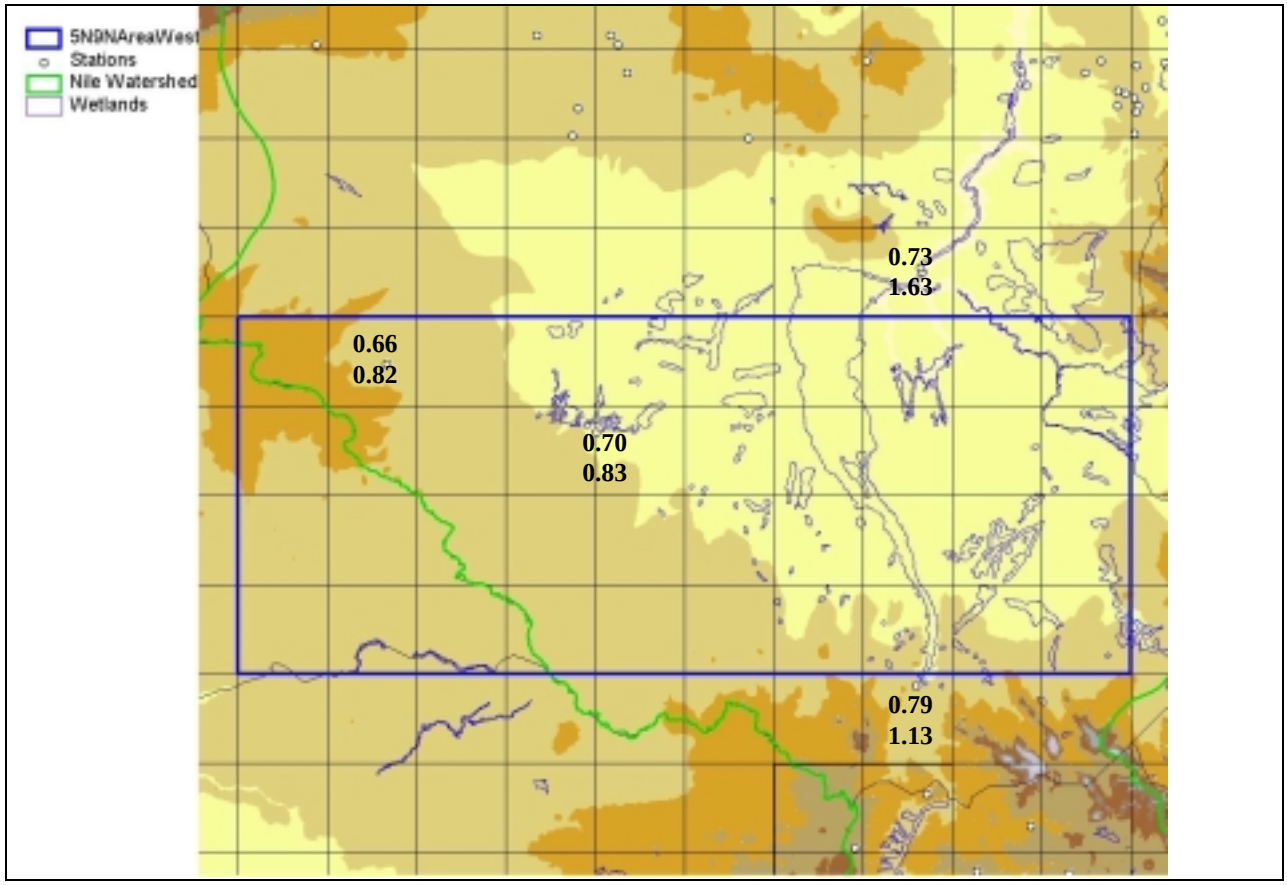


Figure 6.2: Spatial distribution of the GPI algorithm's verification CORR (top row) and BIAS (lower row); [5°N, 9°N] and [24°E, 34°E] region

6.3 Constrained CCD Calibration Based on Infra Red (IR) Radiation

The constrained IR only calibration of the CCD model produces an IR threshold of 233 °K and a rain rate of 1.464 mm/hr. The average BIAS improves to 0.067, the average CORR remains unchanged to 0.846, while the MAE reduces to 0.343. Figure 6.3 shows that the model is performing well over all stations: BIAS is contained to 0.17 and lowest correlation is 0.66. The value may seem low, but it is the correlation between rain gage and satellite precipitation estimations over a *single* pixel. Precipitation estimation over any number of pixels around the rain gage would resemble much more closely the precipitation field.

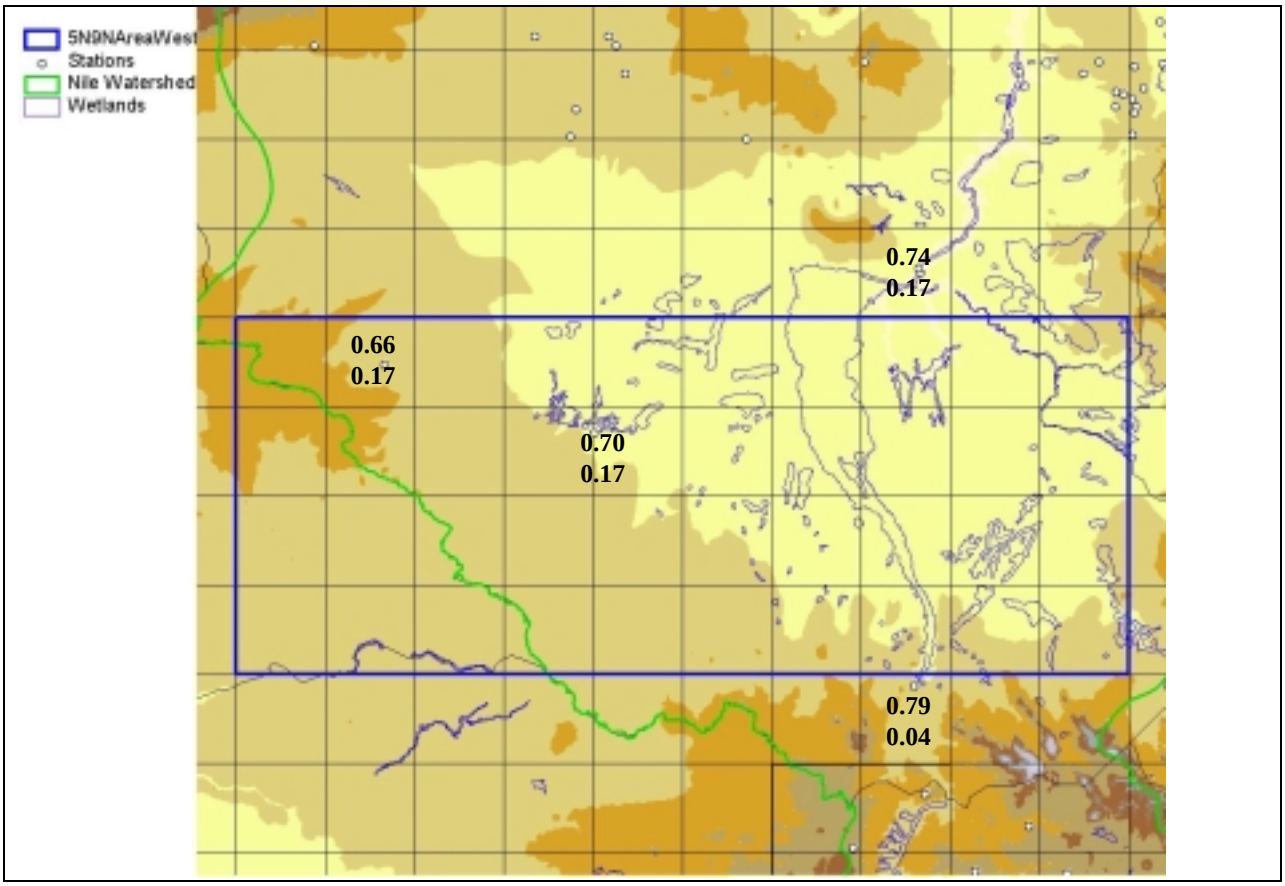


Figure 6.3: Verification Statistics for the Constrained CCD/IR Calibration; [5°N, 9°N] and [24°E, 34°E] region.

6.4 Constrained CCD Calibration Based on Infra Red (IR) and Visible (VIS) Radiation

The CCD parameters obtained with the constrained IR and VIS calibration (Table 6.4.1) attain improved average performance compared to both the GPI and IR only CCD. BIAS is kept to 0.069, like for the IR only model, but CORR reaches 0.861 while MAE decreases to 0.329.

Table 6.4.1: Constrained CCD/IR model's optimal parameters; [5°N, 9°N] and [24°E, 34°E] region

Latitude	Longitude	Nighttime IR Threshold (K)	Daytime IR Threshold (K)	VIS Threshold (Albedo, %)	Rain rate (mm/h)
[5°N, 9°N]	[24°E, 34°E]	231	251	36	1.452

As shown in Figure 6.4, the improvement in model performance pertains mainly to the southwestern part of the calibration area. It is also noted that while correlation improves in most stations, bias degrades in most stations, despite the fact that the average value remains practically the same.

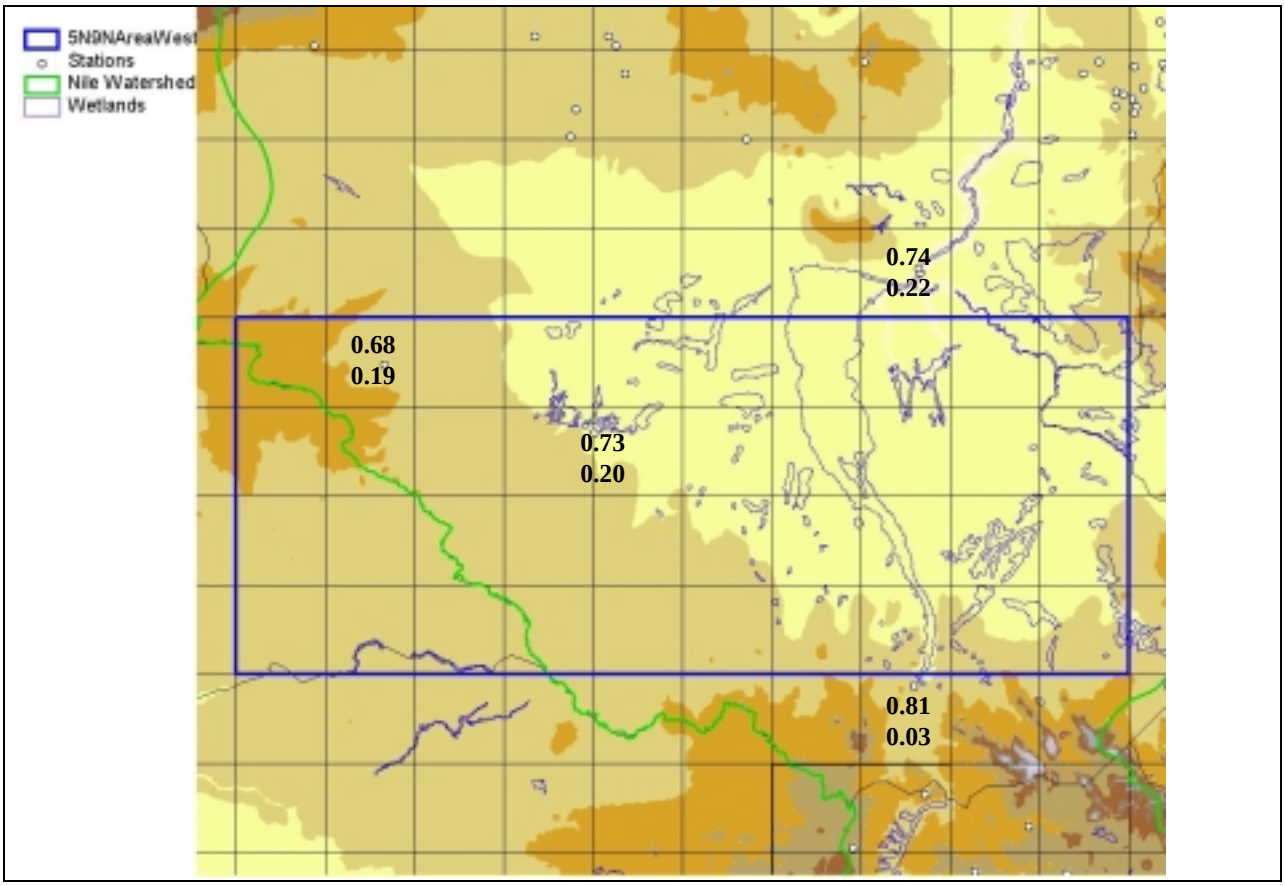


Figure 6.4: Spatial distribution of the constrained CCD IR + VIS model's CORR (top row) and BIAS (bottom row) in the [5°N, 9°N] and [24°E, 34°E] region.

Overall, satellite and rain gages match well considering the small number of stations and the large calibration area.

7. CCD Model Calibration for the [2°N, 5°N] and [34°E, 40°E] Region

7.1 Precipitation Records Available for Calibration

The eastern part of this latitude belt features a very diverse orography, but also a dense network of rain gages, albeit less so in its southern and western parts. The area includes two distinct subareas: the eastern two thirds, occupied by the southern half of the Ethiopian Plateau, and the western third, occupied by lower hills descending towards the Sudanese plains. The climate in the Ethiopian Plateau is strongly influenced by altitude. This would suggest proceeding with calibration at the finer possible spatial resolution. On the other hand, the number of ground stations available for calibration is limited, especially in the southwestern half of the area where two of 2^{o}x2^{o} squares do not host any station. Local calibration for these squares is possible (Figure 7.1), but only by considering the contribution of stations in neighboring squares. For this reason, it would be wise to avoid calibration at this spatial resolution.

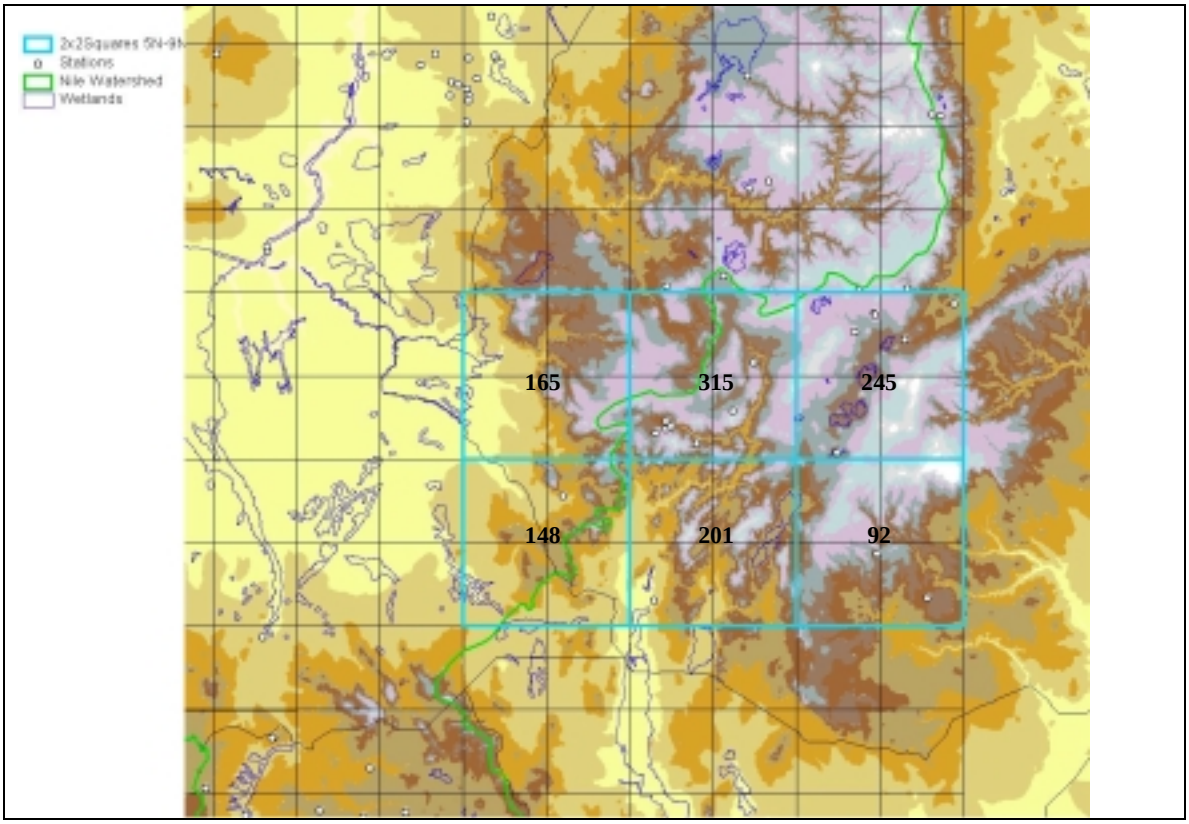


Figure 7.1: Spatial subdivision of the [5°N, 9°N] and [34°E, 40°E] region into 2°x2° squares and distribution of stations reporting at least one month of valid data in 1996-1998. The numbers represent the decades available for calibration in 1996-1998 within the enlarged 2°x2° squares. Periods available for verification are twice those available for calibration.

7.2 Original GPI Model

The GPI model is performing surprisingly well over this part of the basin. Its average BIAS is only 2%, the average CORR is 0.77, and the MAE is 35%. An inspection of the CORR and BIAS spatial distribution reveals that model performance over the area is somewhat uneven. In the western third of the area, where the Plateau descends toward the Sudanese plains, there appears to be a 15 to 17% precipitation underestimation and the correlation is lower than in the remaining two thirds of the calibration area. The poor rain gage density in the western part of the calibration area makes these performance assessment somewhat tentative.

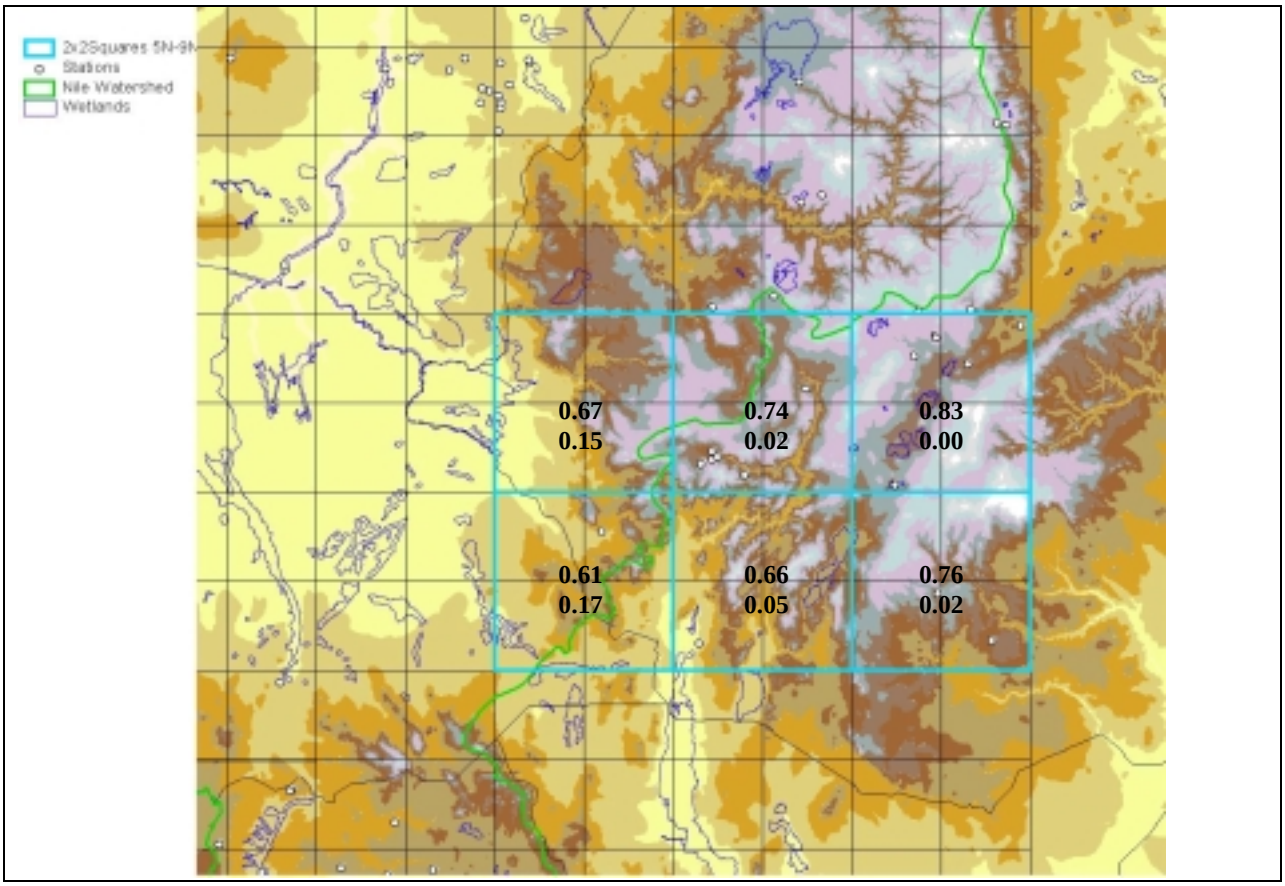


Figure 7.2: Spatial distribution of the GPI algorithm's verification CORR (top row) and BIAS (lower row); [5°N, 9°N] and [34°E, 40°E] region.

7.3 Constrained CCD Calibration Based on Infra Red (IR) Radiation

The behavior described in the previous section suggests partitioning the area in two parts: the western third and the eastern two thirds. Models are developed for both these areas. The optimal IR only parameters are 267 °K and 0.628 mm/h in the western part and 263 °K and 0.643 mm/h in the eastern part. Average BIAS increases to 11%, but average CORR improves to 0.84 and MAE to 0.28.

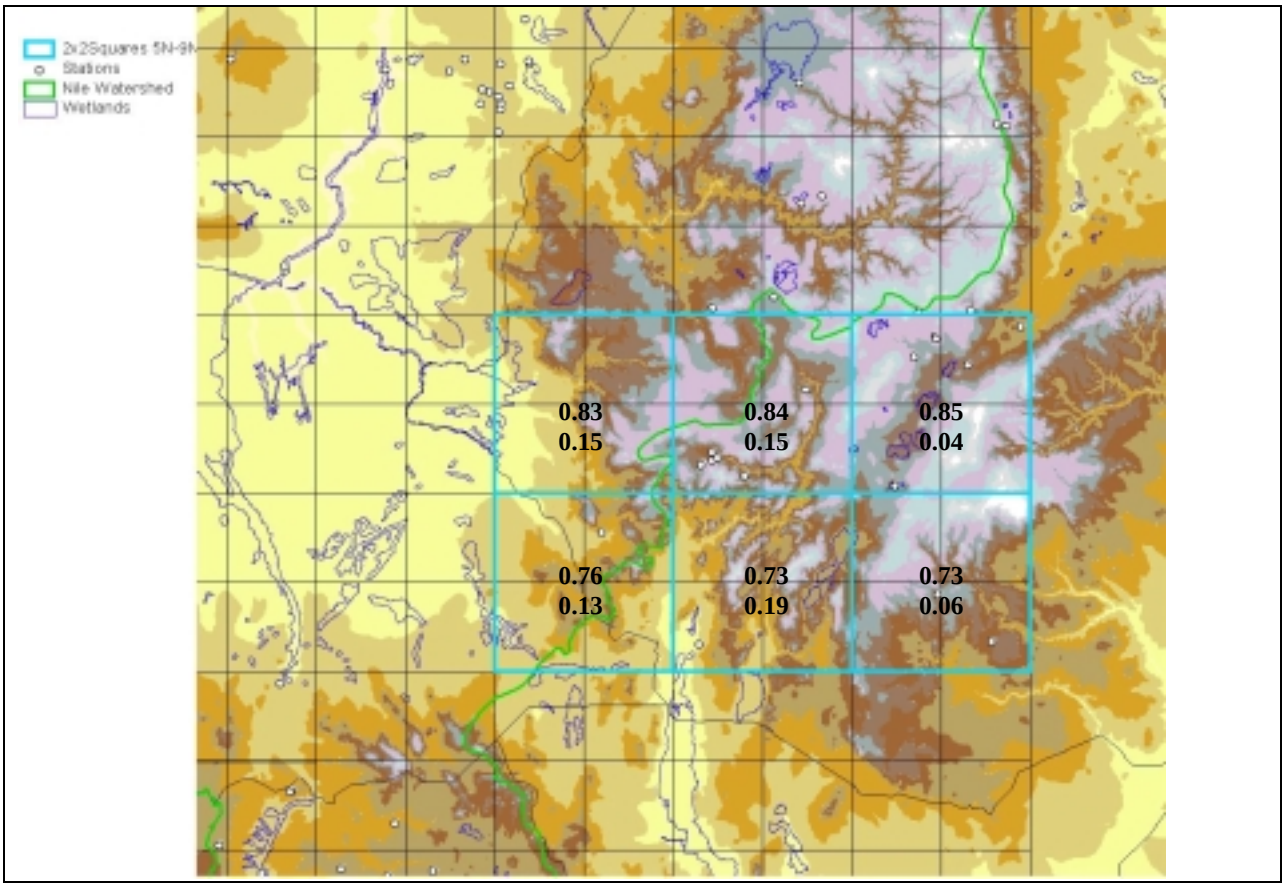


Figure 7.3: Spatial distribution of the constrained CCD/IR algorithm's verification CORR (top row) and BIAS (lower row); [5°N, 9°N] and [34°E, 40°E] region.

Figure 7.3 shows that model performance improves over the western part, especially with respect to correlation. In the eastern calibration area, correlation improves, but at the expense of a BIAS increase. However, BIAS is still very good.

7.4 Constrained CCD Calibration Based on Infra Red (IR) and Visible (VIS) Radiation

Table 7.4.1: Constrained CCD IR + VIS model's optimal parameters; [5°N, 9°N] and [34°E, 40°E] region

Latitude	Longitude	Nighttime IR Threshold (K)	Daytime IR Threshold (K)	VIS Threshold (Albedo, %)	Rain rate (mm/h)
[5°N, 9°N]	[34°E, 36°E]	263	273	22	0.683
[5°N, 9°N]	[36°E, 40°E]	253	273	22	0.790

Table 7.4.1 is typical of mountainous areas. The IR threshold is very high (actually the constraints have been relaxed to 273°K) while the VIS threshold is almost at a minimum. As discussed earlier, this may be the contribution of monsoonal and orographic precipitation.

The constrained IR and VIS CCD further improves the IR only CCD performance. Average BIAS is at 10%, CORR slightly improves to 0.85, and MAE decreases to 0.26. The improvements take place mainly in the eastern part as shown in Table 7.4.1. The optimal calibration parameters are given in Figure 7.4.1.

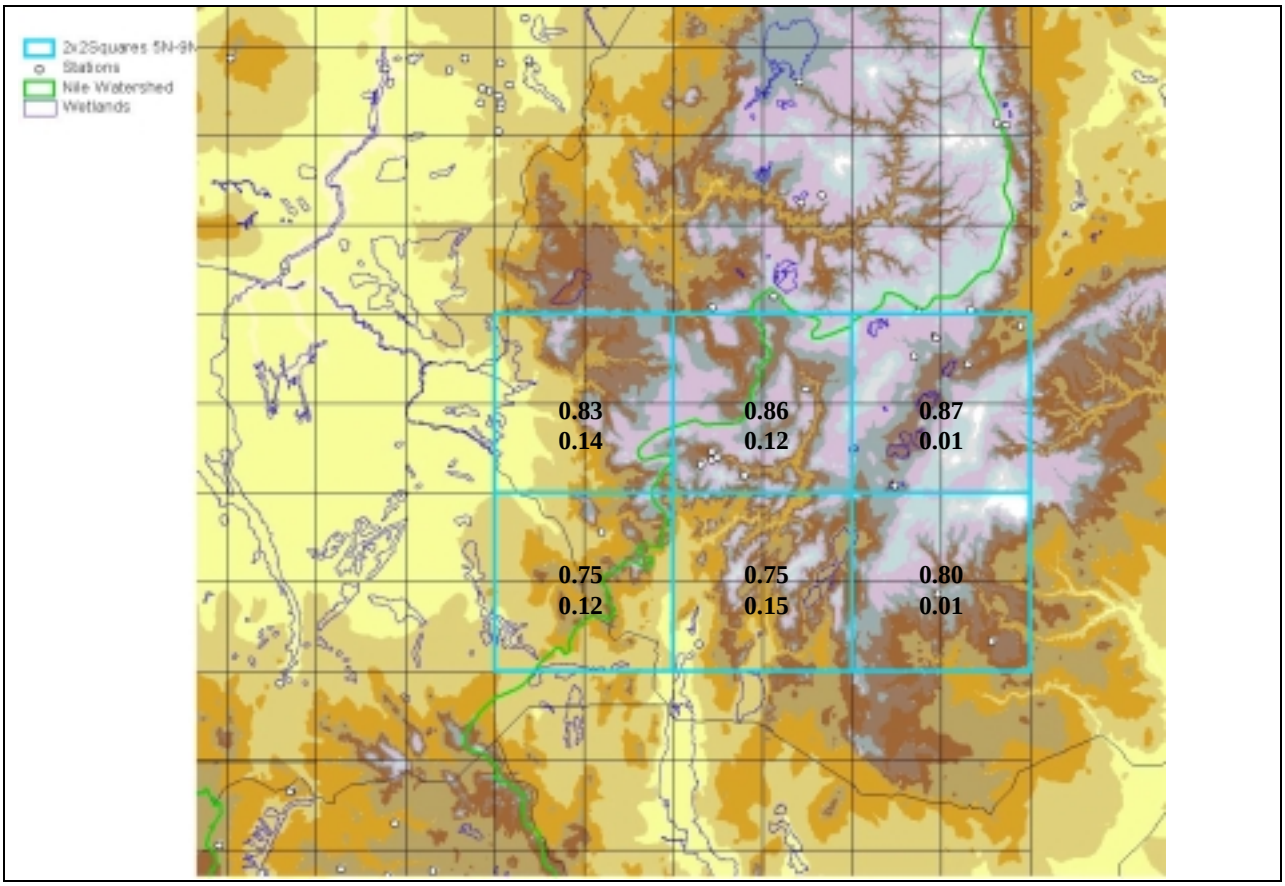


Figure 7.4: Spatial distribution of the constrained CCD/(IR + VIS) algorithm's verification CORR (top row) and BIAS (lower row); [5°N, 9°N] and [34°E, 40°E] region

Overall, the models perform very well with very low values of BIAS and MAE. The correlation is pretty high, varying between 0.75 and 0.87. A limitation for model calibration is that the number of rain gages over which to base the calibration process is low, especially in an area characterized by such dramatic changes in elevation and climate. Additional data from Ethiopia and the Lake Turkana area in Kenya (even if outside the Nile River Basin) would be very helpful for model calibration.

8. CCD Model Calibration for the [9°N, 15°N] and [24°E, 36°E] Region

8.1 Precipitation Records Available for Calibration

The number of rain gages in central Sudan is relatively high, but unfortunately their spatial distribution is not as uniform. Further, the precipitation patterns, especially in the northern part of this latitude belt, are heterogeneous. These elements combine to undermine the calibration process. The area has been divided into 3°x3° squares for calibration purposes. The extent of these areas and the scarcity of rain gages make the concepts of average BIAS, COR, and MAE within the calibration areas very tentative. To understand model performance, it is better to concentrate on model performance within 1°x1° sub-regions, within the larger calibration areas. Figure 8.1 shows that the number of available precipitation records for calibration is inadequate in the western part of this portion of the Nile river basin. It is also inadequate over Ethiopia (southeastern corner).

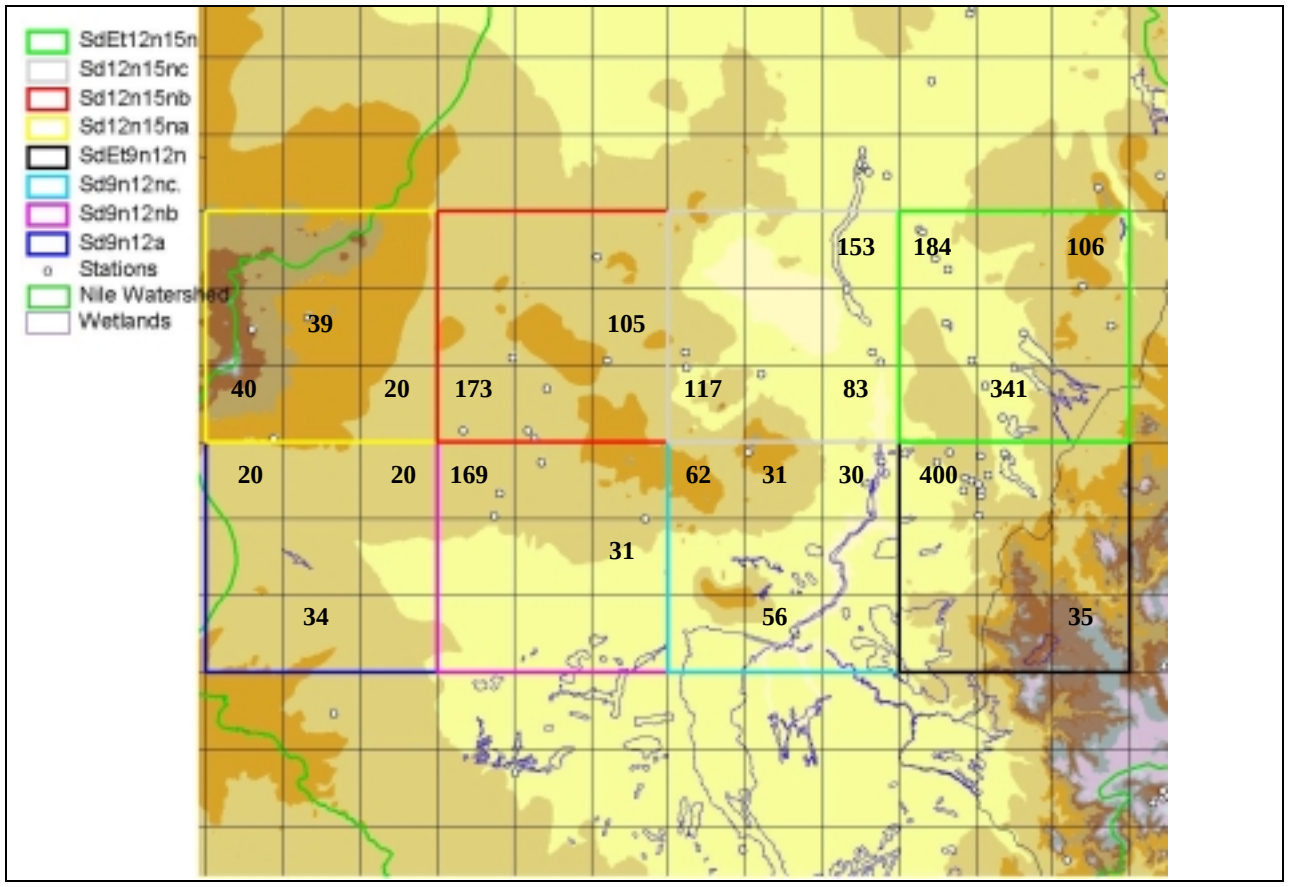


Figure 8.1: Spatial subdivision of the [9°N, 15°N] and [24°E, 36°E] region into 3°x3° squares and distribution of stations reporting at least one month of valid data in 1996-1998. The numbers represent the decades available for calibration in 1996-1998 within the enlarged 1°x1° squares within the calibration areas. Periods available for verification are twice those available for calibration.

It is noted that not all of the 1°x1° squares have been chosen, but it was tried to partition the rain gages within each 3°x3° square into independent enlarged 1°x1° squares.

8.2 Original GPI Model

The correlation between GPI and rain gages is very high, especially considering that in many cases it is actually correlation between satellite and raingage precipitation estimates over one or two pixels. The "averaging effect" that takes place when there are several stations within the same 1°x1° area and improves model performance is confined here to very few areas.

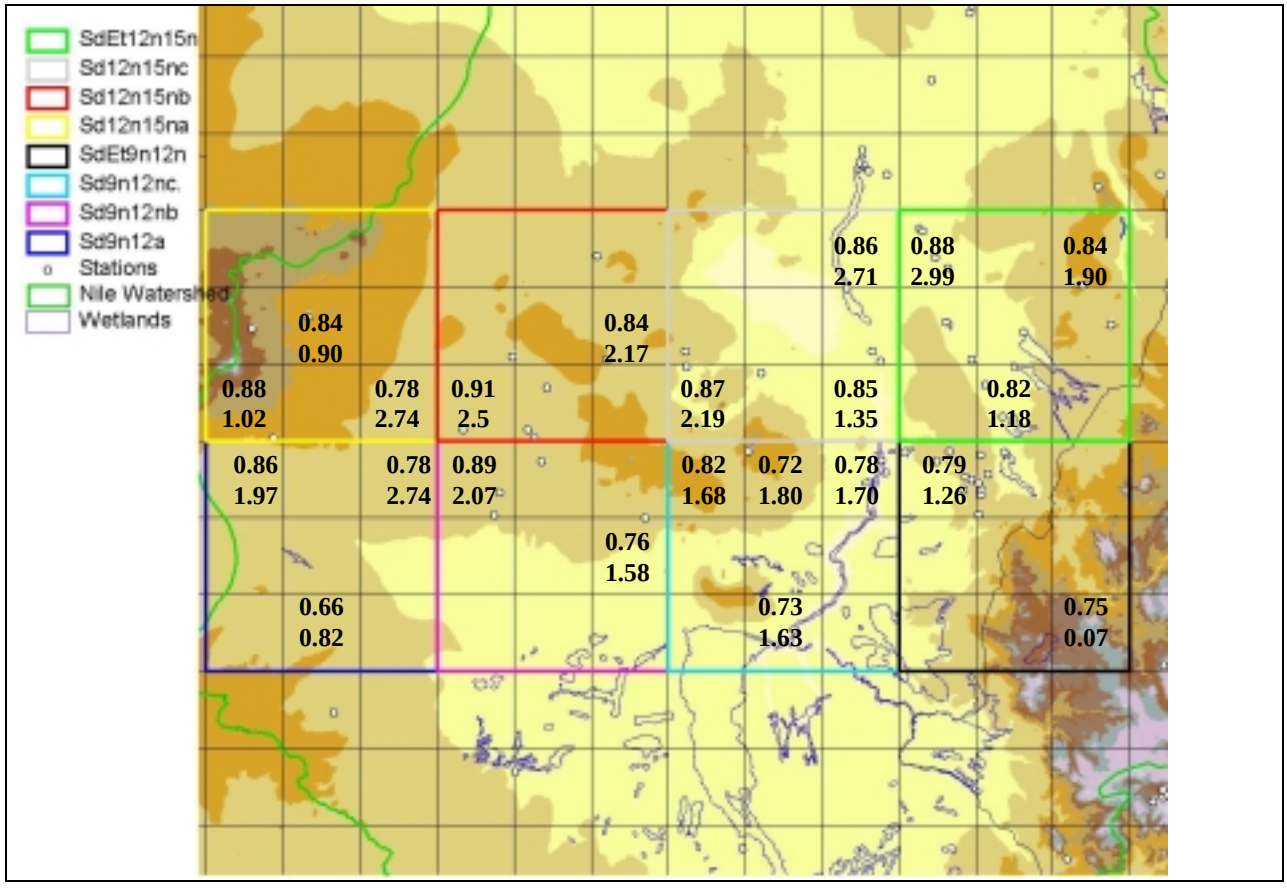


Figure 8.2: Spatial distribution of the GPI algorithm's verification CORR (top row) and BIAS (lower row); [9°N, 15°N] and [24°E, 36°E] region

Some trends are apparent from Figure 8.2:

- BIAS in the extreme west (blue and yellow squares) is high (~ 100%);
- BIAS in the central areas (eastern sides of the blue and yellow squares, purple, red, cyan, gray squares, and some parts of the green and black squares) is very high (~ 120-270%);
- BIAS over the Ethiopian plateau is very low (7%).

High BIAS, accompanied by relatively high correlation, indicates that precipitation is associated with cold clouds, but that its intensity is low, much lower than the standard rain rate of 3.0 mm/hr. The dryness of the air over this area could cause the evaporation of rainfall generated by cold clouds before it reaches the ground.

Whatever the explanation, the natural partition of the area would be in three parts: the western extreme, the central part, and the eastern part. However, calibration of the western part did not perform well. The most likely explanation is the low number of stations (4), half of which include short records. Consequently, the four westernmost 3°x3° squares of Figure 8.1 have been bundled together has a unique calibration area.

The higher number of rain gages in the remaining four 3°x3° squares makes it possible to compute a separate set of parameters for each.

8.3 Constrained CCD Calibration Based on Infra Red (IR) Radiation

The optimal IR thresholds and rain rates reported in Figure 8.3 show a nucleus of low thresholds on the western part and a nucleus of high IR thresholds over the eastern part.

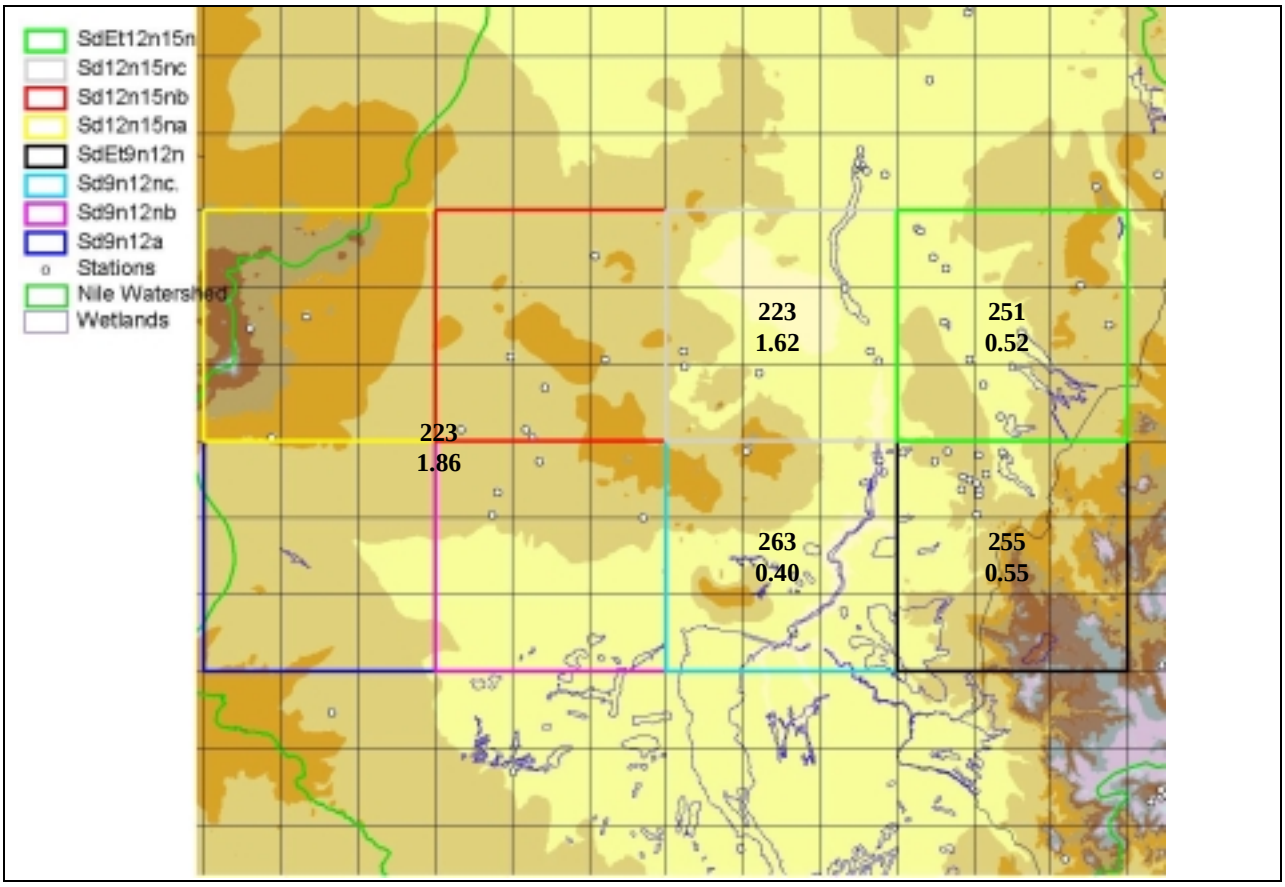


Figure 8.3: Spatial distribution of the constrained CCD/IR model's IR thresholds (°K) and rain rates (mm/h) in the [9°N, 15°N] and [24°E, 36°E] region.

Interestingly, the warm IR thresholds and corresponding rain rates here are similar to those found over the highlands in the western part of the Lake Victoria. On the other hand, the very cold thresholds here are paired to less intense rain rates than in the Lake Victoria area (Figure 4.3).

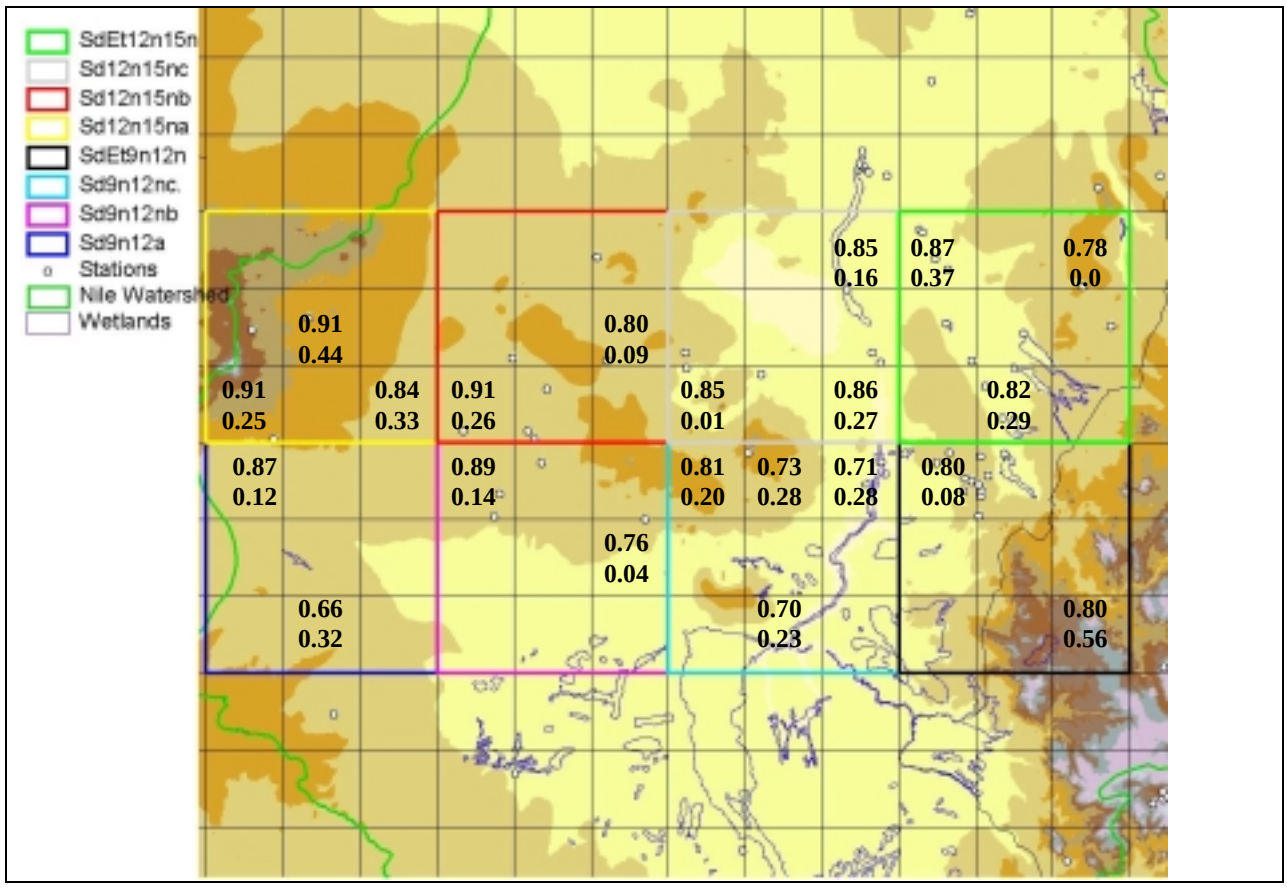


Figure 8.4: Spatial distribution of the constrained CCD/IR model's CORR (top row) and BIAS (lower row) in the [9°N, 15°N] and [24°E, 36°E] region.

The calibration is relatively successful: calibration is very similar to that achieved by GPI, almost always above 0.7, but BIAS is much lower. On the other hand, BIAS is frequently over 25%, a value that begins to be significant. There are three aspects to consider in evaluating the IR only CCD performance:

- BIAS is almost as frequently positive as it is negative, so that the total BIAS over the calibration areas is likely lower;
- The number of stations in a 1°x1° area is often only one or two, making the matching more difficult;
- The lower is the average precipitation, the higher absolute error becomes in percentage terms. That is, 0.5 mm/h error over an average rain of 5 mm/hr corresponds to 10% BIAS, while the same absolute error over a 2 mm/hr is equal to 25% BIAS. The impact of this error in a hydrologic simulation exercise, however, is likely less than BIAS would otherwise suggest.

The BIAS value of 0.56 in the southeastern corner of the black square indicates that the model is not valid for this region. The reason is that in the model calibration, the 400 decadal values from the drier northwestern corner heavily outnumber the 35 decadal values from the wetter southeastern corner. The result is that the model simulates the lowland rain patterns but cannot be successfully applied to the highlands. To somewhat remedy this problem, the coefficients determined in Chapter 7 are applied to the southeastern corner of this area.

8.4 Constrained CCD Calibration Based on Infra Red (IR) and Visible (VIS) Radiation

The constrained calibration of the CCD based on IR and VIS is reproducing the IR only pattern. Low IR and high VIS thresholds are found over the western part, and high IR and low VIS thresholds are found over the southeastern part (Figure 8.5). The effect of the very high VIS threshold and the very low IR threshold over the western part of the basin is to screen out most clouds, except those that are particularly deep and cold.

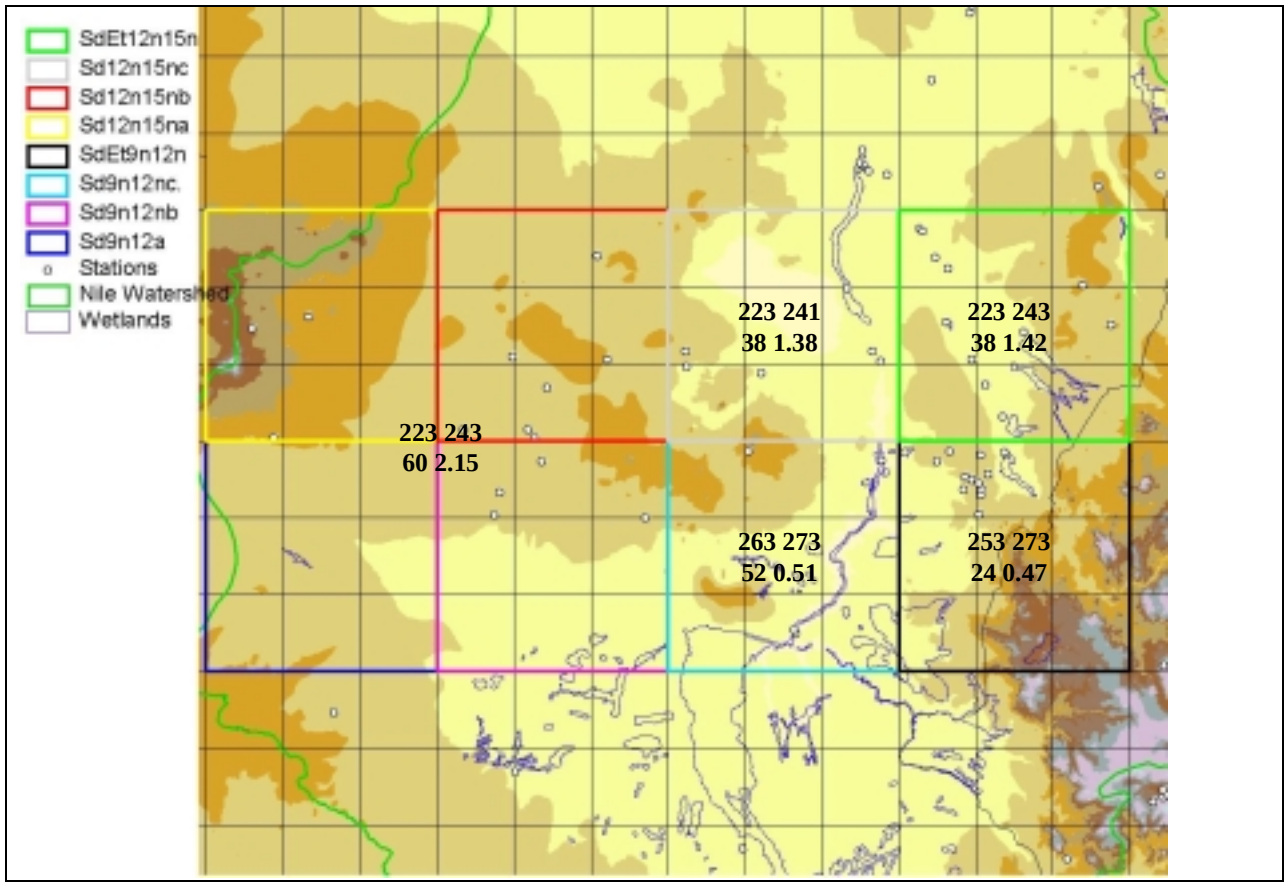


Figure 8.5: Spatial distribution of the constrained CCD IR + VIS model's parameters (top row nighttime IR and daytime IR thresholds (°K); lower row VIS threshold (albedo, %) and rain rate (mm/h)) in the [9°N, 15°N] and [24°E, 36°E] region

This model performance is somewhat better than the IR only performance (Figure 8.6), especially in terms of CORR and MAE.

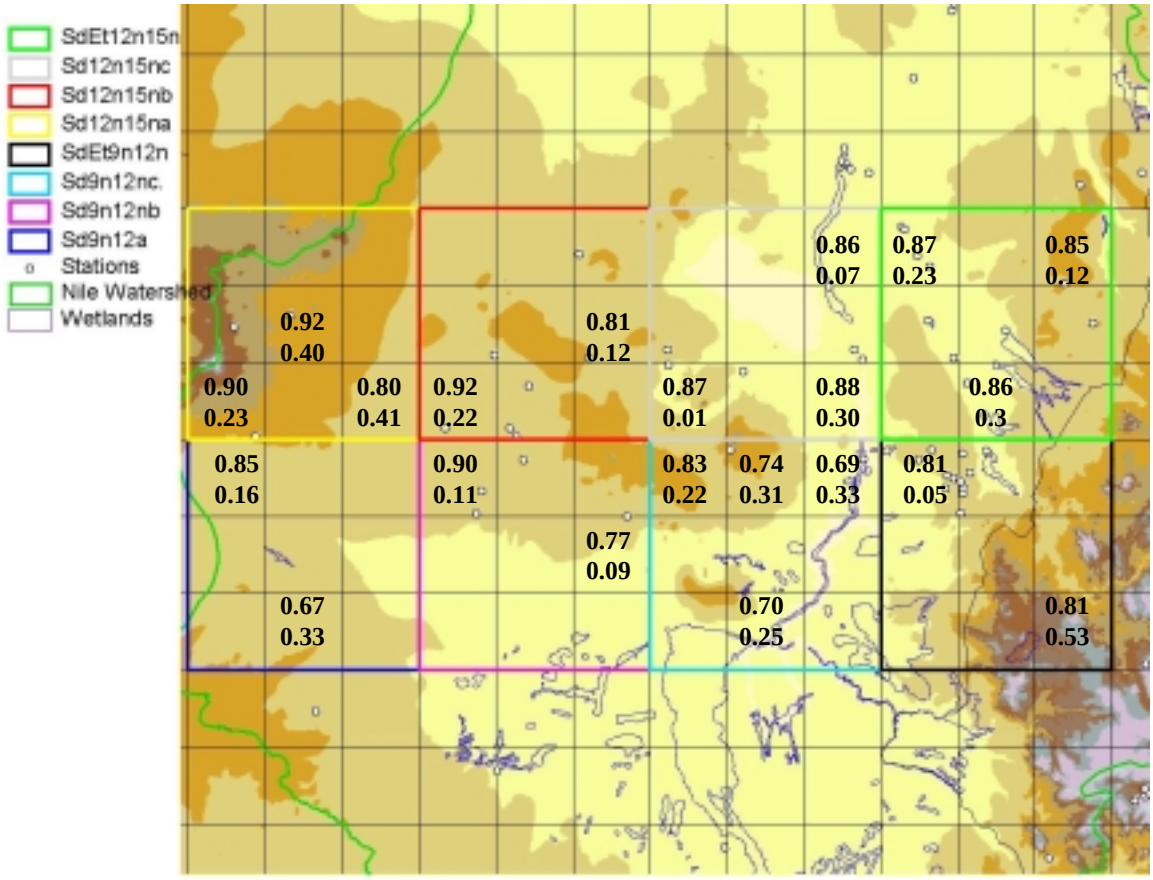


Figure 8.6: Spatial distribution of the constrained CCD/(IR + VIS) model's CORR (top row) and BIAS (bottom row) in the [9°N, 15°N] and [24°E, 36°E] region.

9. CCD Model Calibration for the [9°N, 15°N] and [36°E, 40°E] Region

9.1 Precipitation Records Available for Calibration

Nile DST rain gage density in this part of the Nile River basin is very low, especially over the northwestern foothills, where there is no precipitation record (Figure 9.1). The low rain gage density impacts model calibration, especially because the high orographic heterogeneity and geographical location of the area (being between the sea and the more arid interior), likely produce a variety of precipitation patterns.

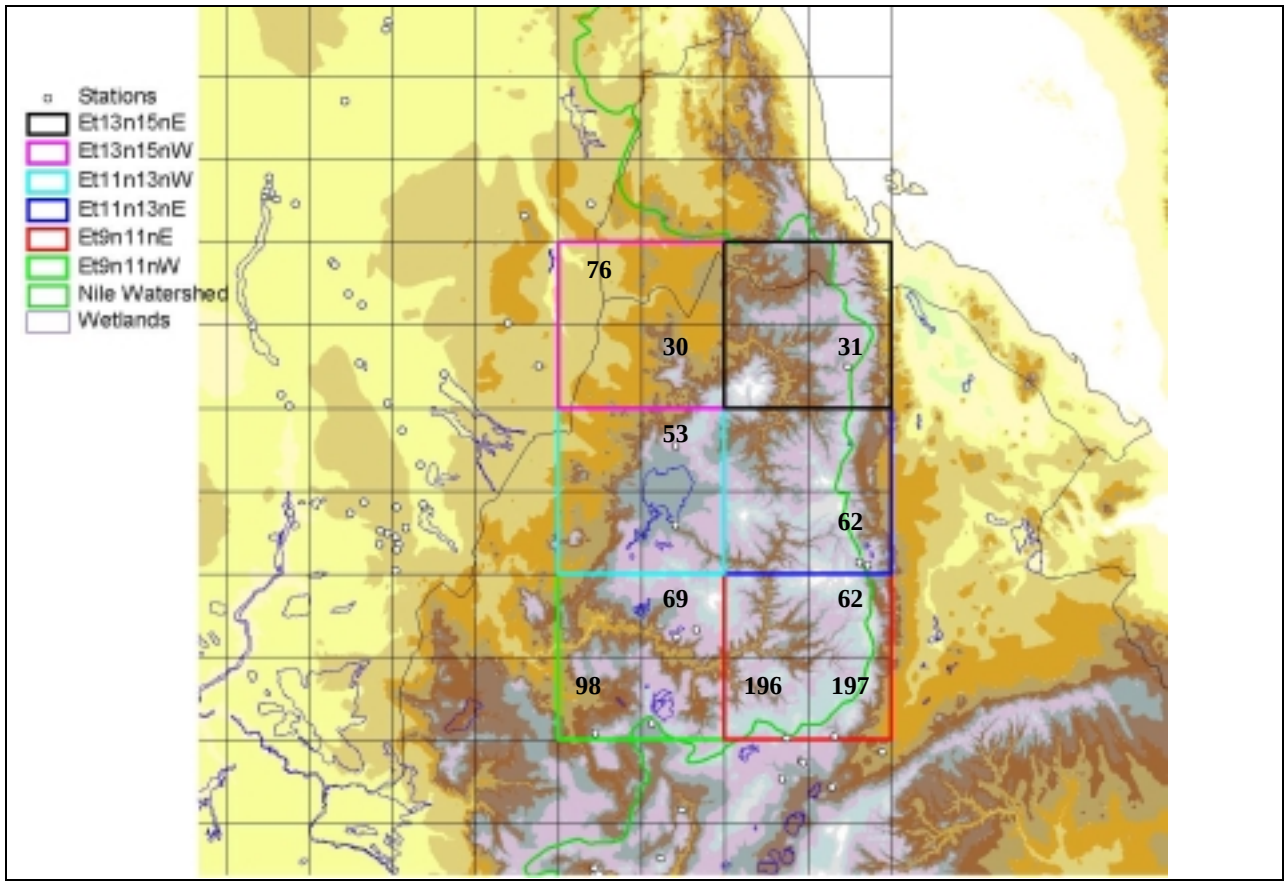


Figure 9.1: Spatial subdivision of the [9°N, 15°N] and [36°E, 40°E] region into 2°x2° squares and distribution of stations reporting at least one month of valid data in 1996-1998. The numbers represent the decades available for calibration in 1996-1998 within the enlarged 1°x1° areas. Periods available for verification are twice those available for calibration.

9.2 Original GPI Model

The GPI algorithm works well over the eastern parts were the BIAS is contained to just 20% (mostly underestimation) and the correlation with rain gages is high (>0.79). Over the western side, on the other hand, GPI estimates have a decreasing, bust still satisfactory correlation with

rain gages (>0.72). GPI grossly overestimates precipitation, however, reaching an unacceptable BIAS of 2.81 over the northwestern foothills (Figure 9.2).

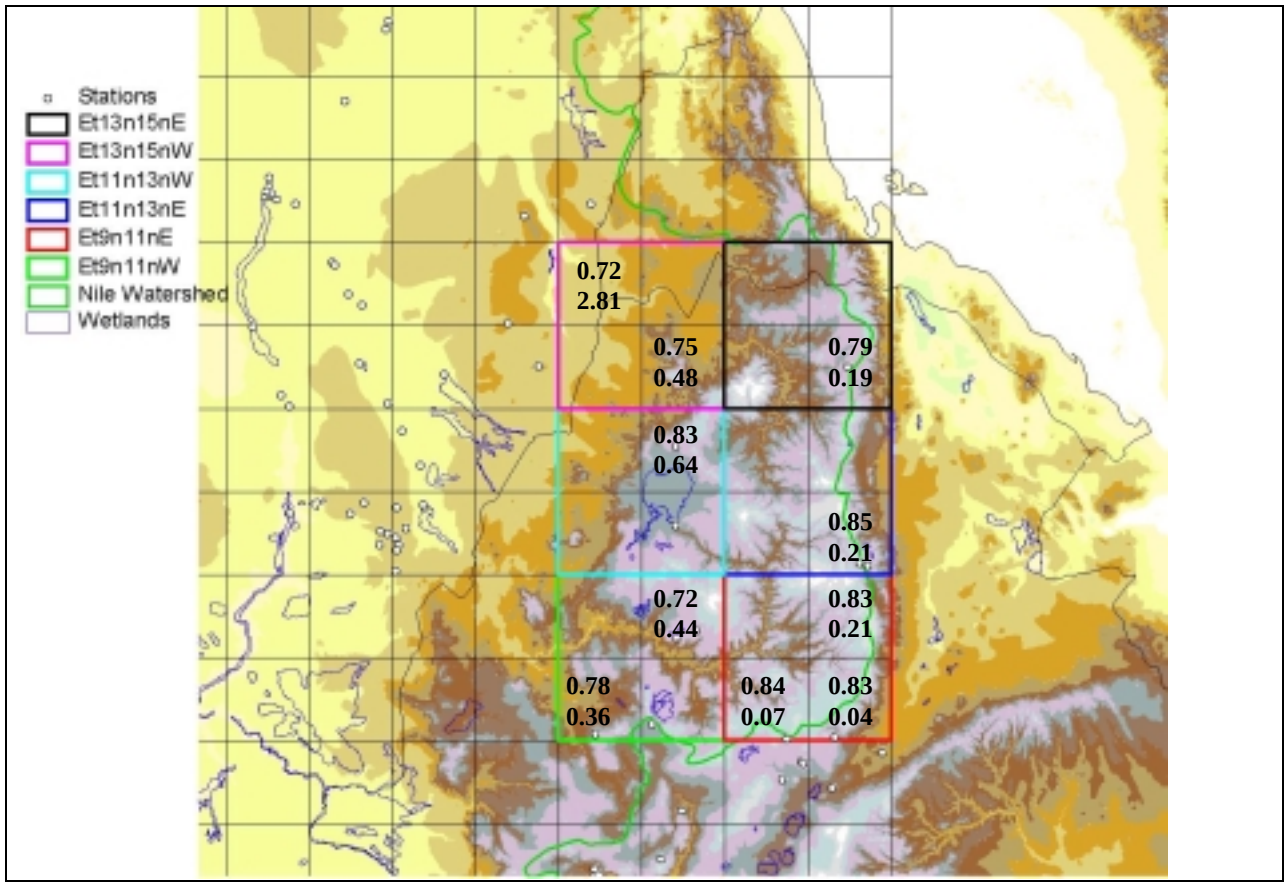


Figure 9.2: Spatial distribution of the GPI algorithm's verification CORR (top row) and BIAS (lower row); [9°N, 15°N] and [36°E, 40°E] region

Given that the scarcity of rain gages in this area prevents a meaningful CCD calibration over single 2°x2° areas, GPI model performance suggests partitioning this area into two regions, one extending over its western half and a second extending over its eastern half.

9.3 Constrained CCD Calibration Based on Infra Red (IR) Radiation

The optimal IR threshold and rain rate are respectively 257°K and 0.723 mm/hr for the western half and 255°K and 1.054 mm/h for the eastern half, indicating that the eastern part is wetter than the western part.

The CCD model is much better than GPI over the western half of the area both with respect to CORR and with respect to BIAS (Figure 9.3). The only exception is the northwestern corner of the area, where CCD improves with respect to BIAS but at the expense of CORR. Given that the BIAS is still too high, it is preferred to adopt the calibration set determined in Section 8.3 for the three northwestern 1°x1° areas.

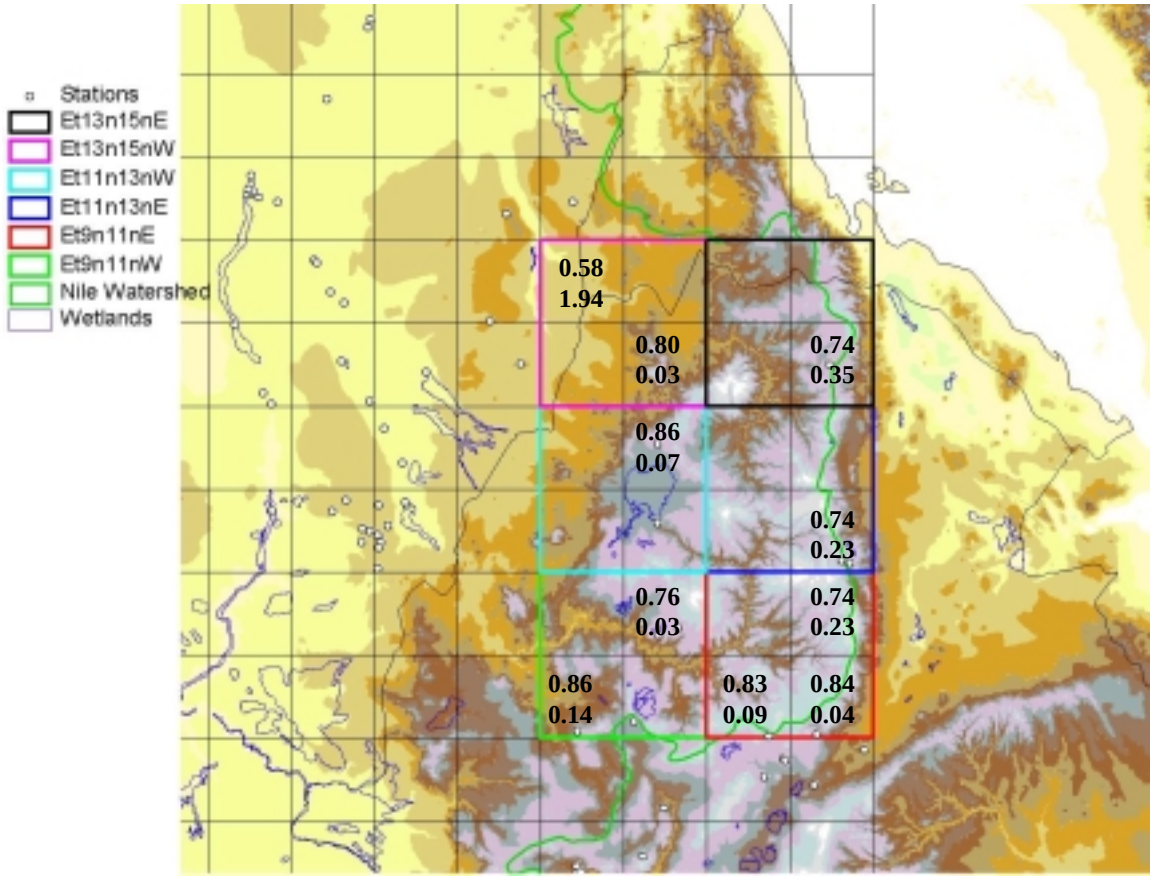


Figure 9.3: Spatial distribution of the constrained CCD/IR model's CORR (top row) and BIAS (lower row) in the [9°N, 15°N] and [36°E, 40°E] region.

In the eastern part model performance is worse than the GPI both with respect to CORR and BIAS. A further analysis for all the precipitation records (calibration and verification) shows that model performance is almost the same. It is thus preferred to keep the GPI set of parameters.

9.4 Constrained CCD Calibration Based on Infra Red (IR) and Visible (VIS) Radiation

The calibration of the CCD/(IR + VIS) model has been repeated with the same division of the area in two parts used for the CCD/IR. The two sets of parameters reported in Table 9.4.1 are notably different, thus justifying a posteriori the division in two halves of this calibration area.

Model performance is better than the CCD/IR both over the western and eastern halves (Figure 9.4). CORR is generally increased over all areas; BIAS shows an average improvement with the exception of some 1°x1° areas. Thus, this model is better than GPI not only over the western part, but also over the eastern part of the region.

Table 9.4.1: Constrained CCD IR + VIS model's optimal parameters; [9°N, 15°N] and [36°E, 40°E] region

Latitude	Longitude	Nighttime IR Threshold (K)	Daytime IR Threshold (K)	VIS Threshold (Albedo, %)	Rain rate (mm/h)
[9°N, 15°N]	[36°E, 38°E]	255	271	32	0.708
[9°N, 15°N]	[38°E, 40°E]	247	267	42	1.4

As for GPI and CCD/IR, CCD/(IR + VIS) does not perform well over the northwestern corner of the calibration area. As for the IR only CCD, the parameters determined in Section 8.4 will be used in the three northwestern 1°x1° areas.

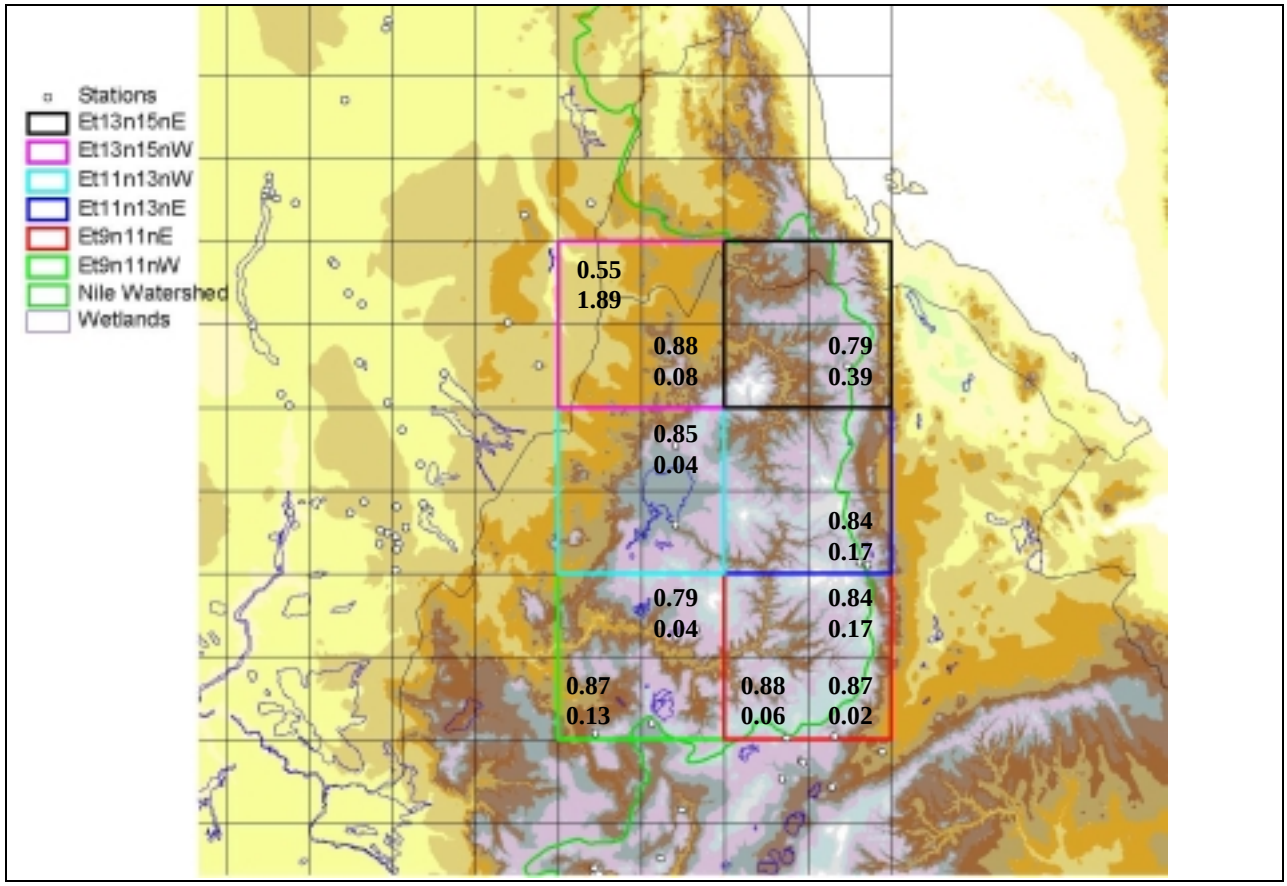


Figure 9.4: Spatial distribution of the constrained CCD/(IR + VIS) algorithm's verification CORR (top row) and BIAS (lower row); [9°N, 15°N] and [36°E, 40°E] region

10. CCD Model Calibration for the [15°N, 22°N] and [26°E, 38°E] Region

10.1 Precipitation Records Available for Calibration

Figure 10.1 shows that rain gage in this region are few and concentrated mainly along the banks of the Nile River. While it is probably safe to assume that precipitation over the northwestern portion of the area is relatively negligible, the same cannot be said about the southwestern portion of the basin or the ridge between the Red Sea and the Sudanese plains.

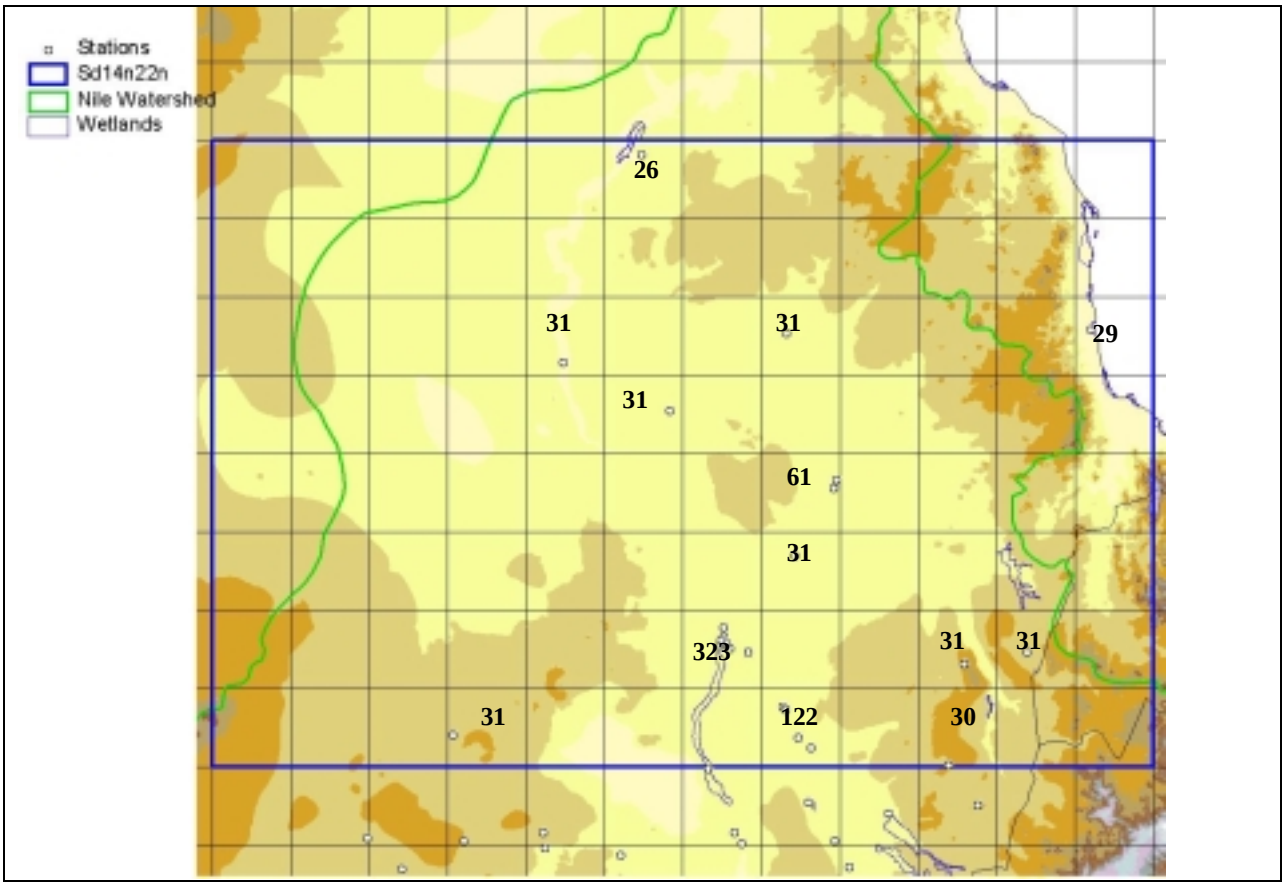


Figure 10.1: Spatial into distribution of stations reporting at least one month of valid data in 1996-1998 for the [15°N, 22°N] and [26°E, 38°E] region. The numbers represent the decades available for calibration in 1996-1998 within the enlarged 1°x1° squares. Periods available for verification are twice those available for calibration.

The calibration of the CCD model for this area is difficult because rainfall events are infrequent, concentrated over few months, and their variability is extremely high. A case in point is station 8009 (21.82°N, 31.48°E), where in the 1996-1998 period there is only one rain event, that happened to takes place during a gap in satellite coverage. In this case, the precipitation for the station is zero making the concepts of correlation and relative error meaningless. Even over the southern side, precipitation variability is such that calibration using the second decade of every month rather than the first yields completely different coefficients. For these reasons, it was

decided to use all available records for calibration instead of partitioning them into calibration and verification sets. Therefore, the results shown here are only calibration results. It was also preferred to extend the calibration area southward, including part of the area for which the model was calibrated in Chapter 7.

10.2 Original GPI Model

The correlation between GPI precipitation and rain gage precipitation in the southern part of the area shows a tenuous relation between cold clouds and precipitation. This relation is simply nonexistent for the upper two thirds of the area. BIAS is above 200% for all stations. It is noted, however, that because of the very dry climate, these unacceptable BIAS levels could be less important that they seem. For example, the average precipitation recorded at station 8009 (21.82°N, 31.48°E) is zero, while the satellite estimate is 0.16 mm/day. The difference is practically negligible for most purposes.

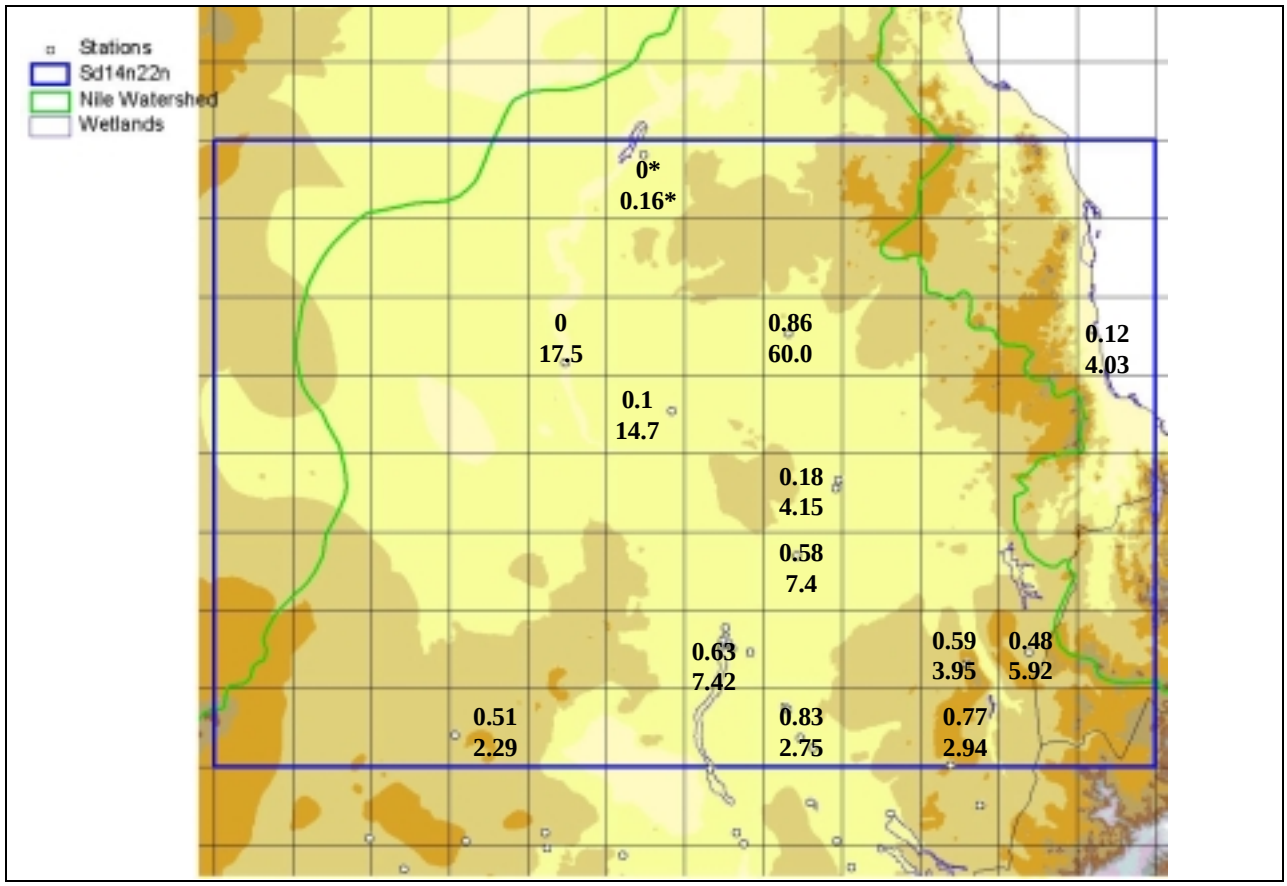


Figure 10.2: Spatial distribution of the GPI algorithm's verification CORR (top row) and BIAS (lower row); [5°N, 9°N] and [34°E, 40°E] region. *The value in the upper row is the average rain gage recorded precipitation, while the value in the lower row is the satellite based satellite estimate in mm/day.

From the map of GPI performance, it is possible subdivide the area in four zones:

- The northern portion of the basin, which includes the three northernmost stations;
- The eastern coastland, that is not part of the basin;
- The southeastern part of the basin, including Eritrea and part of the area west of the divide between the Red Sea and the interior;
- The remaining area.

10.3 Constrained CCD Calibration Based on Infra Red (IR) Radiation

The optimal CCD/IR IR thresholds reported in Table 10.3.1 are very low for the interior zone. This is probably because precipitation in such a dry region must be associated with rare strong convective phenomena. Furthermore, the rain rate associated with it is very low. (A comparison with Figure 4.3 is very useful in this respect.) Along the coast, precipitation is not tied to the migration of the Inter-Tropical Convergence Zone, as it is in the interior, but it is likely tied to masses of moist air blown over the sea by local or monsoonal winds that precipitate when reaching the coast. The associated clouds are not particularly cold and the rains exhibit mild intensity.

Table 10.3.1: Constrained CCD/IR model's optimal parameters; [15°N, 22°N] and [26°E, 38°E] region

Latitude	Longitude	IR Threshold (K)	Rain rate (mm/h)
[15°N, 19°N]	[26°E, 35°E]	223	1.4
[19°N, 22°N]	[26°E, 36°E]	228	0.3
[15°N, 19°N]	[35°E, 38°E]	223	1.7
[19°N, 22°N]	[36°E, 38°E]	248	0.25

Model performance (Figure 10.3) as measured by the statistics used for the rest of the basin is poor, although it shows an improvement over the CCD. In the southern part of the turquoise and purple areas, correlation is poor but acceptable. In the northern part and in the remaining areas, correlation is simply not meaningful. Precipitation is too scarce and sporadic to be reproducible. On the other hand, the models perform well with respect to BIAS. However, high BIAS does mean much in such arid climates, as the absolute difference between rain gage and satellite estimates may be as low as 0.01 mm/day. BIAS is negative for some stations and positive for others.

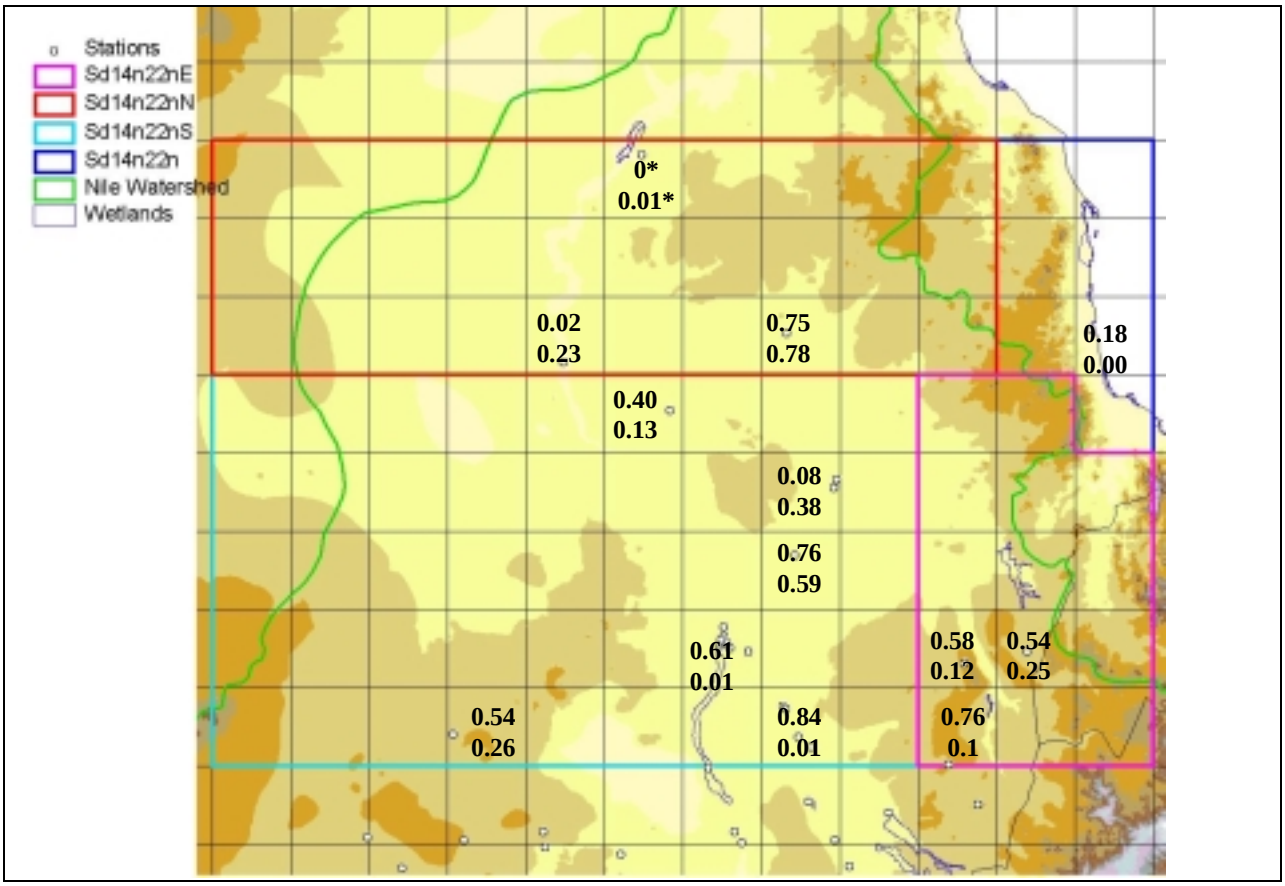


Figure 10.3: Spatial distribution of the constrained CCD/IR algorithm's verification CORR (top row) and BIAS (lower row); [15°N, 22°N] and [26°E, 38°E] region. *The value in the upper row is the average rain gage recorded precipitation, while the value in the lower row is the satellite based satellite estimate in mm/day.

10.4 Constrained CCD Calibration Based on Infra Red (IR) and Visible (VIS) Radiation

The optimal parameters for the CCD/(IR + VIS) follow the same trend seen for the CCD/IR. IR thresholds are very low over the interior and relatively high over the coast (Table 10.4.1).

Table 10.4.1: Constrained CCD/(IR + VIS) model's optimal parameters; [15°N, 22°N] and [26°E, 38°E] region

Latitude	Longitude	Nighttime IR Threshold (K)	Daytime IR Threshold (K)	VIS Threshold (Albedo, %)	Rain rate (mm/h)
[15°N, 19°N]	[26°E, 35°E]	223	238	52	1.42
[19°N, 22°N]	[26°E, 36°E]	231	241	44	0.086
[15°N, 19°N]	[35°E, 38°E]	225	245	44	1.566
[19°N, 22°N]	[36°E, 38°E]	248	255	29	0.235

Model performance improves compared to both CCD/IR and GPI but is still poor. The performance over the wetter southern third of the area is better than the upper part, where it cannot be fairly judged due to the arid climate. BIAS is contained for all the stations, especially in consideration of the little precipitation in the area.

Despite these shortcomings, the set of CCD/IR and CCD/(IR + VIS) coefficients from the northern sub-region of this area is used for the ungaged area between 22°N and 28°N.

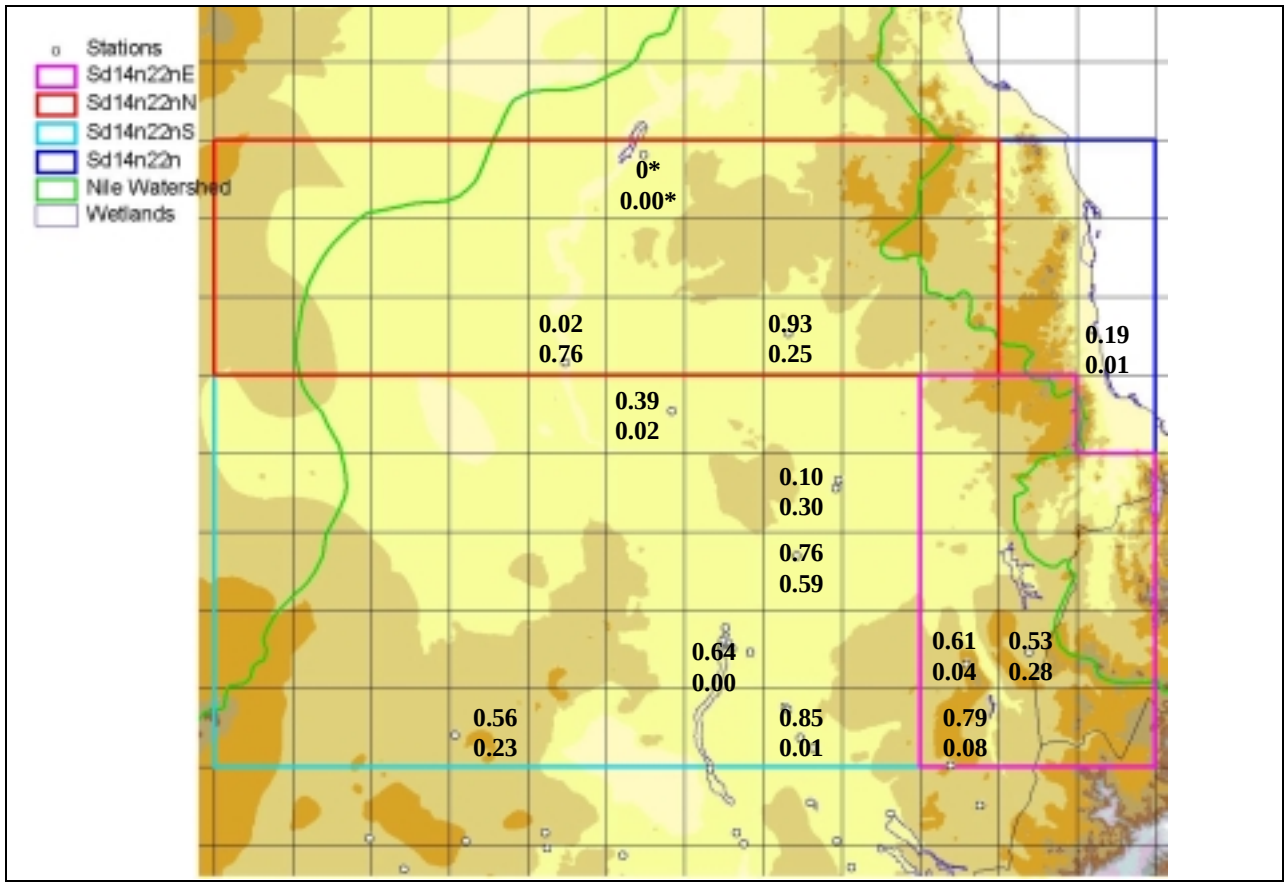


Figure 10.4: Spatial distribution of the constrained CCD/IR model's CORR (top row) and BIAS (lower row); [15°N, 22°N] and [26°E, 38°E] region. *The value in the upper row is the average rain gage recorded precipitation, while the value in the lower row is the satellite based satellite estimate in mm/day.

11. CCD Model Calibration for the [30°N, 32°N] and [30°E, 32°E] Region

11.1 Precipitation Records Available for Calibration

Precipitation records from several rain gages are available for the Nile Delta area. However, despite their relatively high number, they fail to adequately represent rainfall distribution because they are concentrated mainly along the coast since rainfall patterns vary significantly from the coast southward. Furthermore, satellite images for this area are not available in the periods July-December 1997 and August-December 1998, restricting the pool of records that can be used for model calibration.

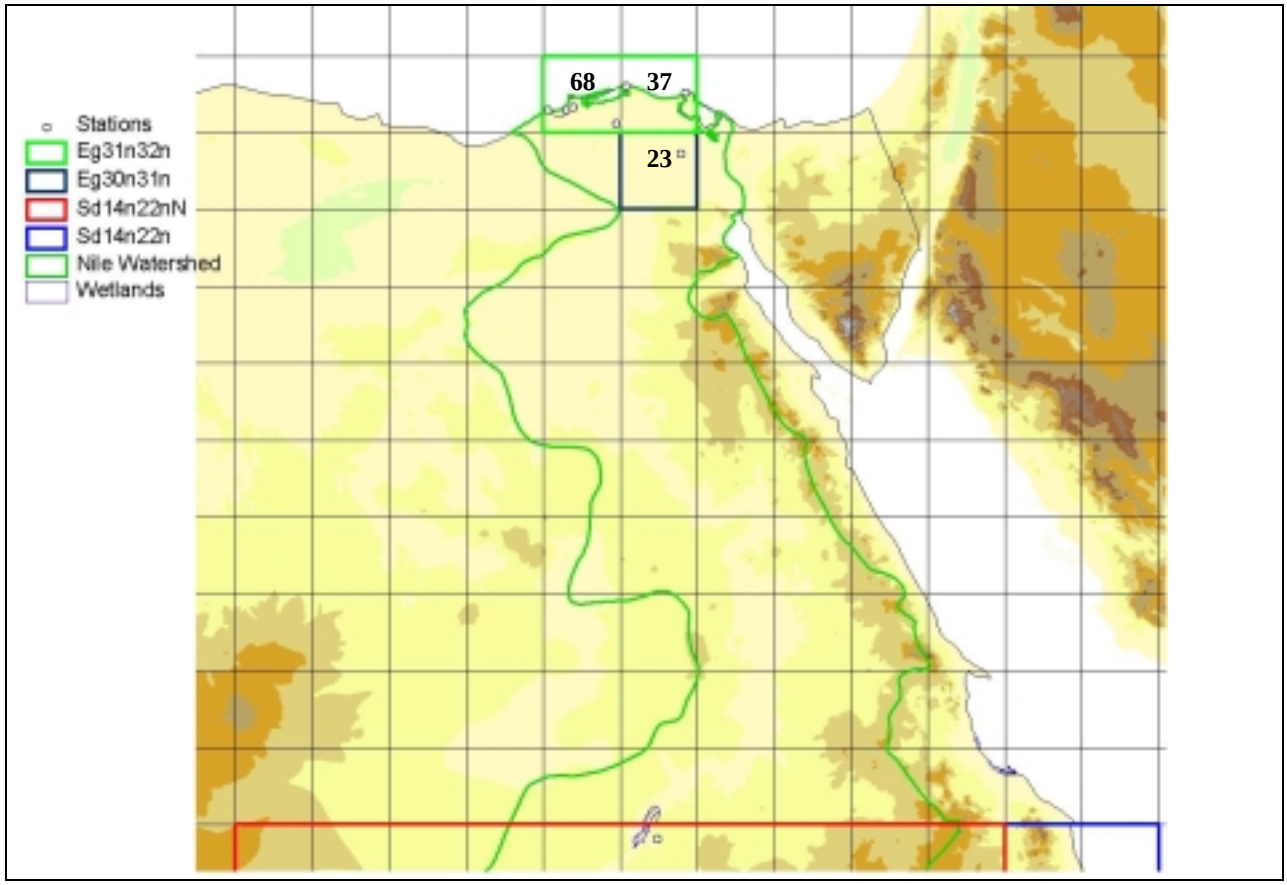


Figure 11.1: Spatial subdivision of the [30°N, 32°N] and [30°E, 32°E] region into 1°x1° squares and distribution of stations reporting at least one month of valid data in 1996-1998. The numbers represent the decades available for calibration in 1996-1998 within the 1°x1° squares. Periods available for verification are twice those available for calibration.

11.2 Original GPI Model

Precipitation is over the this part of the Nile River basin is mainly generated by the tails of extra-tropical fronts passing over the Mediterranean Sea during the winter. These cloud systems are not very high and feature mainly light rain. CCD models are not fitted for this type of

precipitation. Furthermore, additional light precipitation could be caused by the displacement of masses of moist air from the sea towards the interior that is also not fitted for CCD applications. Given these premises, it is not surprising that GPI is not performing well in this part of the Nile River basin. Correlation is slightly negative, a sign that there is no relationship between precipitation and highly convective clouds. The BIAS over the coastal area is high, but within normal model performances, while over the interior is very high. This difference reflects the climatic differences of these two areas and suggest to partition the calibration into two sets: one for the coastal area and one for the interior, including the two innermost rain gages.

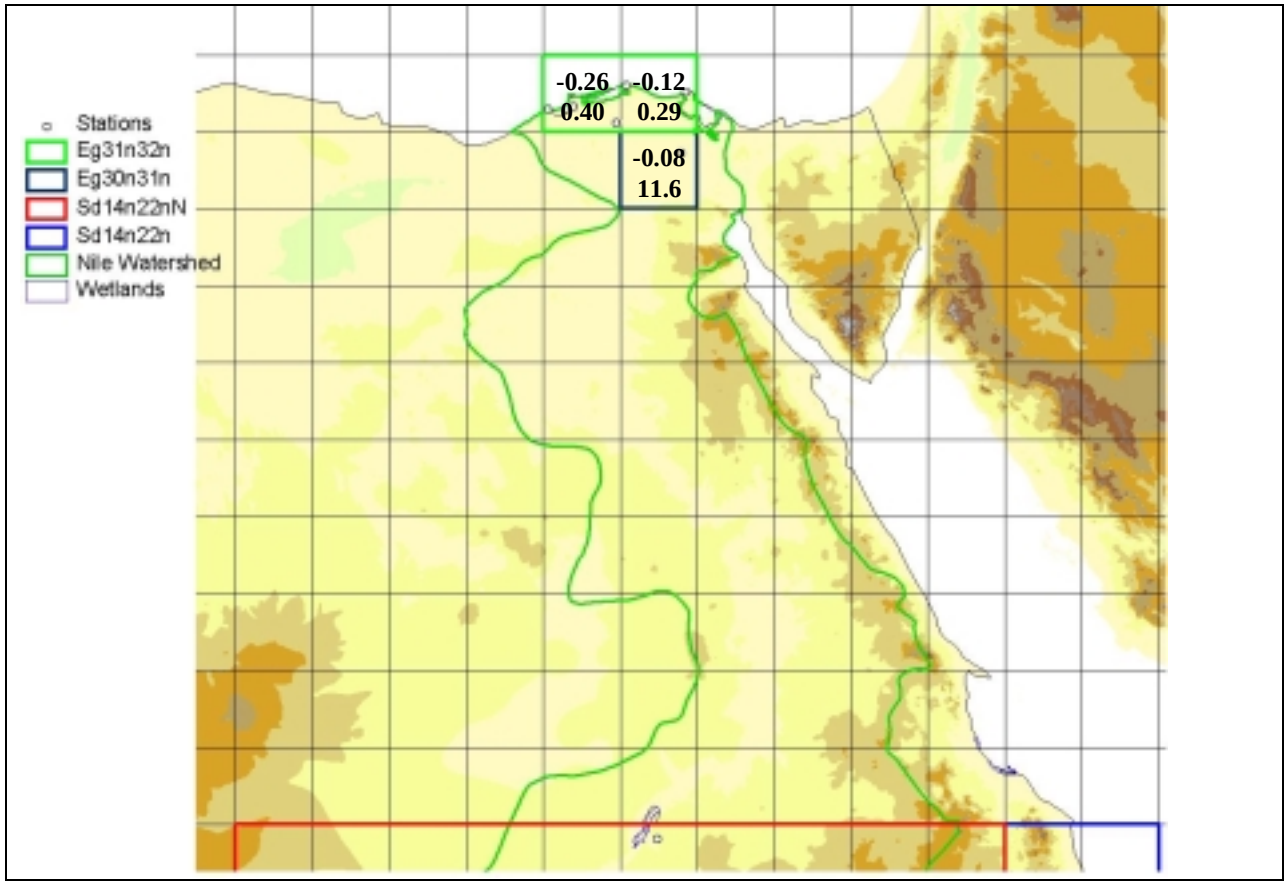


Figure 11.2: Spatial distribution of the GPI algorithm's verification CORR (top row) and BIAS (lower row); [30°N, 32°N] and [30°E, 32°E] region.

11.3 Constrained CCD Calibration Based on Infra Red (IR) Radiation

The optimal parameters determined in the CCD/IR calibration confirm that precipitation in this area of the basin is not tied to the presence of very high clouds and that precipitation is very low, especially inland (Table 11.3.1).

Table 11.3.1: Constrained CCD/IR model's optimal parameters; [30°N, 32°N] and [30°E, 32°E] region

Latitude	Longitude	IR Threshold (K)	Rain rate (mm/h)
[31°N, 32°N]	[30°E, 32°E]	273	0.15
[30°N, 31°N]	[31°E, 32°E]	273	0.04

Model correlation with rain gages is acceptable along the coast, but not inland. On the other hand, BIAS is much better inland than along the coast (Table 11.3.2). The poor correlation with rain gages inland is probably due to the low total precipitation in this area.

Table 11.3.2: Constrained CCD/IR model's verification CORR and BIAS; [30°N, 32°N] and [30°E, 32°E] region

Latitude	Longitude	BIAS	CORR
[31°N, 32°N]	[30°E, 32°E]	0.25	0.69
[30°N, 31°N]	[31°E, 32°E]	0.02	0.31

11.4 Constrained CCD Calibration Based on Infra Red (IR) and Visible (VIS) Radiation

The introduction of the VIS channel improves model performance in terms of correlation, but increases the BIAS along the coast. Inland correlation improves, but is still far from being acceptable, while BIAS remains very good (Table 11.4.2).

Table 11.4.1: Constrained CCD/(IR + VIS) model's optimal parameters; [30°N, 32°N] and [30°E, 32°E] region

Latitude	Longitude	Nighttime IR Threshold (K)	Daytime IR Threshold (K)	VIS Threshold (Albedo, %)	Rain rate (mm/h)
[31°N, 32°N]	[30°E, 32°E]	273	273	60	0.22
[30°N, 31°N]	[31°E, 32°E]	273	273	46	0.057

Table 11.4.2: Constrained CCD/(IR + VIS) model's verification CORR and BIAS; [30°N, 32°N] and [30°E, 32°E] region

Latitude	Longitude	BIAS	CORR
[31°N, 32°N]	[30°E, 32°E]	0.30	0.74
[30°N, 31°N]	[31°E, 32°E]	0.02	0.36

Overall, CCD performance over this area of the Nile Basin is not satisfactory because this model is not suitable to simulate the precipitation patterns (light rain produced mainly by the fringes of

extra tropical systems) and the arid climate of the region. In this region, more ground station data are needed to lead to meaningful satellite-based rainfall estimation methods.

12. Remote Sensing of Lake Victoria Rainfall

Lake Victoria level fluctuations have puzzled hydrologists and meteorologists for many years. Lake levels vary as a result of rainfall and evaporation to and from the lake surface, inflow from the lake watersheds, outflow at Owen Falls, and possibly underground seepage. In spite of various monitoring campaigns, considerable uncertainty still remains regarding the relative magnitude of the lake water balance components. Measuring tributary inflow is problematic because many of the tributaries enter the lake through extensive wetland systems. There is also little knowledge about the lake seepage losses or gains and some controversy about lake outflow. The main uncertainty source, however, pertains to the precipitation and evaporation over the lake itself. In the past, precipitation over the lake has been traditionally assessed using the rain gages situated along the lake shores and over a handful of islands. However, in view of the vast lake surface area (~69,000 square kilometers), the reliability of the rainfall estimates continues to be debated.

This study does not aim to fully resolve this debate; it does, however, aim to demonstrate that remote sensing technology and models can be used to better understand the rainfall patterns on the lake surface and assess their contribution to the overall water balance. Given the availability of longer satellite records, a more definitive answer can be provided and used in water resources planning and management.

In this study, precipitation over the lake is estimated using conventional and remote sensing methods (described earlier) for two periods: June 1996 – June 1997, and June 1997 – June 1998. These periods have been selected because the beginning of June marks the end of the long rains and the beginning of the annual recession in lake level. It is thus easier to assess the effect of annual precipitation to lake level variation (Figure 12.1).

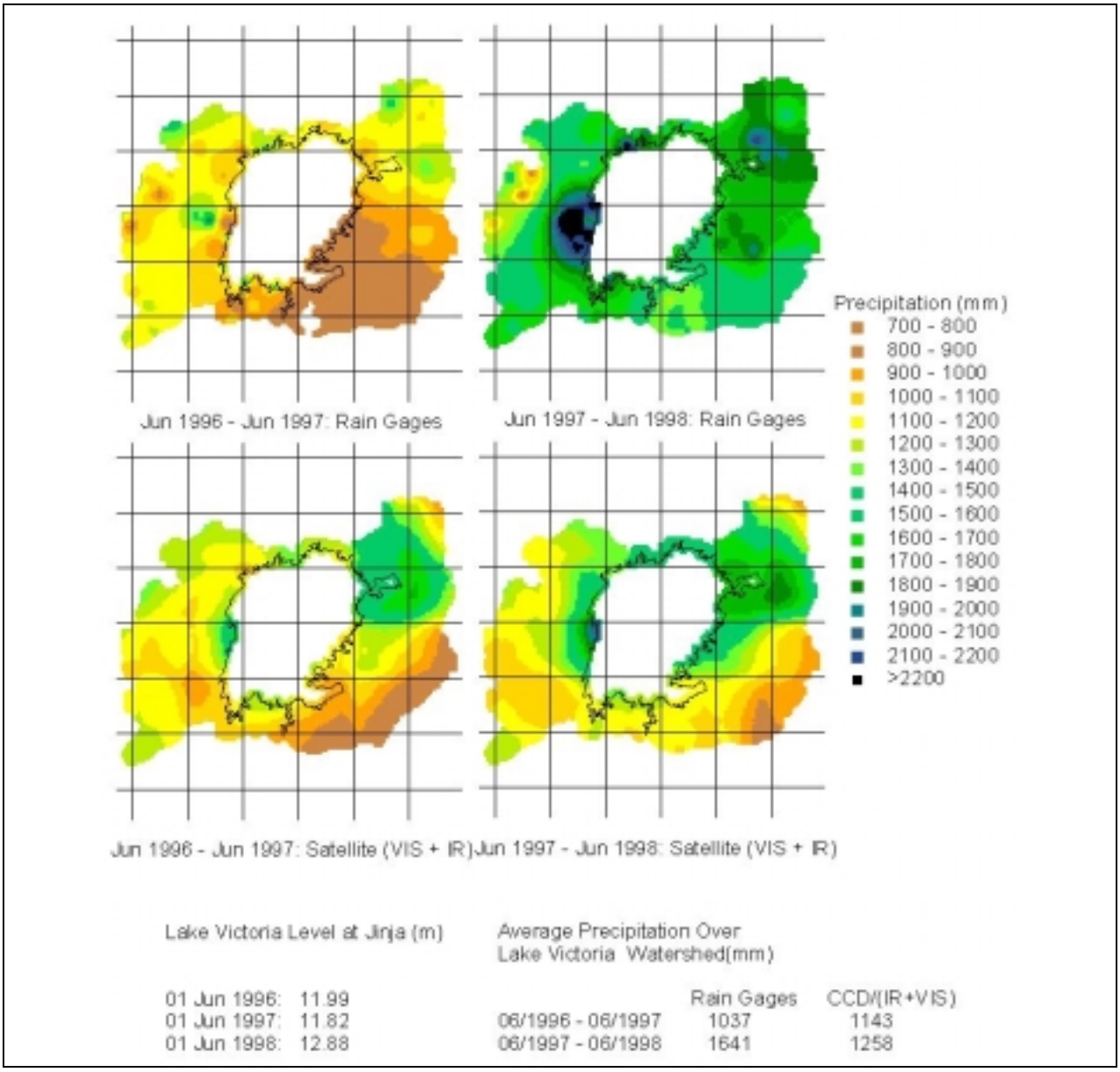
Rain gage information includes available measurements on lake islands and over a band 25 km wide along the lake shore. Using this data, lake precipitation is estimated by an inverse distance square scheme and summed over the lake area and period of interest. Remote sensing precipitation is generated from the CCD/(IR + VIS) procedures at a daily time scale for each lake pixel and again summed over the lake surface and the above-noted periods.

Figure 12.1 and Figure 12.2 summarize the results of the two precipitation estimation methods. For comparison, the results are presented separately for the lake and the lake watershed for the 1996-1997 and 1997-1998 periods

1996-1997 Time Period

Between June 1996 and June 1997, the lake level exhibited a decline of 110 mm indicating that the year is slightly drier than normal. Both the rain gage and the CCD (remote sensing) precipitation estimates indicate that watershed precipitation was low, especially in the southeastern part (Figure 12.1). There is a good agreement between CCD and rain gages estimates over the watershed, with the exception of the northeastern corner, where the CCD estimates higher precipitation than the rain gages. The difference in average precipitation over the watershed is only 10%.

Along the lake shore and over Ukerewe Island the agreement is good, although rain gage measurements are "spotty" (i.e., neighboring rain gages show significant differences). Over the eastern shore, rain gages indicate less precipitation than CCD.



Lake Victoria Level at Jinja (m)		Average Precipitation Over Lake Victoria Watershed(mm)		
			Rain Gages	CCD/(IR+VIS)
01 Jun 1996:	11.99			
01 Jun 1997:	11.82	06/1996 - 06/1997	1037	1143
01 Jun 1998:	12.88	06/1997 - 06/1998	1641	1258

Figure 12.1: Cumulative precipitation over Lake Victoria watershed using the CCD/(IR + VIS) model and rain gage inverse square distance interpolation over Lake Victoria during the hydrologic years June 1996 - June 1997 and June 1997 – June 1998.

Rain gage and CCD precipitation, however, are completely different over the lake area. The rain gages estimate an area of very low precipitation exactly where the CCD shows a maximum. CCD precipitation is more credible than rain gage precipitation because several studies of Lake Victoria precipitation indicate the western shore and the part of the lake immediately east of it as

the wettest in the area. It is therefore very unlikely that the same area during this period becomes the focus of the lowest precipitation in the lake, even less than along the southern shore. Further, precipitation in Ukerewe is higher than in the area immediately to its west.

Over the lake, the CCD precipitation is 1418 mm, while the rain gages are somewhat above 1000 mm. Overall, rain gage precipitation over the lake appears too low for producing a lake drawdown of only 110 mm.

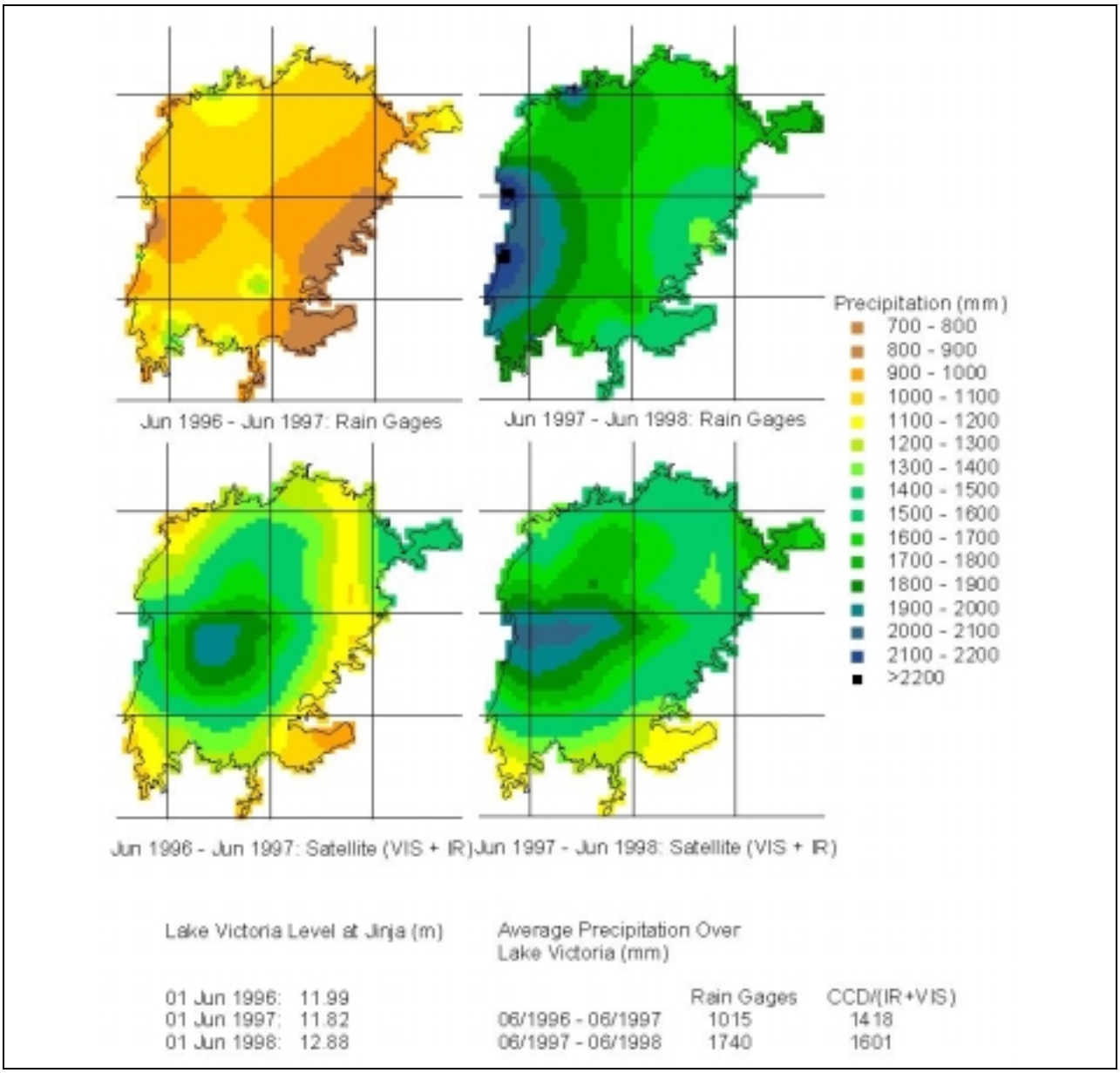


Figure 12.2: Cumulative precipitation over Lake Victoria watershed using the CCD/(IR + VIS) model and rain gage inverse square distance interpolation over Lake Victoria during the hydrologic years June 1996 - June 1997 and June 1997 – June 1998.

1997-1998 Time Period

The period from June 1997 – June 1998 corresponds to a major El Niño event which produced extensive flooding in several parts of the basin. Not surprisingly, the lake level increased by more than one meter during this 12-month period. Rain gage estimates indicate very intense precipitation over most of the watershed, with an average of 1640 mm, which represents a rainfall increase of 60% over the previous year. This precipitation estimate is most likely higher than actual because no data were available for most of the rain gages in Tanzania during 1998. The southern and southwestern sides of the lake are normally the driest in the basin.

CCD precipitation is 10% higher than in the previous year, reaching an average of 1258 mm. The major differences with rain gage precipitation are over Tanzania and Rwanda (for which no precipitation records could be used). Over the lake, however, CCD and rain gage estimations are in average and spatial distribution relatively close (Figure 12.2). The southern area shows a major difference between the two estimates. Even in this case, however, the scarcity of rain gages date during the 1998 short rain in Tanzania is definitely an uncertainty source affecting the rain gage estimates.

The results show clearly that for both periods, rain gage lake precipitation can only be an extension (extrapolation) of shoreline and watershed measurements. However, remote sensing, regional meteorology, and local experience indicate that precipitation patterns over the lake are often distinctly different from those of the lake watershed. Thus, the added value of the remote sensing methods is to better define the relative contribution of lake and watershed precipitation in "closing" the lake water balance. As this study shows, the relative contribution of the two components may be changing significantly from year to year.

13. Conclusions and Further Work Recommendations

A Cold Cloud Duration model using Meteosat infared (IR) digital images to estimate precipitation has been calibrated for the entire Nile River basin. A modification of the CCD model incorporating Meteosat visible (VIS) digital images to better delineate precipitating clouds during daytime has also been calibrated over the entire Nile Basin. The two models have been validated against independent data for most areas of the watershed.

Although the model produces daily rainfall estimates at the pixel level, it has been calibrated and validated for estimating precipitation over 1ºx1º areas for ten day periods. It is therefore strongly suggested to aggregate model outputs over space and time in a way consistent with model calibration.

Model performance is good for most of the Lake Victoria basin, northern Uganda and southern Sudan. Model performance is also good over most of Ethiopia. As shown in Chapter **Error! Reference source not found.**, the model is a valuable tool for better understanding precipitation patterns over the basin and for supporting water balance assessments. Model performance is acceptable in central Sudan, but substantially degrades north of Khartoum.

The remote sensing model can be improved in several ways

Data Availability and Quality

- The most important gaps in 1996-1997 Meteosat data can be filled with data obtained from Eumetsat. The process of replacing the missing images for the northern part of the basin, however, will require around one or two years. Meanwhile, Meteosat images for 1995 can be converted into the Nile DST Grid files to be used for further CCD improvements. Furthermore, Meteosat data can be acquired for a longer time frame (e.g., 1980 to present) create powerful data base for long term assessments.
- Rain gage data availability and their poor quality affect model calibration, especially in regions featuring strong variations in precipitation patterns. Any further improvement of the CCD, or other precipitation estimation procedures, will depend strongly on quality control and expansion of the available data. Data needs exist for Rwanda; Kenya (Lake Victoria basin); Tanzania (the present database does not include all data for 1998 and recent records for some key stations such as Musoma, Bukoba, and Kagondo); Southern Sudan; Northern and western Ethiopia; and Western Sudan. These data needs can only be filled with a concerted effort by the country agencies.

Modeling Improvements

- Seasonal re-calibration of CCD parameters based on 1995 – 1998 data;
- Implementation of more sophisticated remote sensing methods (e.g., the Georgia Tech Remote Sending procedure that also incorporates information from the TRMM (tropical rainfall measuring mission) satellite.

References

De Marchi, C. and A. Georgakakos (2003). *Nile DST Report: Remote Sensing (Beta Version).* Pp. 25. Georgia Water Resources Institute, Atlanta, GA, U.S..

Eumetsat (2000). *The Meteosat System, Revision 4*. Eumetsat Technical Report EUM TD 05, pp. 66. Darmstadt, Germany.

Richards, F. and P. Arkin (1981). On the relationship between satellite observed cloud cover and precipitation. *Mon. Wea. Rev.*, **109**: 1081-1093.