Preliminary Data Compilation for the Nile Basin Decision Support System

FINAL REPORT

March 2010

FutureWater Report 93

Commissioned by
Nile Basin initiative

Authors
Dr. Peter Droogers
Dr. Walter Immerzeel
M. Sc. Belete Berhanu

FutureWater
Costerweg 1G
6702 AA Wageningen
The Netherlands

+31 (0)317 460050

info@futurewater.nl

www.futurewater.nl

Preface

The Nile Basin Decision Support System will provide the necessary knowledge base and analytical tools to support the planning of cooperative joint projects and the management of the shared Nile Basin water resources on an equitable, efficient and sustainable manner. The Water Resources Planning and Management Project (WRPM) is one of the eight projects under the Shared Vision Program of NBI. The WRPM project is preparing the development of the Nile Basin Decision Support System (NB-DSS) in 3 Work Packages:

1. WP1: Software Development and Implementation
2. WP2: Data Compilation and Pilot Test Applications
3. WP3: Supervision and Monitoring

The bulk of the data compilation activities shall be carried out under Work Package 2. However, the project envisages the need to conduct preliminary data collection and compilation. These data collection and compilation is the basis for the subsequent, more extensive, work on data compilation under WP2. A consultancy is undertaken focusing on this preliminary data collection and compilation under the title “Support Data Compilation for the Development of the Nile Basin DSS”. The consultancy has been divided into three phases:

1. Inception phase
2. Analysis phase
3. Synthesis phase

This report describes the results of the third phase of the project (Synthesis phase) and is a result of the following activities:

- Adding additional local data from the Ethiopian Water Master Plan.
- Add attributes required for modeling purposes to the soils, vegetation and land use maps.
- Gap filling of climate and flow time series data for stations at category 1 (main stations) and 2 (local stations).

The consultants wish to acknowledge the support, fruitful discussions and useful comments from NBI staff and in particular: Dr. Abdulkarim Seid, Mr. Ephrem Getahun, Dr. Mekuria Beyene, Dr. Elnaser Abdelwahab.

Table of contents

1 Introduction

1.1 The Nile Basin

The Nile, which is home and source of livelihood to approximately 160 million people, is the longest river in the world having a total length of about 6700 km, traversing an extremely wide band of latitude, from 4-degree south to 32-degree north. The area draining into the Nile river system, about 3 million km^2 extends over 10 African countries.

The two main river systems that feed the Nile are the White Nile, with its sources on the Equatorial Lake Plateau (Burundi, Rwanda, Tanzania, Kenya, Democratic Republic of the Congo and Uganda) and fed by substantial flow from the Baro-Akobo-Sobat system that originates in the foothills of southwest Ethiopia, and the Blue Nile, with its sources in the Ethiopian highlands. Tekeze-Setit-Atbara system contributes to the flow further downstream of Khartoum.

Annual runoff potential of the Nile Basin is estimated to be approximately 85 billion cubic meters (BCM). Such estimates of average annual runoff may vary depending upon the length of records used for the estimation. Compared to other major river basins, the Nile Basin's disparity in water availability differs sharply among sub-basins. Arid portions (about one-third of the area of the Basin) yield negligible flows; whereas, the Highland of Ethiopia, comprising as little as 15–20 percent of the land area of the overall Basin, yields 60–80 percent of the annual flow in the lower Nile.

1.2 The Nile Basin Initiative

The NBI is a partnership initiated and led by the riparian states of the Nile River through the Council of Ministers of Water Affairs of the Nile Basin states (Nile Council of Ministers, or Nile-COM). The NBI seeks to develop the river in a cooperative manner, share substantial socioeconomic benefits, and promote regional peace and security. The NBI started with a participatory process of dialogue among the riparian countries that resulted in their agreeing on a shared vision: to "*achieve sustainable socioeconomic development through the equitable utilization of, and benefit from, the common Nile Basin water resources*".

The NBI involves complex multi-country projects under the two programs, namely, the Shared Vision Program, which deals with projects aimed at creating enabling environment and capacity development for advancing cooperation among riparians, and the Subsidiary Action Programs (SAP), the Eastern Nile and the Nile Equatorial Lakes – Subsidiary Action Programs (EN – and NELSAP), which deal with cooperative investment oriented projects. Considerable progress has been made in the implementation of the various projects under the two NBI programs.

1.3 The Water Resources Planning and Management Project

The water resources planning and management (WRPM) project is one of the eight projects under the Shared Vision Program of NBI. The primary objective of this project is to enhance analytical capacity for a basin-wide perspective to support the development, management, and protection of Nile Basin water resources in an equitable, optimal, integrated, and sustainable manner. The project has the following three technical components:

1. Water Policy Good Practice Guides and Support: The output of this component is to strengthen capacity to formulate and implement effective national policies and strategies for integrated water resources management (IWRM) in Nile Basin countries. This will be a country- and needs-driven component aimed at enabling all basin countries to operate on an equal footing.
2. Project Planning and Management Good Practice Guides and Support: The output of this component is the enhanced capacity in Nile Basin countries for planning and managing multi-country projects, thus contributing to an improved IWRM in the region. These skills will become particularly important as NBI cooperation grows and cooperative investment projects are developed through the subsidiary action programs.
3. Nile Basin Decision Support System: The output of this component is an operational Nile Basin DSS supported by trained staff. The Nile Basin DSS will provide a common, basin-wide platform for communication, information management, and analysis of Nile Basin water resources. Coupled with human resources development and institutional strengthening, the Nile Basin DSS will provide a framework for sharing knowledge, understanding river system behavior, evaluating alternative development and management schemes, and supporting informed decision-making from a regional perspective, thus contributing to sustainable water resources planning and management in the basin.

The WRPM project is managed from the Project Management Unit (PMU) hosted by the Government of Ethiopia and located at Addis Ababa.

1.4 The Nile Basin Decision Support System

The Nile Basin DSS, which is a component of the Water Resources Planning and Management Project, is expected to provide the necessary knowledge base and analytical tools to support the planning of cooperative joint projects and the management of the shared Nile Basin water resources on an equitable, efficient and sustainable manner.

The primary objective of the Nile Basin DSS is to develop a shared knowledge base, analytical capacity, and supporting stakeholder interaction, for cooperative planning and management decision making for the Nile River Basin. An essential feature of the Nile Basin DSS should be that it is an agreed upon tool that will be accepted and used by all riparian countries in the management of the shared Nile water resources.

To support the development and continued use of the Nile Basin DSS, a Nile Basin Regional Decision Support System Center (DSS Center) is established at the Project Management Unit (PMU), in Addis Ababa, Ethiopia. The Regional DSS Center is responsible for developing and operational use of the Nile Basin Decision Support System. The regional DSS Center is supported by national DSS units in every NBI member country.

A recently finished consultancy described the design and requirements for the DSS based on a rigorous process of stakeholder consultations, analysis of available models and technology and an assessment of different areas of concern related to water resources in the riparian countries (Fedra, 2008). The DSS is designed to contain three major components:

- An information management system that provides a common and shared information basis for the planning and decision making processes, locally, sub-regionally, and basin wide, directly accessible for all stakeholders;
- A modular river basin modeling and economic evaluation system built around a dynamic water budget and allocation model, that helps to design and evaluate possible interventions, strategies and projects in response to the problems and challenges identified and prioritized in the stakeholder consultations;
- Tools for a participatory multi-criteria analysis to rank and select alternative compromise solutions for win-win strategies.

The initial first phase of the DSS development is designed to address a basic set of main concerns and priority issues comprising efficient water resources management and allocation, water quality, extreme events (floods and droughts), agriculture, hydropower, and navigation as well as watershed management and erosion, considering simultaneously hydrological, environmental and socio-economic criteria and objectives. This shall be extended to include other areas of concern in subsequent phases.

The DSS shall be implemented at a central location (NBI PMU), two-sub-regional locations (covering the Eastern Nile and the Nile Equatorial Lakes sub-regions) and at the country levels. The data structure and DSS components are identical at all scales, however the contents may vary between the central, sub-regional and national installations of the DSS. Clearly this requires a well designed generic database structure that can deal with a variety of thematic fields, across different spatial and temporal scales and feeding different types of models.

The development of the DSS is organized in three different work packages:

- Work Package 1 (WP1): Software Development and Implementation
- Work Package 2 (WP2): Data Compilation and Pilot Test Applications
- Work Package 3 (WP3): Supervision and Monitoring coordinates/synchronizes these parallel activities and organizes quality assurance processes

The work packaging of the development of the Nile Basin DSS mainly aims at obtaining a high quality product (software + data) in a comparatively short period and delivering the first release of the Nile Basin DSS with the core functionality at an early stage. In order to meet these requirements all relevant data need to be compiled prior to the DSS development insofar as to expedite data compilation activities in WP2 and avoid problems that could arise in data compilation during execution of WP2. The bulk of the data compilation activities shall be carried out under Work Package 2. However, the project envisages the need to conduct preliminary data collection and compilation.

This data collection and compilation is the basis for the subsequent, more extensive, work on data compilation under WP2 and has been divided into three phases:

1. Inception phase
2. Analysis phase
3. Synthesis phase

This report describes the results of the third and final phase of the project (Synthesis phase) and will focus on a description of the spatial and time-series data.

2.1 Land cover

2.1.1 Overview

The GLC2000 land cover / land use data have been compiled, structured and processed as described in the Analysis Report. Details of the dataset can be found in Mayaux et al. (2003). The resulting land cover dataset has the following characteristics:

Columns_and_Rows	2769, 3370
Cellsize__X._Y_	0.011641111, 0.011641111
Extent	
Top	33.4687862225
Left	19.014870712
Right	51.2491077473
Bottom	-5.76175867072
Spatial_Reference	GCS_WGS_1984

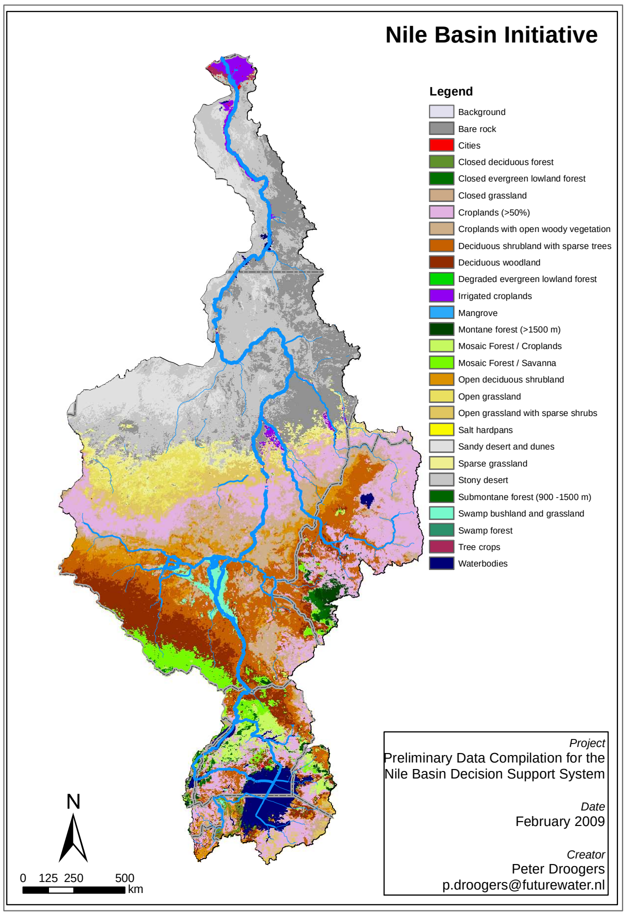

Figure 1. Land cover / land use dataset.

2.1.2 Hydrological attributes

Attributes relevant for hydrological processes as required for modelling purposes are not included in the original GLC2000. Therefore, attributes originating from the commonly applied basin scale model SWAT has been linked to the GLC2000. The following attributes are included in the dataset (full details about these attributed are beyond the scope of this document but can be found in Neitsch et al. (2002)):

ID_NAME	DEFINITION	MIN	MAX
CPNM	Four character code to represent the land cover/plant name.	0.00000	99.00000
IDC	Land Cover/Plant Classification.	1.00000	7.00000
BIO_E	Biomass/Energy Ratio.	10.00000	90.00000
HVSTI	Harvest index.	0.01000	1.25000
BLAI	Max leaf area index.	0.50000	10.00000
FRGRW1	Fraction of the plant growing season corresponding to the 1st. Point on the optimal leaf area development curve.	0.00000	1.00000
LAIMX1	Fraction of the max. leaf area index corresponding to the 1st. point on the optimal leaf area development curve.	0.00000	1.00000
FRGRW2	Fraction of the plant growing season corresponding to the 2nd. point on the optimal leaf area development curve.	0.00000	1.00000
LAIMX2	Fraction of the max. leaf area index corresponding to the 2nd. point on the optimal leaf area development curve.	0.00000	1.00000
DLAI	Fraction of growing season when leaf area starts declining.	0.15000	1.00000
CHTMX	Max canopy height.	0.10000	20.00000
RDMX	Max root depth.	0.00000	3.00000
T_OPT	Optimal temp for plant growth.	11.00000	38.00000
T_BASE	Min temp plant growth.	0.00000	18.00000
CNYLD	Fraction of nitrogen in seed .	0.00150	0.07500
CPYLD	Fraction of phosphorus in seed.	0.00010	0.01500
BN1	Fraction of N in plant at emergence .	0.00400	0.07000
BN2	Fraction of N in plant at 0.5 maturity.	0.00200	0.05000
BN3	Fraction of N in plant at maturity.	0.00100	0.27000
BP1	Fraction of P at emergence.	0.00050	0.01000
BP2	Fraction of P at 0.5 maturity.	0.00020	0.00700
BP3	Fraction of P at maturity.	0.00030	0.00350
WSYF	Lower limit of harvest index.	-0.20000	1.10000
USLE_C	Min value of USLE C factor applicable to the land cover/plant.	0.00100	0.90000
GSI	Max stomatal conductance (in drough condition).	0.00000	5.00000
VPDFR	Vapor pressure deficit corresponding to the fraction maximum stomatal conductance defined by FRGMAX	1.00000	6.00000
FRGMAX	Fraction of maximum stomatla conductance that is achievable at a high vapor pressure deficit.	0.00000	1.00000
WAVP	Rate of decline in radiation use efficiency per unit increase in vapor pressure deficit.	0.00000	50.00000
CO2HI	Elevated CO2 atmospheric concentration.	300.00000	1000.0000 0
BIOEHI	Biomass-energy ratio corresponding to the 2nd. point on the radiation use efficiency curve.	5.00000	100.00000
RSDCO_PL	Plant residue decomposition coefficient.	0.01000	0.09900
Cropname	Crop description name.	0.00000	0.00000
CN2	SCS runoff curve number for moisture condition II.	25.00000	98.00000
OV_N	Manning's "n" value for overland flow.	0.01000	30.00000

2.1.3 Data archiving

The land cover database is stored in the DSS database as one single ArcGis grid file and a DBF attribute file. The join is defined as:

- Target field (in grid file): SWAT_ID
- Join table (in Access): SWAT_LANDCOVER.dbf
- Join field (in Access): ICNUM

The ArcGis grid file and the associated attribute file are stored in:
\DSS\DataBase\LandCover

2.2 Soils

2.2.1 Overview

The Harmonized World Soil Database has been compiled, structured and processed as described in the Analysis Report. Details of the dataset can be found in FAO (2008). The resulting soil database has the following characteristics:

Columns_and_Rows	2280, 4680
Cellsize__X._Y_	0.0083333333, 0.0083333333
Extent	
Top	33.0083333333
Left	21.9999999999
Right	40.9999999999
Bottom	-5.9916666667
Spatial_Reference	GCS_WGS_1984

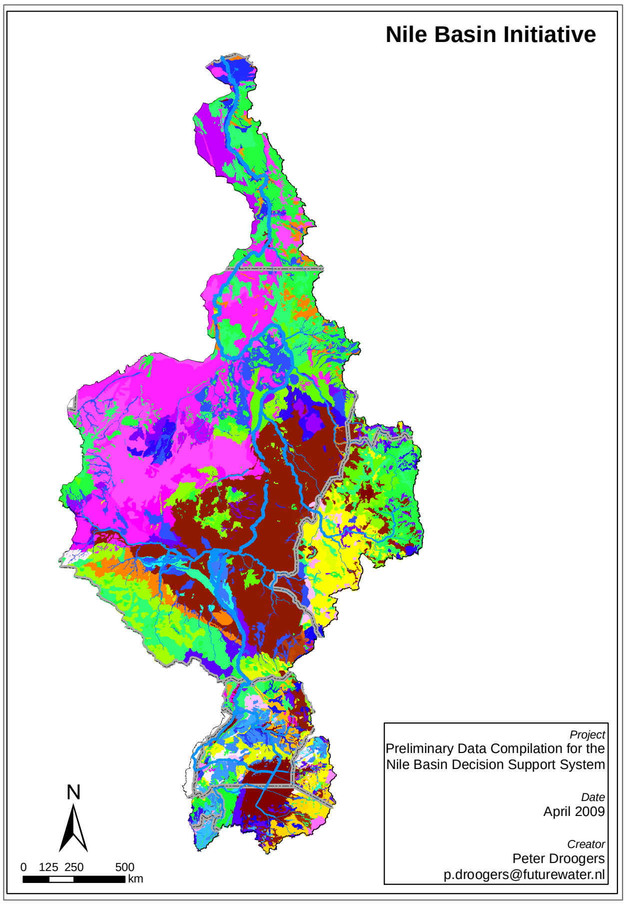

Figure 2. Soils dataset according to Harmonized World Soil Database.

2.2.2 Hydrological attributes

Attributes have been derived as described in the HWSD documentation that are relevant for the hydrological processes as required for modelling purposes. The following attributes are included in the dataset:

Field	**Description**	**Unit**
T_TEXTURE	Topsoil Texture	code
REF_DEPTH	Reference Soil Depth	code
DRAINAGE	Drainage class	code
AWC_CLASS	AWC Range	code
T_GRAVEL	Topsoil Gravel Content	%vol.
T_SAND	Topsoil Sand Fraction	% wt.
T_SILT	Topsoil Silt Fraction	% wt.
T_CLAY	Topsoil Clay Fraction	% wt.
T_USDA_TEX_CLASS	Topsoil USDA Texture Classification	name
T_REF_BULK_DENSITY	Topsoil Reference Bulk Density	kg/dm3
T_OC	Topsoil Organic Carbon	% weight
T_PH_H2O	Topsoil pH (H2O)	log(H +)
T_CEC_CLAY	Topsoil CEC (clay)	cmol/kg
T_CEC_SOIL	Topsoil CEC (soil)	cmol/kg
T_BS	Topsoil Base Saturation	%
T_TEB	Topsoil TEB	cmol/kg
T_CACO3	Topsoil Calcium Carbonate	% weight
T_CASO4	Topsoil Gypsum	% weight
T_ESP	Topsoil Sodicity (ESP)	%
T_ECE	Topsoil Salinity (Elco)	dS/m
S_GRAVEL	Subsoil Gravel Content	%vol.
S_SAND	Subsoil Sand Fraction	% wt.
S_SILT	Subsoil Silt Fraction	% wt.
S_CLAY	Subsoil Clay Fraction	% wt.
S_USDA_TEX_CLASS	Subsoil USDA Texture Classification	name
S_REF_BULK_DENSITY	Subsoil Reference Bulk Density	kg/dm3
S_OC	Subsoil Organic Carbon	% weight
S_PH_H2O	Subsoil pH (H2O)	log(H +)
S_CEC_CLAY	Subsoil CEC (clay)	cmol/kg
S_CEC_SOIL	Subsoil CEC (soil)	cmol/kg
S_BS	Subsoil Base Saturation	%
S_TEB	Subsoil TEB	cmol/kg
S_CACO3	Subsoil Calcium Carbonate	% weight
S_CASO4	Subsoil Gypsum	% weight
S_ESP	Subsoil Sodicity (ESP)	%
S_ECE	Subsoil Salinity (Elco)	dS/m

2.2.3 Data archiving

The soils database is stored in the DSS database as one single ArcGis grid file and an MS-Access attribute file. The join is defined as:

- Target field (in grid file): Value
- Join table (in Access): HWSD_DATA
- Join field (in Access): MU_GLOBAL

The ArcGis grid file and the attribute file can be found in:
\DSS\database\soils

2.3 Vegetation Indices

2.3.1 Overview

The vegetation indices as observed by the MODIS sensor has been compiled, structured and processed as described in the Analysis Report. Details and background of the dataset can be found in https://lpdaac.usgs.gov/lpdaac/products/modis_products_table/vegetation_indices/monthly_l3_global_1km/v5/terra). The resulting vegetation indices database has the following characteristics:

Vegetation Indices	NDVI, EVI, NIR, MIR
Columns_and_Rows	1200, 1200
Cellsize__X._Y_	926 x 926 m
Extent	
Top	4447802
Left	2223901
Right	3335852
Bottom	3335852
Spatial_Reference	Sinusoidal
Temporal resolution	1 month
Time covered	2001..2008

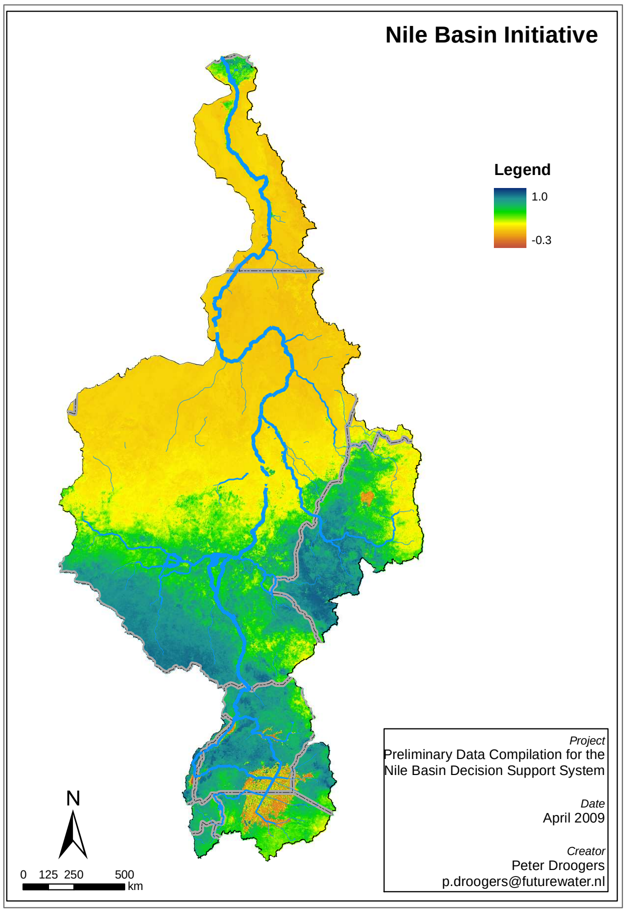

Figure 3. MODIS NDVI (Jun-2008).

2.3.2 Hydrological attributes

Attributes relevant for hydrological processes as required for modelling purposes are included in the dataset. Since final model selection has not been defined yet a wide range of attributes are included in the dataset:

- NDVI : Normalized Difference Vegetation Index
- EVI: Enhanced Vegetation Index
- RR: Red Reflectance
- NIR: Near Infrared Reflectance
- BR: Blue Reflectance
- MIR: Medium Infrared Reflectance

2.3.3 Data archiving

The ArcGis HDF files are stored in:
\DSS\DataBase\NDVI

Total file size of all vegetation data is about 17 GB (20 MB * 9 tiles * 8 years * 12 months). Data are stored in standard compressed HDF format, which can be used directly in ArcGis. Data have been stored in folders for each month (e.g. 2008.06.01 for June 2008). Each folder contains the 9 tiles (h20v05; h20v06; h20v07; h20v08; h20v09; h21v06; h21v07; h21v08; h21v09).

To reduce total storage all information is stored in integers. In order to convert to real values, all data should be multiplied by 0.0001.

2.4 Precipitation FEWS

2.4.1 Overview

The precipitation data according to FEWS-RFE has been compiled, structured and processed as described in the Analysis Report. Details of the dataset can be found in http://earlywarning.usgs.gov/adds/overview.php. The resulting database has the following characteristics:

Columns_and_Rows	751, 801
Cellsize__X._Y_	0.1, 0.1
Extent	
Top	40.05
Left	-20.05
Right	55.05
Bottom	-40.05
Spatial_Reference	GCS_Sphere_Clarke_1866_Authalic
Temporal resolution	1 day
Time covered	2001..2008

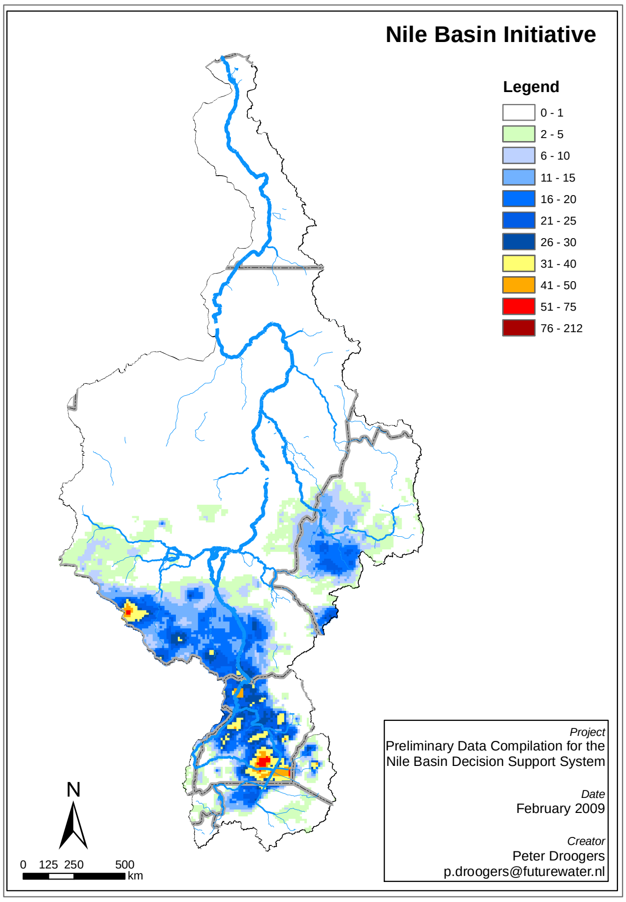

Figure 4. Precipitation according to FEWS in mm d^{-1} (1-Jun-2008).

2.4.2 Data archiving

The ArcGis grid files are stored in:
\DSS\DataBase\FEWS

Given the large number of files (365 days * 8 years). Data have been stored in compressed files for each year (2000 to 2008):

- Annual collection of daily rainfall as YYYY.zip , where YYYY is the 4 digit year.
- Daily grids as as YYYYMMDD.

2.5 Time Series Climate Data

2.5.1 Overview

The climatic parameters like precipitation, temperature, and evaporation at a place are a random event and when it arranged in time represent a time series. The point climatic data with different time scale are compiled in the PostGreSQL database. As it discussed in the Analysis report climatic data accessed with different local and global sources. From these sources a total of 20,662,403 climatic data records compiled in the database.

2.5.2 Precipitation

Precipitation is essential climatic parameter for any water resource project and so it has paramount role in the DSS. So a total of 15,289,931 precipitation data have been compiled from five global and local sources, which include NileDST, Ethiopian master plan study, GSOD, RVIDIS, SMMDSS and RUSUMO. Out of them Nile DST has the highest share of the data sources. These precipitation data are compiled in three forms as daily, monthly and mean monthly form with 15,246,972, 40703, and 2256 records respectively.

2.5.3 Temperature

In the database the second most essential climatic parameter, that is temperature, compiled in different form as min and max temperature, Dry and wet bulb temperature, soil temperature in different depth and mean temperature. A total of 2,805,292 temperature data compiled from different sources.

2.5.4 Evaporation

Similarly Evaporation also compiled in different form as external and Internal Evapometer -Evaporation, pan Type A Evaporation, mean daily Evaporation, and highest daily Evaporation with 15727, 73176, 79228, 1068, and 48 records respectively.

2.5.5 Other climatic parameter

In the database other climatic parameters that are essential for different climatically analysis also compiled. These parameters include atmospheric pressure, radiation, relative humidity, sunshine hours, wind speed, and wind direction. In general 2,397,933 records covered by these climatic parameters.

2.6 Time Series Hydrological Data

2.6.1 Overview

The hydrological data that mainly include the different form of flow and sediment data also have been compiled in the PostGerSQL. As it discussed in the above section the hydrological data also accessed from different local and global sources. A total of 5,091,260 records are covered in the data table of the PostGreSQL database.

2.6.2 Data Compiled

Actually the large share of the hydrological data that compiled in the database is covered by different forms of flow data. And the sediment data holds the smallest portions of it. In general the compiled hydrological data with their records is presented in the following table.

Parameter	Records
Discharge	323,717
Water Level	4,759,152
Annual Max Flood	140
Maximum Discharge	965
Minimum Discharge	964
Sediment Load (Discharge)	2,673
Sediment Flow	3,649

3 Metadata and data integrity

The term “metadata” is used in various ways and often vaguely specified as “data that provides information about other data” (Webster) or in short “data about data”. Since metadata crosses several fields of specialization there is no generally accepted model to describe types of metadata. Data integrity is used as term to describe the quality of individual records as well as the entire database. The term data integrity is therefore related to redundant and conflicting data.

According to Wikipedia relational databases has their own mechanisms for storing metadata. Examples of relational-database metadata include:

- Tables of all tables in a database, their names, sizes and number of rows in each table.
- Tables of columns in each database, what tables they are used in, and the type of data stored in each column.

The degree to which metadata is captured is referred to as its granularity. Metadata with a high granularity contains more information and enables greater levels of technical manipulation however, a lower level of granularity means that metadata can be created for considerably lower costs but will not provide as detailed information.

The World Meteorological Organisation (WMO, 2009) specifies metadata as “Metadata can be extensive and all-inclusive, or it can be specific to a more limited function. Metadata are used to provide documentation for data sets or services. In essence, metadata answers who, what, when, where, why, and how about every data that are being documented”. WMO advises to create a single XML file for each metadata record, that is, one XML document describes one data set.

However, the CUAHSI (Consortium of Universities for the Advancement of Hydrologic Science; www.cuahsi.org) initiative focuses on much more advanced systems. They mention that several metadata standards have been developed during the past few years of which the most important ones are: International Standards Organization Technical Committee 211, 19115 norm, the Dublin Core Metadata Initiative, the Federal Geographic Data Committee, the Global Change Master Directory, and the EOSDIS Core System. Currently, the various metadata standards are converging towards compliance with the ISO 19115 norm. CUAHSI finally concludes that defining and implementing appropriate metadata standards is important and that no standardized system exists and will never be developed as this is application specific.

Probably the best source for an in-depth discussion regarding database can be found in the book: “Hydroinformatics: Data Integrative Approaches in Computation, Analysis, and Modeling”, by Kumar et al. Especially Chapter 4 by Michael Piasecki is completely devoted to metadata. This book is attached as separate appendix to this report.

Hydrologic Metadata

Michael Piasecki

Contents

The database as compiled during this study has specific characteristics with special relevance for metadata and data integrity. The far most important one is that most discussions regarding metadata and integrity are relevant to observational issues. However, data collected for the NBI-DSS are all based on existing databases that have undergone already data quality control at various levels.

A major concern in hydro-meteorological data is the frequently occurring gaps. These gaps can occur by a broad spectrum of reasons such as: instrumental failure, operational problems, computer crashes, lost papers, non-accessibility of data. It is therefore that a the scientific literature is focusing on this aspect. However, there is no single acceptable method to fill data gaps in general as it all depends on issues like: can the stations being checked, distance between stations, is database sufficient small to undertake visual inspection. In principle there are two procedures for filling data gaps:

- Use observations from the nearest station to fill gap (location-infill)
- Use observation from the same station of the same season to fill gaps (time-infill).

The first method, replacing missing data from the nearest station, is the most applied one as it is straight forward and easy to implement. There is however a major restriction to use this method for the NBI-DSS data: stations should be within 25 km distance and 50 m height difference (Klok et al., 2006). If this approach would be followed, almost no gap filling should be possible as stations are often further apart or at different heights.

Since the location-infill is not an option, the time-infill method might be considered in the DSS time-series data. The time-infill has no standardized procedure and options often seen are:

- Fill with the value of the previous day, in case single days are missing.
- Fill with the mean of the previous and the next day, in case single days are missing.
- Fill with the mean of the particular month, in case more days are missing.
- Fill with the data from the previous year, in case more days are missing.
- Correlate missing parameter with correlated parameter (e.g. mean temperature with maximum temperature).
- Apply a hydrological model to fill hydrological data gaps.

All of the above methods depend on: (i) for what purpose is the data going being used, (ii) how severe are the data gaps. The first point, purpose of the data, should be considered in the use of the data in the DSS. Climate data is required for forcing the models in the DSS, while hydrological data is going to be used for calibration and validation of the models and gap filling is therefore not critical. Even more important is that gap-filled records should never be used for calibration and validation of models.

In the context of this study it is advisable to use metadata at two levels of granularity. The first level is a description of the activities of data added to the database. This is a relatively straight forward table including the following topics:

- datasetTitle
- datasetResponsibleParty
- datasetTopicCategory
- datasetRecords
- datasetLocation
- datasetUploadDate
- datasetUploadContact
- spatialResolution
- ReferenceSystem

This metadata is shown in Appendix 1.

The second level of metadata is a dynamic one that is automatically extracted from the database by Queries or Views. A typical example of this second level of metadata is shown in Append 2 and might have the following topics:

- series_id
- station_name
- country_name
- parameter_name
- FirstYear
- LastYear
- NrRecords
- Lowest
- Highest

Additional work on metadata and data integrity is foreseen in WP2 of the NBI-DSS project.

4 Conclusion

The Water Resources Planning and Management Project of the Nile Basin Initiative has as primary objective to enhance analytical capacity for a basin-wide perspective to support the development, management, and protection of Nile Basin water resources in an equitable, optimal, integrated, and sustainable manner. The Nile Basin Decision Support System (NBI-DSS) is one of the activities to fulfill this primary objective.

The Water Resources Planning and Management Project (WRPM) is one of the eight projects under the Shared Vision Program of NBI. The WRPM project is preparing the development of the Nile Basin Decision Support System (NB-DSS) in 3 Work Packages:

- WP1: Software Development and Implementation
- WP2: Data Compilation and Pilot Test Applications
- WP3: Supervision and Monitoring

The bulk of the data compilation activities will be carried out under Work Package 2. However, the project envisages the need to conduct preliminary data collection and compilation. These data collection and compilation is the basis for the subsequent, more extensive, work on data compilation under WP2. This preliminary data collection task has been undertaken in the period from January 2009 to March 2010 and will form a solid base for subsequent work.

The activities are described in three reports (Inception, Analysis en Final) and two types of databases have been developed:

- A spatial database including relevant hydrological properties (attributes). A total of five different type of spatial data have been included, each which various time frames. A total of over 3000 spatial layers are included in the database.
- A time-series database developed in PostGreSQL including over 25 million records with climatic and hydrological data.

This work will be used for the additional activities as described under Work Package 2 of the NBI-DSS development "Data Compilation and Pilot Test Applications":

- Data compilation and pilot test applications will be executed in parallel but independently, involving end users, the national DSS units and the DSS team at the PMU.
- The objectives of this work package include both a support function and a test function:
    - providing quality assured data for software testing and applications;
    - usability testing of the IMS and data analysis functions;
    - pilot test applications (system usability);
    - model calibration and validation.
- The tasks of WP2 are designed to test the operation and usability of the integrated system and its components together with the data in realistic applications of immediate basin wide interest, including specifically
    - Preparatory phase, data catalog and meta data, pilot study definitions;
    - Data compilation, processing and import;
    - Usability testing of all components;
    - Pilot test cases and model calibration/validation;
    - Stakeholder, user and core team training (data analysis, application building).
- Final system certification eventually by external auditors.

5 References

Andrighetti, M. G. Rampanelli, D. Zardi. 2007. Reconstruction and climatological analysis of the temperature series of Verona (1741 – 2006). Proceedings/ICAM2007.

Brunetti, M., Maurizio Maugeri, Fabio Monti, Teresa Nanni. 2006. Temperature and precipitation variability in Italy in the last two centuries from homogenised instrumental timeseries. Int. J. Climatol. 26: 345–381.

FAO/IIASA/ISRIC/ISSCAS/JRC, 2008. Harmonized World Soil Database (version 1.0). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Klok, L., Albert Klein Tank, Michael Begert, Evelyn Zenklusen. 2006. Assessment of the available station density for the gridding and daily data quality/homogeneity.

Kumar, P., Mike Folk, Momcilo Markus, J.C. Alameda. 2005. Hydroinformatics: Data Integrative Approaches in Computation, Analysis, and Modeling. ISBN: 9780849328947

Mayaux, P. et al., 2003. A land cover map of Africa. European Commission Joint Research Center.

Neitsch, S.L., J.G. Arnold, J.R. Kiniry, J.R. Williams, K.W. King. 2002. Soil and Water Assessment Tool, Theoretical documentation. Version 2000.

Ramos-Calzado, P., J. Gómez-Camacho, F. Pérez-Bernal, M. F. Pita-López. 2008. A novel approach to precipitation series completion in climatological datasets: application to Andalusia. International Journal of Climatology. Volume 28 Issue 11, Pages 1525 – 1534.

Stockli. R. LBA-MIP driver data gap filling algorithms. 2007. Department of Atmospheric Science. Colorado State University.

WMO. 2009. World Meteorological Organisation Information System. http://www.wmo.int

World Meteorological Organization: Guidance on Metadata and homogenization, WCDMP-No. 53, WMO-TD No. 1186, 1045, WMO, Geneva, 2004.

Appendix 1: Metadata level 1

datasetTitle	datasetResponsibleParty	datasetTopicCategory	datasetRecords	datasetLocation	datasetUploadDate	datasetUpload(	spatialResolu	ReferenceSyste
GRDC_Stations	Global Runoff Data Center	Stations	77	PostGreSQL	16/Feb/2009	Peter Droogers	N/A	LatLon
RivDis_Stations	RivDis	Stations	2	PostGreSQL	16/Feb/2009	Peter Droogers	N/A	LatLon
LEGOS_Stations	TOPEX/POSEIDON/JASON	Stations	32	PostGreSQL	16/Feb/2009	Peter Droogers	N/A	LatLon
DST-Climate_Stations	FAO Nile-DST	Stations	2,333	PostGreSQL	16/Feb/2009	Peter Droogers	N/A	LatLon
DST-Flow_Stations	FAO Nile-DST	Stations	529	PostGreSQL	16/Feb/2009	Peter Droogers	N/A	LatLon
GSOD	Global Summary of Day Climate	Stations	215	PostGreSQL	16/Feb/2009	Peter Droogers	N/A	LatLon
GRDC day	Global Runoff Data Center	Flow	14,581	PostGreSQL	12/Mar/2009	Peter Droogers	N/A	LatLon
GRDC month	Global Runoff Data Center	Flow	16,536	PostGreSQL	12/Mar/2009	Peter Droogers	N/A	LatLon
RivDis	RivDis	Flow	92	PostGreSQL	12/Mar/2009	Peter Droogers	N/A	LatLon
LEGOS	Laboratoire d'Etudes en Géophys	Flow	2,068	PostGreSQL	12/Mar/2009	Peter Droogers	N/A	LatLon
DST-Discharge	FAO Nile-DST	Flow	115,151	PostGreSQL	12/Mar/2009	Peter Droogers	N/A	LatLon
DST-GaugeHeight	FAO Nile-DST	Flow	2,889,463	PostGreSQL	12/Mar/2009	Peter Droogers	N/A	LatLon
DST-Precipitation_01	FAO Nile-DST	Climate	9,766,774	PostGreSQL	16/Mar/2009	Peter Droogers	N/A	LatLon
DST-Precipitation_01	FAO Nile-DST	Climate	843,326	PostGreSQL	16/Mar/2009	Peter Droogers	N/A	LatLon
GSOD Precipitation	Global Summary of Day Climate	Climate	120,806	PostGreSQL	16/Mar/2009	Peter Droogers	N/A	LatLon
GSOD Tavg	Global Summary of Day Climate	Climate	122,897	PostGreSQL	16/Mar/2009	Peter Droogers	N/A	LatLon
GSOD Tmin	Global Summary of Day Climate	Climate	122,639	PostGreSQL	16/Mar/2009	Peter Droogers	N/A	LatLon
GSOD Tmax	Global Summary of Day Climate	Climate	122,691	PostGreSQL	16/Mar/2009	Peter Droogers	N/A	LatLon
SRTM	NASA	DEM	N/A	\Spatial\SRTM	25/Mar/2009	Peter Droogers	90 m	Sinusoidal
HYDRO1K	USGS	DEM	N/A	\Spatial\HYDRO1K	25/Mar/2009	Peter Droogers	1000 m	LAEA
NileOutline	NBI	Background	N/A	\Spatial\Background	25/Mar/2009	Peter Droogers	Shape	LatLon
Nile Rivers	USGS	Background	N/A	\Spatial\Background	25/Mar/2009	Peter Droogers	Shape	LatLon
Nile Countries	ESRI	Background	N/A	\Spatial\Background	25/Mar/2009	Peter Droogers	Shape	LatLon
HWSD	Harmonized Word Soil Database	Soils	N/A	\Spatial\Soil	25/Mar/2009	Peter Droogers	1000 m	LatLon
GLC2000	EU JRC	Vegetation	N/A	\Spatial\LC	30/Mar/2009	Peter Droogers	1000 m	Sinusoidal
NDVI	MODIS	Vegetation	N/A	\Spatial\LC	30/Mar/2009	Peter Droogers	1000 m	Sinusoidal
LAI	MODIS	Vegetation	N/A	\Spatial\LC	30/Mar/2009	Peter Droogers	1000 m	Sinusoidal
NPP	MODIS	Vegetation	N/A	\Spatial\LC	30/Mar/2009	Peter Droogers	1000 m	Sinusoidal
TRMM-rainfall	NASA	Climate	N/A	\Spatial\TRMM	30/Mar/2009	Peter Droogers	25 km	LatLon
CRU-rainfall	University of East Anglia	Climate	N/A	\Spatial\CRU	30/Mar/2009	Peter Droogers	50 km	LatLon
CRU-temperature	University of East Anglia	Climate	N/A	\Spatial\CRU	30/Mar/2009	Peter Droogers	50 km	LatLon
ENTRO_OSI	ENTRO	Climate	17,808	PostGreSQL	12/Dec/2009	Belete Berhanu	N/A	LatLon
ENTRO_OSI	ENTRO	Flow	9,540	PostGreSQL	12/Dec/2009	Belete Berhanu	N/A	LatLon
SMMDSS Docs	SMMDSS	Rainfall	526,866	PostGreSQL	12/Dec/2009	Belete Berhanu	N/A	LatLon
SMMDSS Docs	SMMDSS	Flow	112,495	PostGreSQL	12/Dec/2009	Belete Berhanu	N/A	LatLon
RUSUMO	RUSUMO	Rainfall	955,407	PostGreSQL	12/Dec/2009	Belete Berhanu	N/A	LatLon
RUSUMO	RUSUMO	Flow	280,630	PostGreSQL	12/Dec/2009	Belete Berhanu	N/A	LatLon

Appendix 2: Metadata level 2

Example of metadata at granularity level 2

series_id	station_name	country_name	parameter_name	FirstYear	LastYear	NrRecord:	Lowest	Highest
1	Arua Met Station	Uganda	Precipitation	1922	1999	18535	0	110
2	Koboko St. Charles L	Uganda	Precipitation	1942	1980	10923	0	115.3
3	Abi Estate	Uganda	Precipitation	1949	1979	10286	0	148
4	Olovu	Uganda	Precipitation	1957	1980	8095	0	125
5	Lokiragodo	Uganda	Precipitation	1957	1980	8310	0	114.8
6	Manibe Omuazire	Uganda	Precipitation	1957	1968	4049	0	96.8
7	Ovujo	Uganda	Precipitation	1957	1978	7578	0	91.7
8	Arua Central Govt Prison	Uganda	Precipitation	1964	1994	8070	0	104.1
9	Wandi BAT Uganda Ltd	Uganda	Precipitation	1967	1980	4322	0	98
10	L. Victoria at Entebbe	Uganda	Water Level	1948	2001	37234	10.04	12.87
11	Yumbe Hospital (Aringa)	Uganda	Precipitation	1938	1979	13998	0	111.8
12	Adjumani Dispensary	Uganda	Precipitation	1942	1978	11937	0	132.8
13	Ladonga VFM	Uganda	Precipitation	1938	1957	5483	0	101.6
14	Laropi Dispensary	Uganda	Precipitation	1939	1980	14060	0	127.5
15	Moyo Boma	Uganda	Precipitation	1938	1998	15308	0	118.4
16	Terego Dispensary	Uganda	Precipitation	1938	1966	9648	0	151.1
17	Obongi Dispensary	Uganda	Precipitation	1939	1979	13817	0	119.4
18	Udupe Dispensary	Uganda	Precipitation	1942	1962	7457	0	98.6
19	Zaipi Dispensary	Uganda	Precipitation	1942	1980	11989	0	135.9
20	Pakelli Dispensary	Uganda	Precipitation	1943	1980	13026	0	109
21	L. Victoria at Jinja	Uganda	Water Level	1948	2000	37177	10.54	13.41
22	Otrevu	Uganda	Precipitation	1948	1973	8431	0	152
23	Utumbari	Uganda	Precipitation	1957	1980	8555	0	123.7
24	Ivu	Uganda	Precipitation	1957	1980	7668	0	101.6
25	Moyo Senior Secondary School	Uganda	Precipitation	1967	1998	4802	0	130.5
26	Mount Kei Forest Station	Uganda	Precipitation	1968	1978	3805	0	107.2
27	Adjumani Prisons Farm	Uganda	Precipitation	1968	1982	3834	0	150.2
28	Bileafe Tobacco Station	Uganda	Precipitation	1971	1980	2862	0	89.6
29	Kitgum Centre VT	Uganda	Precipitation	1914	2001	28658	0	153.1
30	Atiak Dispensary.	Uganda	Precipitation	1942	1977	9252	0	123
31	Palabek Divisional Hqs	Uganda	Precipitation	1939	1981	13721	0	140
32	R. Kakinga Index Catchment	Uganda	Discharge	1969	1998	19	0.01	9.75
33	R. Kakinga Index Catchment	Uganda	Water Level	1968	2000	9021	0	2.99
34	Padibe	Uganda	Precipitation	1942	1983	13568	0	123.2
35	Patiko	Uganda	Precipitation	1965	1985	7151	0	122
36	Aringa Valley Coffee	Uganda	Precipitation	1967	1983	5446	0	99.1
37	Acholi Ranch	Uganda	Precipitation	1970	1985	5510	0	140.2
38	Kitgum Matidi	Uganda	Precipitation	1943	1982	12263	0	108.5
39	Kalongo Hospital	Uganda	Precipitation	1956	1981	8710	0	99.9
40	Paimol	Uganda	Precipitation	1942	1980	11838	0	101.8
41	Agoro	Uganda	Precipitation	1942	1984	13422	0	114.3
42	Orom	Uganda	Precipitation	1943	1983	13293	0	149.4
43	Karenga	Uganda	Precipitation	1952	1977	8002	0	128.5
44	R. Ruizi at Mbarara water works	Uganda	Water Level	1954	2000	31869	0.96	5.18
45	Naam	Uganda	Precipitation	1942	1983	12781	0	129.5
46	Mucwini Gombolola	Uganda	Precipitation	1963	1978	5543	0	104.1
47	Madi Opei	Uganda	Precipitation	1965	1998	4959	0	113
48	Kacheri	Uganda	Precipitation	1977	1991	2858	0	90
49	Kotido PWD	Uganda	Precipitation	1947	1980	10954	0	101.9
50	Kaabong	Uganda	Precipitation	1946	1966	7396	0	133
51	Kotido	Uganda	Precipitation	1947	2001	13049	0	101.8
52	Loyoro [County Dodoth]	Uganda	Precipitation	1947	1963	5965	0	127
53	Kuluva	Uganda	Precipitation	1957	1977	7275	0	93.4
54	Payidha	Uganda	Precipitation	1942	1982	9373	0	110.7
55	R. Nyakizumba at Maziba	Uganda	Discharge	1956	2000	203	1.05	23.36
56	R. Nyakizumba at Maziba	Uganda	Water Level	1956	2000	27601	0.91	3.09
57	Vurra Customs	Uganda	Precipitation	1942	1977	9071	0	157
58	Warr Dispensary	Uganda	Precipitation	1942	1965	8368	0	111.8
59	Nyapea St.Aloysius	Uganda	Precipitation	1943	1997	8338	0	109.7
60	Awung	Uganda	Precipitation	1957	1971	4958	0	77
61	Usi Forest Station	Uganda	Precipitation	1957	1979	8066	0	140
62	Lendu Forest Station	Uganda	Precipitation	1957	1978	7818	0	125
63	Okavu Forest Station	Uganda	Precipitation	1971	1981	3500	0	80
64	Nyara TWGCS	Uganda	Precipitation	1972	1980	3016	0	93.2
65	Buliisa Gomborora Hqs	Uganda	Precipitation	1943	1977	5144	0	127
66	Okollo Dispensary	Uganda	Precipitation	1943	1979	9161	0	110.8
67	R. katonga at Kampala-Masaka Roa	Uganda	Discharge	1965	2000	93	0.36	35.73
68	R. katonga at Kampala-Masaka Roa	Uganda	Water Level	1965	2000	13118	6.2	7.72
69	Pakwach Dispensary	Uganda	Precipitation	1943	1980	9158	0	93.2
70	Rhino Camp Dispensary	Uganda	Precipitation	1942	1978	10832	0	172.7
71	Goli African.Inland.	Uganda	Precipitation	1943	1964	7854	0	130.3
72	Nebbi UTC	Uganda	Precipitation	1943	2001	16819	0	110
73	Anaka	Uganda	Precipitation	1948	1982	7271	0	106