Assimilation of remote sensed soil moisture data into a hydrological model using the Ensemble Kalman Filter

A West Africa case study

By: Jaïr Smits

Delft, May 2008

Graduation Committee:
Prof. Dr. Ir. N.C. van de Giesen, TU Delft, Water Resource Management
Prof. Dr. Ir. A.W. Heemink, TU Delft, Applied Mathematical Analysis
Ir. H.C. Winsemius, TU Delft, Water Resource Management

TU Delft
Delft University of Technology
Assimilation of remote sensed soil moisture data into a hydrological model using the Ensemble Kalman Filter

A West Africa case study

By: Jaïr Smits

Delft, May 2008

Graduation Committee:
Prof. Dr. Ir. N.C. van de Giesen, TU Delft, Water Resource Management
Prof. Dr. Ir. A.W. Heemink, TU Delft, Applied Mathematical Analysis
Ir. H.C. Winsemius, TU Delft, Water Resource Management
Preface

This is my MSc thesis report. The MSc thesis is credited for 42 ECTS points and forms the final part of the master Water Resource Management. I started my thesis in September 2007. This is a thesis on the hydrological processes and spatial features that influence the accuracy of the ERS-satellite estimated soil moisture states. The study focusses on West Africa.

During my thesis I received some valuable help from several people. I like to thank a few people in particular.

In my thesis I made use of a conceptual hydrological model. This model was built by Rens van Beek, who is a lecturer and researcher of the University of Utrecht. He provided me with the model data and helped me to understand how to use all the files and since the model was written in PCRaster, this was not a very easy task. For his support I thank Rens van Beek.

The model was fed by actual evaporation data which was provided by EARS. I thank Marjolein De Weirdt, Steven Foppen and Andries Rosema from EARS very much for this data and the information and help they have given me.

Also, I like to thank Susan Steele who was of great support during the implementation of the Ensemble Kalman Filter.

Furthermore I thank my graduation committee. Each member was very helpful throughout the completion of my thesis. At almost any given moment I could come by their office to discuss a matter or two, this I find very special. I would like to especially thank Martine Rutten, my daily supervisor, for her extraordinary help and support.

Lastly, I would also like to acknowledge my family for their support and encouragement during the completion of my thesis.

I wish you a pleasant reading.

Jaïr Smits

Delft, May 2008
Abstract

Soil moisture plays a very important role in environmental processes. By knowing soil moisture states, flood- and drought predictions can be improved and more insight into climate change can be obtained. The ERS-2 satellite has a scatterometer on board which can be used for moisture state estimation for large scale areas. However, some discussions exist on how to interpret the satellite measurements.

This thesis aims to determine by what spatial features and hydrological processes the accuracy of the ERS-satellite estimated soil moisture states is influenced. The study focuses on West Africa with data of 2003-2006.

The ERS-satellite measurements have been compared with model results from a conceptual hydrological model called PCR-GLOBWB. The model requires precipitation and actual evaporation as input. Actual evaporation data is provided by EARS (http://www.ears.nl) and precipitation data is obtained from the TRMM-satellite. Thereafter the ERS-satellite measurements are assimilated into the model using the Ensemble Kalman Filter (EnKF). Ensembles of model states were generated by perturbing four variables used by the model: Actual evaporation, precipitation, saturated hydraulic conductivity and the thickness of the first soil layer.

The ERS-satellite measurements correlate well with the upper layers of the model, which have an average thickness of ~0.20 m and ~1.00 m respectively. The correlation is the highest for the first store and during the wet summer periods.

The application of the EnKF does not lead to better results. The difference between the ERS-satellite measurements and the assimilated model states is larger than the difference between satellite measurements and the original -non assimilated- model states. It is thought that this is mainly due to the used method to create the model ensembles. The method used to create the ensemble members for precipitation introduces a bias because the perturbations leading to negative precipitation are set to zero. Only those perturbations leading to more positive precipitation are kept, which lead to wetter soil conditions. Updates are made in the assimilation process but remain small. This is due to the large uncertainty of the ERS-satellite measurements compared to the model uncertainty which is based on the perturbations of the four variables.

To improve the results of the data assimilation, it is recommended to choose a simpler method to create the model ensembles. The ensembles must be created in such a way that no bias and unrealistic values are created.
Samenvatting

Bodemvocht speelt een belangrijke rol in veel hydrologische en klimatologische processen. Als het bodemvochtgehalte bekend is, dan kunnen droogte- en overstroom voorspellingen worden verbeterd en wordt meer inzicht verkregen in klimaatverandering. De ERS-2 satelliet is in staat om voor grote gebieden de bodemvocht te schatten. Er is echter nogal wat discussie over hoe de metingen geïnterpreteerd moeten worden.

Dit afstudeerwerk heeft tot doel om te bepalen door welke ruimtelijke objecten en hydrologische processen de nauwkeurigheid van de door de ERS-satelliet geschatte bodemvochtigheidsgraad wordt bepaald. De studie is uitgevoerd voor West Afrika met data van 2003-2006.

De metingen van de ERS-satelliet zijn vergeleken met resultaten van PCR-GLOBWB, een conceptueel hydrologisch model. Als invoer gebruikt het model de actuele verdamping data van EARS (http://www.ears.nl) en de neerslag data afkomstig van de TRMM-satelliet. Hierna zijn de ERS-satellietmetingen gebruikt om het model te updaten met behulp van de Ensemble Kalman Filter (EnKF). De verschillende versies (ensembles) van het model zijn gecreëerd doormiddel van het variëren van vier invoerparameters. De vier parameters zijn: actuele verdamping, neerslag, hydrologische conductiviteit bij verzadiging en de dikte van de eerste laag van het model.

De ERS-satelliet metingen correleren goed met de bovenste twee lagen van het model. Deze hebben een gemiddelde dikte van respectievelijk ~0.20 m en ~1.00 m. De correlatie is het grootste voor de bovenste laag en tijdens de natte zomer periodes.

De toepassing van de Ensemble Kalman Filter heeft niet tot betere resultaten geleid. De verschillen tussen de ERS-satelliet metingen en de geassimilereerde modelresultaten zijn groter dan de verschillen tussen de satellietmetingen en de originele - niet geassimilereerde- modelresultaten. Dit komt waarschijnlijk door de methode die is gebruikt om de verschillende versies van het model te creëren. De methode die is gebruikt om verschillende versies van regendata te creëren introduceert een bias, waardoor de grond natter lijkt. In het assimilatie proces wordt het model weliswaar geupdate, maar de invloed is gering. Dit komt voornamelijk doordat een grote onzekerheid is toegekend aan de ERS-satelliet metingen ten opzichte van de onzekerheid van het model, dat gebaseerd is op de variatie van de vier parameters.

Om de huidige resultaten te verbeteren zal een simpelere methode gebruikt moeten worden om de model versies te creëren. Deze versies moeten op een dusdanige wijze worden gecreeërd dat er geen bias en geen onrealistische waarden ontstaan.
# Table of contents

Preface ....................................................................................................................... i  
Abstract ..................................................................................................................... iii  
Samenvatting ............................................................................................................ v  
List of acronyms ....................................................................................................... 1  
List of symbols ......................................................................................................... 1  
1. Introduction ........................................................................................................... 3  
  1.1 The relevance of soil moisture ........................................................................... 3  
  1.2 Purpose ............................................................................................................. 4  
  1.3 Readers guide .................................................................................................. 4  
2. Study area, data and method ............................................................................. 5  
  2.1 West Africa ..................................................................................................... 5  
  2.2 Conceptual hydrological model ...................................................................... 7  
  2.3 Data ................................................................................................................. 9  
  2.4 Method .......................................................................................................... 12  
3. Results .................................................................................................................. 23  
  3.1 Data analysis ................................................................................................... 23  
  3.2 Data assimilation ............................................................................................ 27  
4. Discussion ........................................................................................................... 33  
5. Conclusions and recommendations .................................................................. 35  
  5.1 Conclusions ................................................................................................... 35  
  5.2 Recommendations ......................................................................................... 36  
Reference ................................................................................................................. 41  
Appendix .................................................................................................................. CD  
  Appendix A: West Africa  
  Appendix B: Soil moisture theory  
  Appendix C: The PCR-GLOBWB model  
  Appendix D: Satellite details  
  Appendix E: Actual evaporation  
  Appendix F: Model initialization  
  Appendix G: Waterbalance and runoff results  
  Appendix H: Model- and observation results  
  Appendix I: Correlations  
  Appendix J: Sensitivity analysis  
  Appendix K: Data assimilation results
List of acronyms

AGPI   TRMM-adjusted GOES precipitation index  
DA     Data Assimilation       
EARS   Environmental Analysis & Remote Sensing  
EKF    Extended Kalman Filter   
EnKF   Ensemble Kalman Filter   
ERS    European Remote Sensing Satellite (ESA)  
ESA    European Space Agency    
GLCC   Global Land Cover Characterization   
GOES   Geostationary Observatioanl Environmental Satellite  
JAXA   Japan Aerospace eXploration Agency  
KF     Kalman Filter       
NASA   National Aeronautics and Space Administration    
PCR-GLOBWB   PCRaster Global Water Balance    
RMSE   Root-Mean-Squared-Error    
SMOS   Soil Moisture and Ocean Salinity mission (ESA)    
TRMM   Tropical Rainfall Monitoring Mission (NASA/JAXA)  

List of symbols

\( \sigma^0 \)   Backscatter   
\( E_a \)   Actual evaporation   
\( m_s \)   Surface wetness ratio  
\( P \)   Precipitation     
\( Q_{bf} \)   Base flow   
\( Q_{Channel} \)   Modelled river or channel flow  
\( Q_{CR} \)   Capillary rise flux  
\( Q_{DR} \)   Direct surface runoff flux    
\( Q_{Perc} \)   Percolation flux   
\( Q_{iy} \)   Modelled interflow or subflow  
\( Q \)   Discharge  
\( S_c \)   Maximum saturation   
\( S_{res} \)   Residual saturation  
\( S \)   Soil moisture content  
\( S_f \)   Maximum soil moisture content  
\( Z_1 \)   Thickness of the first store
Assimilation of remote sensed soil moisture data into a hydrological model using the EnKF
1. Introduction

Water is essential to life on earth. Not surprisingly, men have always been searching for water and have been trying to understand its behaviour to the fullest. A famous Roman architect and engineer named Vitruvius (born ~100-80 B.C.) was aware of the importance of water and wrote a book on water (The Project Gutenberg EBook, 2006). Vitruvius knew that if fresh surface water was not available, water could be found underground by digging a well. According to him, the right place to dig can be determined by conducting the following test: ‘Before sunrise, lie down flat in the place where the search is to be made, and placing the chin on the earth and supporting it there, take a look out over the country. [...] Then, dig in places where vapours are seen curling and rising up into the air. This sign cannot show itself in a dry spot’.

Although methods to find water have been improved since, it is only in recent decades that methods have been significantly developed. This is mainly due to the remote sensing techniques that have enriched our understanding of water. We have come to the awareness that fresh water is not an unlimited resource and thus requires proper management. For this, soil moisture plays a very important role. Soil moisture is the amount of water held in the ground. If the soil moisture states and flux are known, water management can be improved. In the following paragraph (1.1) the relevance of soil moisture is explained in more detail. The purpose of this thesis is described in 1.2. Boundary conditions and assumptions are also given. This chapter concludes with a readers guide.

1.1 The relevance of soil moisture

Soil moisture plays a very important role in environmental processes (Scipal et al., 2005; Dabrowska-Zielinska et al., 2001; Wagner et al., 1999). It is for example the most important variable affecting climate after sea surface temperature (Berg and Famiglietti, 2003). It determines the partitioning of precipitation into runoff and infiltration. As the soil becomes saturated, less water will infiltrate into the soil, resulting in more direct surface runoff. By knowing soil moisture states, flood- and drought predictions can be improved (Scipal et al., 2005) and more insight into climate changes can be obtained.

Soil moisture can be measured by ground observations and by remote sensing. However, for many applications that use soil moisture data, ground observations cannot meet the data requirements (Wagner and Pathe, 2005). Ground observations are point measurements in a spatial and temporal sense, they are costly (Wagner and Pathe, 2005; Scipal et al., 2005; Wagner et al., 1999) and often require site-specific calibration (Moens, 1989). In remote sensing, information is collected in a systematic way, which allows time series. Remote sensing also covers wide areas such as entire river basins, whereas ground observations are often only available for small pilot areas (Bastiaanssen, 2000).

Numerous studies have identified techniques to observe soil moisture through remote sensing (Berg and Famiglietti, 2003). Data acquisition of soil moisture by optical sensors, however, is often disturbed by unfavourable weather conditions, such as clouds (Dabrowska-Zielinska et al., 2001). According to Woodhouse and Hoekman (2000), there are several studies that highlighted the relationship between low-resolution scatterometer data and surface geophysical parameters such as vegetation cover, surface roughness, and surface soil moisture content. Microwaves are not disturbed by weather conditions. The ERS-1 and ERS-2 satellites,
which have scatterometers on board, can be used for moisture states estimation for large-scale areas. They have been widely applied in many areas around the world (Wagner et al., 1999). According to some scientists, satellite data provide information only for the top few centimetres and will be limited to areas of low to moderate vegetation foliage (Njoku and Entekhabi, 1996). Other scientists however, doubt that satellites provide information of the top few centimetres only. The seasonal variations are higher than one would expect and thus they conclude that satellites either represent soil moisture over greater depth, are influenced by vegetation litter or are otherwise greatly influenced by the effect of scale (NIVR and Vrije Universiteit van Amsterdam, 2007). It can be concluded that there is still discussion on how to interpret the ERS-satellite observations.

1.2 Purpose

Problem description
It is largely unknown by which spatial features and hydrological processes the accuracy of the soil moisture state estimates, from the ERS-satellite, are influenced. In order to make better use of estimates on soil moisture states, more insight is needed in what influences the accuracy of satellite measurements.

Goal
This thesis aims to determine by which hydrological processes and spatial features the accuracy of the ERS-satellite estimated soil moisture is influenced. The estimates will be compared with - and assimilated for a conceptual hydrological model using the ensemble Kalman filter.

The study focuses on a specific study area, West Africa, which is defined as the region between 0˚N, 20˚W to 20˚N, 20˚E. There are large spatial and temporal variations in this region, which are expected to be clearly visible in the ERS-satellite measurements. The period between 1/1/2003 to 31/12/2006 is studied. The time step is a single day (24 hours).

1.3 Readers guide

Chapter 2 first presents a short description of West Africa. The conceptual hydrological model and the various data, such as the ERS-satellite measurements, precipitation and actual evaporation measurements, are described thereafter. Chapter 2 concludes with a description of the method which is used to compare and assimilate the ERS-satellite measurements with the model. The results are presented in chapter 3. These results are discussed in a broader context in chapter 4. Finally in chapter 5 the conclusions and recommendations are presented.
2. Study area, data and method

This chapter starts with a short description of West Africa (2.1). In paragraph 2.2 the conceptual hydrological model is given. The theory and information on the various data sets, such as the ERS-satellite soil moisture estimates, the precipitation- and evaporation measurements are given in 2.3. Finally in 2.4 the method that is used in this thesis is described.

2.1 West Africa

West Africa is a very diverse region. In this paragraph considers the geography, climate and demography of this region. West Africa is chosen as the study area because large spatial and temporal variations occur. Large seasonal variations lead to large variations in the vegetation coverage over the year. These variations are expected to be clearly visible in the ERS-satellite measurements as well.

![Digital Elevation Map West-Africa](image)

Figure 2.1. West Africa, study area. The elevations are given in meters (Hydro1k, 2008).

Geography

The study area is located south of the Sahara in West Africa, see figure 2.1, and extends from 0°N, 20°W to 20°N, 20°E. The total land area is about 6,140,000 km². Mountainous regions are located in the Southwest of the study area, in Guinea (~8-12°N, ~8-12°W), in the Northeast of Nigeria (~10°N, 8°E), in Cameroon (5-7°N, 10-15°E), in the large areas south of Cameroon (2-5°N, 11-15°E) and in the North of the study area in Mali (20°N, ~5°E) and Niger (17-20°N, 9°E). Inselbergs, which are isolated hills where the surrounding plains have eroded away, can be found across the Sahel (Concise Atlas of the World, 2005).

There are about five typical land covers in the study area. In the North, there is mostly barren land, towards the South there is first a thin band of shrub land followed up by a band of grassland. In the South there are savannas and tropical forests. Other land cover types in the region are cropland (located mainly between 10°-14°N), permanent wetlands (mainly located at delta of rivers), mangroves, urban or industrial area and water bodies. On cropland, peanuts and cash crops such as coffee, cotton, cocoa and rubber are grown in large quantities (Concise
Atlas of the World, 2005). There is large tropical deforestation in West Africa. Less than 25% of the original forest still exists in the study area and between 1980 and 1990, the rate of deforestation was 4 - 10% (EarthTrends, 2008).

Within the study area, there are three main watersheds for which the runoff is modelled in the hydrological model (see 2.2). The three main watersheds are the Niger-, Volta- and Senegal-river basins. Figure 2.2 shows the surface area of each watershed (IUCN, IWMI, Rasmar Convention Bureau and WRI, 2003). There are some Ramsar sites, which are ecologically important wetlands, at the downstream end of the Senegal and Volta watersheds (Earthtrends, 2008). The delta of the Niger consists mainly of swamp forests. The most dominant feature of the Volta watershed is the large artificial Lake Volta. Many small reservoirs exists upstream of Lake Volta. The Niger has a large freshwater marsh and wetlands in Mali.

Figure 2.2. The major watersheds in West Africa: the Senegal -, Volta - and Niger Watershed. Surface areas are shown in km² (IUCN, IWMI, Rasmar Convention Bureau and WRI, 2003).

Climate
The climate varies through West Africa. The North is very hot and dry, whereas the South is hot, wet and humid. The average yearly temperature ranges between 24 and 29 degrees Celsius.

The climate is dominated by two air masses. A high pressure cell is located over the Sahara, causing dry Northeast trades, or harmattans. These winds can travel all the way to the south of Western Africa in boreal winter. Another high pressure cell is located over the tropical Atlantic, bringing very humid air to Western Africa and consequently causing heavy rainfall. Winds caused by this Atlantic pressure cell are known as the July Winds. In between the two air masses is the Intertropical Convergence Zone (ITCZ) (Encyclopædia Brittanica, 2007). The annual rainfall ranges from over 1400 mm·year⁻¹ in the South to less than 200 mm·year⁻¹ in the North of the study area.

An overview of the average rainfall and evaporation per day over 2003-2006 is shown in figure A.2 and A.3 of Appendix A. The difference between the average precipitation and actual evaporation per day is shown in figure A.4 of appendix A. There is a surplus of precipitation in most coastal regions below 10°N. The difference is negative for most of the Senegal river basin and for the Upper Niger Basin. For the rest of the study area is the difference close to zero.
Demography and economy
West Africa is one of the poorest regions in the world. Most of the countries have a Gross National Product per capita (GNP) of less than 400 US dollar. In contrast, most of the EU countries have GNP per capita of more than 20,000 US dollar. The population density in the North is less than 340 people km\(^{-2}\), in the South 600-1300 people km\(^{-2}\) and in Senegal and Niger the density is over 1300 people km\(^{-2}\). Over 60% of the workforce is employed in agriculture (EarthTrends, 2008).

2.2 Conceptual hydrological model

In this thesis, the ERS-satellite measurements are compared to model results. The measurements are also used to update the model states. How the measurements are compared with the model and how the model states are updated is explained in paragraph 2.4. This paragraph describes the conceptual hydrological model.

Hydrological models describe the movement of water within the modelled area. The models are simplifications of the real world, meaning that certain processes are simplified or omitted. Nevertheless these models help in the understanding of the behaviour of water. For this particular thesis, a hydrological model can give more insight in the behaviour of soil moisture. The underlying theory on soil moisture is given in appendix B (Ward, 1975).

![Figure 2.3. Model concept of PCR-GLOBWB. The soil compartment is divided into three stores; two upper soil stores and one groundwater store. From each store, there is a drainage component adding to the discharge of the channel. The drainage component of the first store is the direct runoff \(Q_{DR}\), of the second store is the subflow or interflow \(Q_{SF}\) and the drainage component of the third store is the baseflow \(Q_{BF}\). Evaporation \(E_a\) takes place in each store and percolation \(Q_{PC}\) and capillary rise \(Q_{CR}\) may occur between them. The top store receives water from precipitation \(P\). The channel can receive water from precipitation directly, while direct evaporation may take place as well. The PCR-GLOBWB (PCRaster Global Water Balance) model, developed by van Beek (2007), is used for this thesis. It is a conceptual hydrological model for the whole world. Conceptual models represent the most relevant catchment processes through a number of reservoir and](image)
fluxes. The input of the model consists of precipitation- and actual evaporation data and geophysical data such as elevation, soil- and vegetation maps. Soil moisture content and runoff are the output of the model. The output can be compared to soil moisture estimates of the ERS-satellite after these measurements have been transformed to soil moisture states for the top layer of the model, which is explained by formula 2.30 in paragraph 2.4.2. Data assimilation techniques can be used to update the model states with the satellite measurements. This is also explained in 2.4.2.

For this thesis a cut-out of West Africa is used. The PCR-GLOBWB model origin is based on the HBV-model, which is described in Bergström (1992). The model concept is given in figure 2.3. In the model, the soil compartment is divided into three stores. From each store there is a drainage component adding to the total discharge. As rain falls on the topsoil layer, some is intercepted and some reaches the surface. The amount of interception depends on the vegetation cover, the vegetation type and the amount - and duration of precipitation. In the model, the interception process can be seen as a small bucket that fills up during rain events. When the bucket is full, all water is passed on directly to the surface. At the surface, the rain may flow overland as direct runoff or may infiltrate into the first store, increasing its soil moisture content. Water may percolate from store 1 towards store 2 or rise through capillary forces from store 2 to store 1, thus influencing the soil water content in store 2. The mechanisms of percolation and capillary rise are explained in more detail in appendix C. From store 2, there is an interflow component adding to the total discharge. Water may again percolate or rise between store 2 and 3. Store 3 represents the groundwater reservoir. It is fed by the net percolation, which is described as the difference between the percolation from store 2 to store 3 minus the capillary rise from store 3 to store 2. Drainage from store 3 contributes as base flow to the total discharge. The change in water storage \( \frac{dS}{dt} \) [m/d] can be described as:

\[
\frac{dS}{dt} = Q_{\text{Perc};i-1} + Q_{\text{CR};i} - (Q_{\text{Perc};i} + Q_{\text{CR};i} + E_{\text{a};i} + Q_{i}),
\]

(2.1)

where index \( i \) represents the layer number, \( Q_{\text{Perc}} \) the percolation [m/d], \( Q_{\text{CR}} \) the capillary rise [m/d], \( E_{\text{a}} \) the actual evaporation [m/d] and \( Q \) the lateral drainage [m/d].

The hydraulic conductivity determines how quickly water can move through the soil. In this model, the hydraulic conductivity is a function of the saturated hydraulic conductivity, which is fixed over time. In appendix C, the relationship between the two is explained. The hydraulic conductivity influences the soil moisture content directly. The values of the saturated hydraulic conductivity used in the model is based on the FAO soil dataset (Batjes, 2005). The saturated hydraulic conductivity maps for the first and second store are shown in figure C.3 and C.4 of appendix C. A detailed soil map is shown in appendix A.

The model requires precipitation and actual evaporation on a 0.25 by 0.25 degrees resolution as inputs. The water storage in the three stores and the total discharge are the outputs of the model and can be generated per day.
2.3 Data

As explained, the hydrological model requires precipitation and actual evaporation as input. The model states of the first layer are compared with the ERS-satellite estimated soil moisture states. In this paragraph, these data products are described in more detail. It is described how the data is retrieved and what is the spatial and temporal resolution.

2.3.1 ERS-satellite estimated soil moisture

The ERS-satellite has a scatterometer on board and three antennae producing beams looking to 45 degrees backward, 90 degrees sideways and 45 degrees forward with respect to moving direction of the satellite along its orbit. The active microwave scatterometer transmits pulses of microwave energy towards the earth’s surface and measures the reflected energy at different reference angles. After noise is subtracted from the reflected signal, the backscatter signal power (\(\sigma^0\)) is known. In appendix D is explained in more detail how the ERS-satellite measures the \(\sigma^0\)-values (Woodhouse and Hoekman, 2000). The orbital details of the ERS-satellite are presented in table D.1 of appendix D.

In several studies, it is found that the backscatter signal (\(\sigma^0\)) is greatly influenced by vegetation, surface roughness and dielectric profile of the soil (e.g. Ulaby and Batlivala, 1976; Njoku and Entekhabi, 1996; Li et al., 2005). The dielectric profile of the soil depends largely on the water content; soil moisture (Ulaby and Batlivala, 1976). Open water areas have a dielectric constant of about 80. The dielectric constant can differ in the soil between 4 and 40, depending on moisture content and soil type. A dielectric constant of 40 represents a very wet soil and a soil with a dielectric constant of 4 is dry.

Figure 2.4. Schematics of a cross section of West Africa from 5˚N to 20˚N at ~5˚W. The cross section is divided into six regions. Each region influences the ERS (backscatter) signal in a different way. The dielectric constant (d.c.) is ~80 for open water and ranges from ~40 to ~4 for saturated and dry soil respectively. The roughness ranges from very smooth (--) to very rough (++). The bedrock formation is shown in grey.

Figure 2.4 shows a schematics of the cross section of West Africa from 5˚N to 20˚N at ~5˚W. The cross section is divided into six regions, each influencing the backscatter signals in a different way. The first region represents the ocean or large open water areas such as Lake Volta. The ocean and Lake Volta are very flat compared to the earth surface and have a
dielectric constant of about 80. It acts like a mirror; the transmitted signal is not reflected but ‘bounces off’ in a different direction and thus the backscatter is hardly measured. The backscatter signal increases if the ocean surface becomes rougher. The ocean surface becomes rougher due to waves, which are a result of increasing wind speeds. This is why the ERS-satellite (Wind Scatterometer) was initially designed to measure wind speed over oceans (Woodhouse and Hoekman, 2000).

The other regions shown in figure 2.4 are of particular interest for this thesis. They represent different surface types, each having a different dielectric constant and roughness. The higher the dielectric constant and the rougher the surface, the higher the backscatter signal. The dielectric constant generally decreases from the South towards the North, because the soil becomes more dry towards the North. At wetlands, the dielectric constant is close to 40 and thus there is higher backscatter signal (see region 5). Vegetated areas, mountainous regions and desert regions are considered rough. The backscatter signal is higher for these regions. At mountainous and desert regions the bedrock may reach the surface and influence the backscatter signal through azimuthal modulation (Bartalis et al. 2006). Azimuthal effects can occur for example at parts that are very dry. The signal from the ERS-satellite can then penetrate deep into the soil and reach the underlying bedrock. The ERS-satellite actually sees the underlying topography instead of the surface topography. The TU-Wien data has not been corrected for azimuthal modulation.

Figure 2.5. The overpass frequency of the ERS-2 satellite over the covered regions given as the number of days between observations on average.

$$m_i = \frac{\sigma_0^w - \sigma_0^d}{\sigma_0^w - \sigma_0^d}$$

For this thesis, the surface wetness data (version 05-11-2006) from the Vienna University of Technology (TU Wien) is used. The data supposedly represents the surface wetness of the first 0.5 to 2 centimetres of the earth’s surface (Wagner et al., 1999). To retrieve surface soil moisture states, the backscattering coefficient is extrapolated to a reference angle of 40 degrees. From long backscatter series the lowest and highest backscatter are stored (Wagner et al., 1999). The lowest backscatter is considered to represent a completely dry situation ($\sigma_0^{dry}$).

---

1 Azimuthal effects can also occur at large croplands where one crop type is grown in a systematic way or at other predominantly orientated targets. For more information on azimuthal modulations, refer to Bartalis et al. (2006)
and the highest backscatter a saturated situation ($\sigma_{0w}^\sigma$). The soil moisture states are given as a relative degree of saturation (the surface wetness). The surface wetness ($m$) is derived according to formula 2.2. In Wagner et al. (1999) it is concluded that this method allows one to distinguish at least five soil moisture classes: very dry, dry, medium, wet and very wet.

Data is obtained from the end of August 2003 until the end of October 2006 and is obtained every 3-5 days on average. The data is interpolated from a 50 km resolution to 28 km grid, which represents 0.25° spacing in latitude and a latitude dependent spacing in longitude direction (Wagner et al., 1999). The study area is not fully covered, see figure 2.5. In June 2003, the onboard data storage capability of the ERS-2 satellite failed (TU Wien, 2007). Since then, data can only be retrieved when the satellite is in sight of a ground station. Furthermore, some parts in the North have been excluded from coverage in the data set because of the azimuthal anisotropy. The relatively small areas of no coverage near 6°W, 7°N and near 0°E, 6°N represent Lac de Kossou (Ivory Coast) and Lake Volta (Ghana) respectively. Lastly, some pixels that lie in certain curves have no data. It is unknown why these pixels have no data.

### 2.3.2 Remote sensed actual evaporation

This thesis uses actual evaporation data from EARS (http://www.ears.nl). The data are derived from the Meteosat I and II satellites. The orbital details of the Meteosat II satellite are given in appendix D. In appendix E is explained how the actual evaporation is derived. The data was interpolated\(^2\) from 0.05 by 0.05 degrees to 0.25 by 0.25 degrees, so it could be used as input for the model. The data does not contain actual evaporation information over large open water areas such as the ocean and lake Volta. This leads to unknown errors in the water balances for these areas.

### 2.3.3 Remote sensed precipitation

Precipitation data is collected by the TRMM satellite. The Tropical Rainfall Measuring Mission (TRMM) is a joint mission between NASA and the Japan Aerospace Exploration Agency (JAXA) designed to monitor and study tropical rainfall (NASA, 2005). The TRMM satellite consists of several components\(^3\), namely:

- Precipitation Radar (PR)
- TRMM Microwave imager (TMI)
- Visible and InfraRed Scanner (VIRS)
- Cloud and Earth Radiant Energy Sensor (CERES)
- Lightning Imaging Sensor (LIS)

TRMM estimates rainfall intensities on a 3-hourly temporal resolution and a 0.25 by 0.25 degrees spatial resolution for the part of the globe between 50°S and 50°N. The TRMM 3B42 algorithm calculates the precipitation intensity by combining the measurements of all components. The algorithm has four steps (NASA, 2003). Firstly, the microwave precipitation estimates are calibrated and combined, then the infrared precipitation estimates are created using the calibrated microwave precipitation and thereafter the microwave and IR estimates are combined. The final step is to rescale the data to monthly data, but in this thesis the products of this final step are not used. Instead, the daily precipitation was computed from the 3-hourly data.

---

\(^2\) Linear interpolation is used. The algorithm of Mathworks, called `griddata`, is used. Refer to the website of Mathworks for more information on this algorithm (http://www.mathworks.com)

\(^3\) More information on the components, refer to the NASA website of TRMM (http://disc.gsfc.nasa.gov/TRMM/index.shtml)
2.4 Method

In the previous paragraphs, the hydrological model and the datasets have been described. This paragraph explains the method used to analyze and assimilate the data. The method can be divided into two parts:

- Data analysis (2.4.1);
- Data assimilation (2.4.2);

The data analysis (2.4.1) consists of water balances of three watersheds and the correlation between the different data sets. Paragraph 2.4.2 explains what data assimilation (DA) is, what DA-method is used, how the errors of different data are estimated, how the DA-method is applied and how the results are analyzed.

2.4.1 Data analysis

The data is analyzed by checking the waterbalances for the different watersheds, comparing modelled runoff to measured runoff and by computing correlations between the data sets. Note that the hydrological model is run from 1/1/2003 to 31/12/2006 with timesteps of a day.

Water balance

The water balance is computed for the three main watersheds of the study area, namely: the Senegal-, Volta- and Niger basin. The difference between the precipitation and the actual evaporation is first computed, then followed by the computations of the waterbalance. The waterbalance is defined as:

$$\Delta S = (P - E_a - Q) \cdot \Delta t,$$

where $\Delta S$ is the change in water storage in the soil (m$^3$), $P$ is the precipitation (m$^3$), $E_a$ the actual evaporation (m$^3$) and $Q$ the river discharge (m$^3$). The local groundwater recharge is omitted because horizontal groundwater flow is assumed to be zero.

Runoff analysis

One of the outputs of the hydrological model is the daily runoff created per day for each grid cell. A separate routing model simulates how the daily runoff values are routed down to the outlets at lakes and oceans. In appendix C it is explained how the daily runoff values are generated. A comparison is made for the Volta Basin between the modelled runoff and the measured runoff at Sabari (Black Volta) and at Bamboi (Oti) for the period of 2003-2006. This is done because the amount of daily runoff is mainly dependent on the soil moisture content of the three different stores. The final results of the modelled discharge is also dependent on the routing model. The set of variable values is apparently well chosen if the modelled resembles the measured runoff. It may give some confirmation that the soil moisture content is also well modelled. However, if the modelled discharge is not in accordance with the measured discharge, then it may be due to the errors in the routing model, in the hydrological model or in both. It does not necessarily imply that the soil moisture content is modelled wrongly.
Table 2.1. The correlations that are calculated. \( S_1 \) is the modelled soil moisture content of store 1 and \( S_2 \) is the modelled soil moisture content of store 2.

<table>
<thead>
<tr>
<th></th>
<th>( P )</th>
<th>( E_{ax} )</th>
<th>( P-E_{ax} )</th>
<th>( S_1 )</th>
<th>( S_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P )</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( m_t )</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>( S_1 )</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( S_2 )</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Correlation**

Correlation is a measure of how much one random variable depends upon the other. A high correlation between two data sets could imply that there is a relationship between the two. Table 2.1 shows which correlations are calculated. The correlations are computed for seven different periods, which represent summer and winter periods. The summer period is defined as the period from 1 May - 30 October and the winter period is defined as 31 October to 30 April. The periods are chosen as such because the ERS-satellite data was available until 30/10/2006. With these defined periods, only the first period is not equal to half a year, because the first measurement of the ERS-satellite was available in August 2003.

### 2.4.2 Data assimilation

In this paragraph, it is described what data assimilation (DA) is, what DA-method is used, how the errors of the different data sets are estimated, how the DA-method is applied and finally how the results are analyzed. The paragraph is divided as follows:

- Data assimilation theory
- Error estimation
- Application of the Ensemble Kalman Filter
- Result analysis

#### Data assimilation theory

Calibrated hydrological models contain errors in model structure, input, and output data. These errors lead to uncertainties in the model predictions. To address hydrologic uncertainty properly, it is important to understand, quantify and reduce the uncertainty in a systematic way (Liu and Gupta, 2007). Data assimilation techniques provide a way for dealing with input, output and structural uncertainty, and for optimal merging of uncertainty model predictions and observations (Vrugt et al., 2005), which contain errors as well. Data assimilation aims to limit the divergence between simulated and observed states or variables (Aubert et al., 2003). Many different data assimilation techniques exist (Liu and Gupta 2007), each addressing uncertainties and observations in a different way. Most data assimilation techniques are based on the Kalman filter (KF) or the extended Kalman Filter (EKF). The EKF is well explained in Aubert et al. (2003). For this thesis the Ensemble Kalman Filter (EnKF) is used, because it can better cope with non-linear models than the EKF. A brief explanation of the EnKF, based on Evensen (2003) and Aubert et al. (2003), is given here.

#### Kalman Filter (KF)

The Kalman Filter is a sequential filter method. This means that the model predicts the model state of the next time step. Furthermore, whenever a measurement is available, this
measurement is used to update the model state before the prediction for the next time step is made.

The model variables are described by a vector $x_k$. For this particular thesis, the vector consists of only the soil moisture state of the first store because of the ease of computations. In Kalman filtering, when a measurement ($z_k$) is available, it is required to translate the measurement to a model variable ($x_k$) through a constraint equation $H_k$:

$$z_k = H_k \cdot x_k$$

(2.4)

The hydrological model itself is represented by an operator $f$, which is explained in 2.2. So the variable value for the next timestep $k+1$ is calculated as: $x_{k+1} = f(x_k)$

The model variables are updated with available measurements via the Kalman filter. For this it is necessary to identify two error covariance matrices: $R_k$ for the uncertainties in the observations and $Q_k$ for the uncertainties in the model. The latter represents the uncertainties during one time step and does not account for the total uncertainties of the model. To account for the propagation of the total uncertainties, a matrix $P_k$ is defined. This matrix includes the propagation of uncertainties from time $k-1$ to time $k$. On the diagonal of $R_k$ the uncertainties of the measurements are defined. The matrix $P_k$ can be a identity matrix and is updated later. How the uncertainties are defined is explained later.

The calculation of the Kalman filter is implemented in two phases: the adjustment phase and the propagation phase. In the adjustment phase, the gain matrix $K_k$ is calculated:

$$K_k = P_k \cdot H_k^T \cdot (H_k P_k H_k^T + R_k)^{-1}$$

(2.5)

It indicates that for each internal parameter, the correction factor will be applied due to each observation. The calculation takes into account the error matrices of the observation and model. Thereafter the a posteriori state vector $\hat{x}_k$ is equal to the a priori state vector $x_k$ plus the difference between the observation and state vectors multiplied by the Kalman gain:

$$\hat{x}_k = x_k + K_k \cdot (z_k - h(x_k))$$

(2.6)

The error covariance matrix is adjusted as well:

$$P_k^t = (I - K_k \cdot H_k) \cdot P_k$$

(2.7)

The parameter value for $k+1$ is then calculated as:

$$x_{k+1} = f(\hat{x}_k)$$

(2.8)

The error covariance matrix for $k+1$ is calculated by adding the error at time $k$ ($P_k$) propagated via the tangent linear and the error $Q_k$ generated during this time step:

$$P_{k+1} = M_k \cdot P_k^t \cdot M_k^T + Q_k,$$

(2.9)
in which the Jacobian matrix $M$, which is defined by the partial derivative of the model, represents the tangent linear model:

$$M_{k[i,j]} = \frac{\partial f(x_{i,j})}{\partial x_{k[i]}} ,$$

(2.10)

where $i$ and $j$ stand for the matrix rows and columns.

The magnitude of the correction is thus dominated by the ratio of errors on the observations and the model. Very small uncertainties in observations lead to a Kalman gain of almost equal to 1 so that the internal states will be set to the observations. If the uncertainties in the observations are high, the Kalman gain will be almost equal to 0 and the correction will be almost zero.

The Ensemble Kalman Filter (EnKF)

The Ensemble Kalman Filter uses an ensemble of model states to represent the error statistics of the model estimate and uses an analysis scheme which operates directly on the ensemble of model states when observations are assimilated (Evensen, 2008). The tangent linear model and the update of error covariance matrix are not required to be calculated. The EnKF is explained here. It is also well explained in Evensen (2003).

For the EnKF, $N$ ensembles of model states exists. Matrix $A$ is defined and contains the ensembles of model states $x_i \in \mathbb{R}^n$:

$$A = (x_1, x_2, ..., x_N) \in \mathbb{R}^{n \times N} ,$$

(2.11)

where $N$ is the number of ensembles and $n$ is the size of the state vector.

The ensemble mean is stored in each column of $\bar{A}$:

$$\bar{A} = A \cdot 1_N ,$$

(2.12)

where $1_N \in \mathbb{R}^{N \times N}$ is the matrix where each element is equal to $1/N$. The ensemble perturbation matrix is then defined as:

$$A' = A - \bar{A}$$

(2.13)

The ensemble covariance matrix $P_e \in \mathbb{R}^{n \times n}$ is defined as:

$$P_e = \frac{A' \cdot (A')^T}{N - 1}$$

(2.14)

Next, $N$ vectors of perturbed observations are defined as:

$$d_j = d + \varepsilon_j , \ j = 1, ..., N ,$$

(2.15)

which are stored in the columns of matrix

$$D = (d_1, d_2, ..., d_N) \in \mathbb{R}^{m \times N} ,$$

(2.16)

where $m$ is the number of observations. The perturbations are stored in a separate matrix as well:

$$\Upsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_N) \in \mathbb{R}^{m \times N} .$$

(2.17)
The ensemble representation of the measurement error covariance matrix is then defined as:

\[ R_e = \frac{\Upsilon \cdot \Upsilon^T}{N - 1}. \quad (2.18) \]

The Kalman gain becomes:

\[ K_e = P_e \cdot H_e^T \cdot (H_e \cdot P_e \cdot H_e^T + R_e)^{-1} \quad (2.19) \]

Or

\[ K_e = A_e^T \cdot H_e^T \cdot (H_e \cdot A_e^T \cdot H_e^T + \Upsilon_e \cdot \Upsilon_e^T)^{-1} \quad (2.20) \]

To overcome problems with singular matrices, the following is applied after it is checked that \( H \cdot A \cdot \Upsilon' \equiv 0 \):

\[ H \cdot A \cdot \Upsilon' = U \Sigma V^T \quad (2.21) \]

The singular value decomposition (SVD) is then calculated:

\[ H \cdot A \cdot \Upsilon' = U \Sigma V^T \]

Eq. 2.18 becomes:

\[ H \cdot A \cdot \Upsilon' \cdot H_e^T + \Upsilon_e \cdot \Upsilon_e^T = U \cdot \Sigma \cdot V^T \cdot V \cdot \Sigma^T \cdot U^T = U \cdot \Sigma \cdot \Sigma^T \cdot U^T \quad (2.23) \]

The update is calculated by

\[ A_e' = A + A_e \cdot (H \cdot A \cdot \Upsilon')^T \cdot U \cdot \Sigma \cdot \Sigma^T \cdot U^T \cdot D' \quad (2.24) \]

where \( D' \) is \( D \cdot H \cdot A \).

Error estimation
To obtain \( N \) ensembles of the model state, some model variables can be perturbed. In this thesis, \( N \) perturbations of four variables are created. Perturbations are created for:

- Actual evaporation;
- Precipitation;
- Saturated hydraulic conductivity;
- Thickness of the first layer.

It is expected that the soil moisture content of the first layer is mainly influenced by these four variables. Actual evaporation and precipitation are the input variables of the model. The hydraulic conductivity, which is a function of the saturated hydraulic conductivity, determines how quickly water can move through soil. The thickness of the first layer is mainly of interest when comparing the ERS measurements with the model results. In several papers (e.g. Njoku and Enthekabi, 1996; Wagener et al., 1999) it is stated that the ERS-satellite measures the soil moisture content for the first 0.5 to 2 cm or for the first 5 cm at most. The first store in the model has a thickness of 20 cm on average. Comparing the soil moisture content of the model with the ERS-satellite would therefore be not very straightforward. However, it is done directly in this thesis. Therefore it is necessary to investigate what is the influence of the thickness of the first layer and the consequence this thus have on the difference with the ERS-satellite estimated soil moisture states.

The perturbations must be in the order of magnitude of the uncertainty of the parameter. Therefore, first the uncertainties of the variables are determined. How the \( N \) perturbations are created is discussed in the next subparagraph.
Actual evaporation error estimation

The accuracy of the actual evaporation data on a yearly time-scale was estimated in De Weird et al. (2006) at 10%. The actual evaporation was estimated as water balance residual ($E_a = P - Q$) for a part of the Upper Yellow River basin in China. This estimation was then compared to the measured actual evaporation by EARS and deviated up to 10% per year. There are however also uncertainties in the precipitation and runoff measurements, which lead to uncertainties in the water balance residual estimation. For this thesis, it is preferred to use the uncertainty of the EARS product on a daily time-scale, rather than a yearly timescale as estimated by de Weird et al. (2006).

The accuracy of the EARS product was then estimated on the basis of temperature and sensible heat flux validation presented in de Weirdt et al. (2006), see appendix E. The root-mean-squared error (rmse) found in the temperature and sensible heat flux measurements compared to ‘ground truth data’ was thereafter used to define the rmse of the actual evaporation product, which was largely based on these measurements (see Appendix E). Rosema et al. (personal communication, 2008) argued, however, that the ground truth data was in fact not that “true” and contained errors as well. Therefore, the accuracy of the EARS product had to be defined in a different way.

Finally, the variability of the data is determined as an indication of the uncertainty of the data. The variability is influenced by the underlying physics and also by the method used to determine the actual evaporation. In small areas the variability is expected to be small, because the area is (in general) more homogeneous. To determine the variability, the study area is divided into subareas of 2.5° by 2.5° (ten by ten grid cells). The variability is then determined for each subarea for each day (from 1/1/2003 to 31/12/2006). The lowest variability over the whole time-series is stored for each subarea, see figure 2.6. The lowest variability is zero for some subareas in the dry North and for those subareas containing only ocean grid cells. These subareas are filtered out. The mean of the lowest variability of the remaining areas is computed and is 0.18 mm·d⁻¹.

Precipitation error estimation

Precipitation greatly influences the model states. Unfortunately, little validation work has been performed on the TRMM product for (West) Africa. Nicholson et al. (2003) performed the
first validation of the various products of TRMM for West Africa. Validation was carried out for a 2.5˚ and for a 1.0˚ grid resolution over the period May – September 1998. The TRMM-adjusted Geostationary Observational Environmental Satellite precipitation index (AGPI; algorithm B342) has been used in this thesis as input for the hydrological model. The measurements were validated with gauged data from a West African network of 920 stations. Although there are some limitations in this work and in the TRMM satellite, the results are still used as reference to determine the perturbations for the precipitation, because of the lack of any other validation work. The limitations of the validation and the TRMM satellite are:

- The validation was limited to the wet period of 1998. Validation of measurements on gauged data for several years would be preferred;
- Direct validation of satellite measurements and gauged data is not possible, because gauge data represent point measurements. The point measurements are interpolated on the grid size of the satellite measurements and then used to validate the satellite measurements. Nicholson et al. (2003) state that the error of the areal averages from the gauges is about 10% or less when five or more gauges are available for a 2.5˚ grid and two or more for a 1.0˚ grid. It is unknown how the error is determined and what interpolation technique is used. Also the distribution of the errors is unknown;
- It was found that the TRMM products do not recognize coastal effects that lead to a strong degree of rainfall enhancement at Sierra Leone and Guinea. Therefore, no validation was performed for this region. Since this region is known to be very wet, it is surprising that there is a net precipitation deficit, see figure A.4 of appendix A. Validation of the TRMM product for this region is therefore preferred;
- It was also found that AGPI underestimates rainfall at most locations below 14˚N and overestimates rainfall above 14˚N. This is of importance when comparing model results and ERS-satellite measurements.

In the paper by Nicholson et al. (2003), the root mean squared error (rmse) and the bias were determined for the whole season and per month. It was found that the rmse and the bias for a 2.5˚ grid were respectively 0.6 mm · d⁻¹ and 0.2 mm · d⁻¹ for the whole season and about 1.2 mm · d⁻¹ and 0.2 mm · d⁻¹ on average per month. The rmse and bias are larger for the 1.0˚ grid. The rmse is 1.0 mm · d⁻¹ and the bias is 0.2 mm · d⁻¹ for the whole season and 1.34 mm · d⁻¹ and 0.26 mm · d⁻¹ respectively on average per month. It is unknown what the rmse and bias of the AGPI product is for a 0.25˚ grid on a daily time-scale, which is the grid size and time step of the hydrological model. It is expected that the rmse is larger at a daily timescale.

Saturated hydraulic conductivity error estimation
The mean and the standard deviation of the global dataset were given by van Beek (personal communication, 2008). The standard deviation of the saturated hydraulic conductivity for the first store is 1.25 (in ln[cm · d⁻¹]).

Error estimation of the thickness of the first model layer
The uncertainty of the thickness of the first layer is unknown. The increase and decrease of the thickness is set to a relative large value of 0.10 m. This is done, because the difference is large between the average thickness of the first layer of the model (~0.20 m) and the thickness of which the ERS-satellite measurements suppose to represent the degree of saturation (~0.05 m). The original maps of the first and second store of the model are shown in appendix C, see figure C.1 and figure C.2.
Table 2.2. The values with which the original variable values are increased and decreased in order to determine the sensitivity of the model to the change of variable value.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Increase/Decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual evaporation</td>
<td>± 0.18 mm ⋅ d⁻¹</td>
</tr>
<tr>
<td>Precipitation</td>
<td>±1.34 mm ⋅ d⁻¹</td>
</tr>
<tr>
<td>Saturated hydraulic conductivity</td>
<td>± 1.25 ln(cm ⋅ d⁻¹)</td>
</tr>
<tr>
<td>Thickness of first layer</td>
<td>±0.10 m</td>
</tr>
</tbody>
</table>

**Sensitivity analysis**

A sensitivity analysis is used to determine how ‘sensitive’ the hydrological model is to changes in variable values of the model. This helps to understand the model dynamics and may also give insight in what degree of accuracy is needed for a parameter in order to get model results that fit the expectation.

The sensitivity of the model to a variable is determined by letting the model run with data of two times a half year. For the first run, the parameter value is increased and for the second it is decreased by the value shown in table 2.2. The final model state of the first store (at day 181) is compared to the original model state at that same day. Analyzing the deviations per day would possibly give more insight of the sensitivity of the model to a parameter, but this is not done in this thesis.

**Figure 2.7.** The EnKF model setup schema. The N (perturbed) model states are created by feeding those models with the N perturbations of four parameters (precipitation \(P\), actual evaporation \(E_a\), saturated hydraulic conductivity \(K_s\) and the thickness of the first model layer \(Z_1\)). These model states \((x)\), together with N perturbed observations \((d)\) of the ERS-satellite, are then used in the Ensemble Kalman Filter to obtain N adjusted model states \((\hat{x})\).
Application of EnKF
As explained earlier, some scientists believe that the ERS-satellite measures the soil moisture content of the first few centimetres, while others believe it may contain information about soil moisture content of deeper layers as well. In this thesis, it is chosen to only update the soil moisture state of the first layer with the ERS soil moisture state observations.

The state vector $x_k$ in this research consists of model states of only one variable, namely the soil moisture content of the first layer. The size of the state vector ($n$) is thus equal to the total number of grid cells of the study area which is (81 by 161 =) 13041. The number of observations ($m$) is equal to the number of grid cells for which an observation is available. This number can differ per observation. Due to limited computer memory, the study area is divided into eight areas of 40 by 40 grids and to update per area$^4$. The size of the state vector ($n$) becomes 1600.

Prior to using the Kalman Filter, it is important to define how the ensembles members are created and how many ensembles members are created. Unfortunately, little attention is being given in the literature on this subject. Therefore part of this report is dedicated to this subject. How the data assimilation method is applied is explained thereafter.

To obtain $N$ ensembles of the model state, some model variables are perturbed. In this thesis, $N$ perturbations of four parameters are created. Perturbations are created for the actual evaporation and the precipitation and the saturated hydraulic conductivity, the thickness of the first store of the model. The perturbations must be in the order of magnitude of the uncertainty of the parameter. Using different ensembles of the four parameters, $N$ model states are created. These, together with the ERS-satellite observations, are used in the EnKF. See also figure 2.7.

### Actual evaporation
The variance of the actual evaporation data is used to create the ensembles of the actual evaporation. The variance is used as the standard deviation is 0.18 mm $\cdot$ d$^{-1}$, because otherwise the perturbations are expected to have very little effect on the model results. The ensembles are generated via formula 2.25.

$$E_i = E + N_i(0, \sigma_E^2), \ i=1,2,\ldots,N$$  \hspace{1cm} (2.25)

### Precipitation
To generate the perturbations for precipitation, the monthly averages of the rmse and the bias of the 1.0˚ grid from Nicholson et al. (2003) were used. First, the bias is subtracted from the original precipitation data set. Next, a random number is picked from a normal distribution with zero mean and a standard deviation of 1.34 mm $\cdot$ d$^{-1}$ and added to the precipitation data. See formula 2.26.

$$P_i = P - Bias + N_i(0, \sigma_P^2), \ i=1,2,\ldots,N$$  \hspace{1cm} (2.26)

### Saturated hydraulic conductivity
Perturbations of the saturated hydraulic conductivity are generated via formula 2.27:

$$K_{s,i} = K_s + N_i(0, \sigma_{K_s}^2), \ i=1,2,\ldots,N$$  \hspace{1cm} (2.27)

$^4$ Note that one row and one column are not updated, because the initial size of the study area was 81 by 161 gridcells. The combined size of the eight areas is 80 by 160 gridcells.
where the standard deviation is the same as in the sensitivity analysis. The standard deviation is 1.25 (in \(\text{ln}[\text{cm} \cdot \text{d}^{-1}]\)).

**Thickness of the first store**

To generate the ensembles for the thickness of store 1, a random number is drawn from a uniform distribution with zero mean and maximum value of 1 and minimum of -1 and then multiplied by the maximum deviation, which is set at 0.10 m. This is added to the original thickness. The minimum thickness is set at 0.02 m. The ensembles are thus created via formula 2.28:

\[
Z_{1,i} = \max(0.02, Z_1 + \sigma_{Z_1} \cdot \frac{U_i(-0.5,0.5)}{0.5}, i=1,2,\ldots,N)
\]  

(2.28)

**Observations**

As explained in 2.3.1, the ERS-satellite estimated degree of saturation data from TU Wien allow one to distinguish at least five soil moisture classes. If the total water capacity is divided by two times the rmse, a number of about five is obtained (Wagner et al., 1999). For this thesis this is translated as a distribution of the ERS-satellite estimates as is given in formula 2.29.

\[
m_{s,i} = m_s \cdot (1 + N_i(0,\sigma_{m_s}^2)), i=1,2,\ldots,N
\]  

where in a first run \(\sigma_{m_s}\) is set at 0.10 and in a second run this was set to 0.05.

Note that each grid cell is randomly perturbed for all the variables. This implies that it is assumed that the variables are spatially independent. This may not be entirely true when considering the saturated hydraulic conductivity if one looks at the soil map of West Africa (see appendix A).

**Number of ensembles**

The number of ensembles is set at thirty. More ensembles (50-100) could increase the performance of the data assimilation method (Reichle et al., 2002), but also increases the computation time. At a certain amount of ensemble members, the performance does not increase so much anymore. This amount of ensembles differ per case. With respect to this thesis, thirty ensemble members may be sufficient (Heemink and Steele, personal communication, 2008).

**Application of EnKF**

Prior using the EnKF, the N models are “warmed up” to create an initial condition, because the stores have no initial soil moisture content and no runoff is generated yet. The models are run for 20 years (five times for four years) with the perturbed inputs. The final model states after twenty years are used as initial conditions for the data assimilation process. Appendix F describes in more detail how the model is warmed up and explains why twenty years of warming up is enough to obtain an equilibrium result for the second store.

The application of the EnKF consists of three main parts, which are repeated every time-step:

✶ Translation of observations to a model variable;
✶ The update of model states with the translated observations;
✶ The model run.
The observations are given as a degree of saturation. The observation is required to be translated to a model variable. This is done via formula 2.30:

\[
S_{\text{obs}} = m_s \cdot S_c
\]

where \(S_{\text{obs}}\) is the soil moisture content according to the observation [m], \(m_s\) the degree of saturation [-], and \(S_c\) is the maximum storage [m].

The maximum storage is calculated as (van Beek, personal communication, 2007):

\[
S_c = s_c - s_{\text{res}} Z_1
\]

where \(s_c\) is the maximum saturation (\(m^3\cdot m^{-3}\)), \(s_{\text{res}}\) the residual saturation (\(m^3\cdot m^{-3}\)) and \(Z_1\) the thickness of layer 1 [m].

To update the model states, it is first checked which locations have an observation. Thereafter the \(N\) model state vectors are reduced to the size of the observation state vector. If, for example, there are fifty grid cells containing an observation, then the model state vectors are reduced from the initial \(N \times 81\) by 161 matrices to a fifty by \(N\) matrix. The EnKF calculates the updated model state matrices. These matrices are eventually stored and used as initial condition for the model runs of the next day.

If there are less observations (\(m\)) than the number of ensembles (\(N\)), then the Kalman Filter is not used and no update takes place. There is also no update in case \(H \cdot A \cdot Y' \equiv 0\) is not true. The update cannot result in a negative soil moisture content and also cannot exceed the maximum soil moisture content.

**Results analysis**

After the data assimilation is completed, a lot can be analyzed. Due to limited time for research, the analysis is focussed on the following:

- The update process;
- Runoff analysis of updated model.

**The update process**

The update process is analyzed by comparing the updated model states with the initial model states and the observations. First the difference between the initial state and the observation is computed, then the difference between the update and the initial state is computed. These comparisons show how the EnKF works in ‘practice’. The variance over time at a few locations is determined as well. It is expected that after an observation is available, the variance of the model states decreases and that the updated value is somewhere between the initial state and the observation.

**Runoff analysis of updated model**

The mean of the model results is computed. These results can be used by the routing model in order to simulate the resulting runoff. The runoff can be compared to the runoff of the original model and to the measured runoff data of the Black Volta at Sabari and the Oti at Bamboi.
3. Results

This chapter presents the results of the data analysis (3.1) and the data assimilation (3.2).

3.1 Data analysis

The data analysis consists of the waterbalances of the Senegal-, Volta-, and Niger river basin, a runoff analysis for two locations in the Volta basin and an overview of the correlations between different data sets for the three river basins.

3.1.1 Waterbalances

The waterbalances of the Senegal-, the Volta- and the Niger basin have been computed according to formula 2.3. The results are shown in table 3.1. The difference between the precipitation and actual evaporation is given in this table as well. The cumulative difference is plotted in figure G.1 of appendix G.

<table>
<thead>
<tr>
<th>Year</th>
<th>Difference $P-E_a$ (m)</th>
<th>Waterbalance $dS = P-E_a - Q$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Senegal</td>
<td>Volta</td>
</tr>
<tr>
<td>2003</td>
<td>0.11</td>
<td>0.31</td>
</tr>
<tr>
<td>2004</td>
<td>-0.07</td>
<td>0.17</td>
</tr>
<tr>
<td>2005</td>
<td>-0.07</td>
<td>0.10</td>
</tr>
<tr>
<td>2006</td>
<td>-0.18</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>-0.21</td>
<td>0.68</td>
</tr>
</tbody>
</table>

There is a small surplus of precipitation in all regions, but the waterbalance suggest large losses in the Senegal basin in 2005 and in the Niger basin for all years. Such big water losses due only to runoff are impossible. If the runoff is too large for the entire region (for each pixel), the soil moisture content is largely underestimated because less water can infiltrate into the soil (see also appendix C). However, this does not seem to be the case. It appears that only a small region in the Senegal and Niger basin is responsible for extremely high runoff. Figure G.2 of appendix G shows the average runoff over the four years. In both basins large flows are generated from very dry regions in the North. Extremely high and unrealistic peak flows occur twice (2005 and 2006) in the Senegal basin, once (2003) in the Volta basin and every year in the Niger basin, see figure 3.1. The peak flow in the Volta and the peak flow in the Senegal basin of 2006 have a short duration and do not contribute to large errors in the water balance. The other peak flows, however, cause large errors in the waterbalances of the Senegal and Niger river basin. It was found that the errors are caused by the routing model. The errors are a result of numerical instability in the routing scheme. It is currently investigated how these problems can be resolved (van Beek, personal communication, 2008). However, they are of little influence when comparing the measured discharge with the modelled for the Volta river (see the next paragraph).
Figure 3.1. Discharge at outlets of the Senegal river (purple), the Volta river (blue) and the Niger river (yellow), from 1/1/2003 to 31/12/2006. Values on the y-axis are given in m$^3$·s$^{-1}$.

**Discharge at Sabari (Oti)**

Figure 3.2. Average discharge (m$^3$·s$^{-1}$) per day at Sabari (Oti river). The modelled discharge (m$^3$·s$^{-1}$) is shown in blue, the measured (m$^3$·s$^{-1}$) in black and the cumulative difference (100 m$^3$·s$^{-1}$) is given in red.

**Discharge Bamboi (Black Volta)**

Figure 3.3. Average discharge (m$^3$·s$^{-1}$) per day at Bamboi (Black Volta river). The modelled discharge (m$^3$·s$^{-1}$) is shown in blue, the measured (m$^3$·s$^{-1}$) in black and the cumulative difference (100 m$^3$·s$^{-1}$) is given in red.
3.1.2 Runoff analysis
At Sabari, the discharge of the Oti is measured and at Bamboi, the discharge of the Black Volta is measured. The modelled discharges have been compared with the measurements for these sites, see figure 3.2 and figure 3.3 for respectively Sabari and Bamboi. The cumulative difference is also computed. In case no measurement was available, the cumulative difference is not changed. The following can be seen:

- The peakflows are relatively well captured by the model for the Oti, but not for the Black Volta. The modelled peakflows occur several days later than is measured;
- The higher discharges in the Black Volta have a shorter durations in the model compared to the measurements;
- The cumulative difference between the modelled and measured discharge is increasing. For the Black Volta this is mainly due to the overestimation by the model of the baseflow of the Black Volta. At the Oti the model simulates a tail after a period of high discharge, while this was not measured;
- The runoff error in the Volta described earlier, is visible in both figures. The peakflows of ~4000 m$^3$·s$^{-1}$ occur at both locations at about the same moment as the peakflow of the Volta shown in figure 3.1.

Table 3.2. The average significant correlation between different datasets for the Senegal-, Volta- and Niger riverbasin for two periods. In case over ~25% of the watershed contains significant correlation values, the correlation is shown in bold. $P$ is the precipitation, $E_a$ the actual evaporation, $m_s$ is the degree of saturation according to the ERS-satellite, $S_1$ and $S_2$ are the soil moisture content of respectively the first and second store.

<table>
<thead>
<tr>
<th>Watershed/Period</th>
<th>Average Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P - E_a$</td>
</tr>
<tr>
<td>Senegal summer</td>
<td>0.09 0.55 0.14</td>
</tr>
<tr>
<td>Senegal winter</td>
<td>0.27 0.51 0.24</td>
</tr>
<tr>
<td>Volta summer</td>
<td>-0.27 0.54 0.12</td>
</tr>
<tr>
<td>Volta winter</td>
<td>0.26 0.57 0.35</td>
</tr>
<tr>
<td>Niger summer</td>
<td>0.06 0.59 0.49</td>
</tr>
<tr>
<td>Niger winter</td>
<td>0.25 0.53 0.26</td>
</tr>
<tr>
<td>Average</td>
<td>0.11 0.55 0.27</td>
</tr>
</tbody>
</table>
3.1.3 Correlations

The ERS-satellite estimated degree of saturation ($m$) correlate well with the modelled soil moisture content for the first store ($S_1$), see table 3.2. This table shows the average significant correlation between different data sets per watersheds for summer and winter periods. The satellite estimates correlate also, but to a lesser extent, with the soil moisture content of the second store ($S_2$). The correlation between the ERS-satellite estimated degree of saturation and the modelled soil moisture content of the first layer is higher in summer than in winter. The correlations between the other data sets are in some cases (e.g. $P$ and $m$) high, but only represent a small part of the watershed. In appendix I, the correlations are mapped for each watershed. Only the significant correlations are shown, meaning that if the significance is less than 0.05, the correlation is not shown.

![Mean Soil Moisture Content Volta Basin (2003-2006)](image)

**Figure 3.4** Mean soil moisture content at the Volta basin compared to the mean soil moisture content according to the observations, after translated to a soil moisture state of the first store. The precipitation and actual evaporation are shown in the second plot.

Figure 3.4 shows the ERS-satellite measurements over the Volta basin, after those have been translated to a soil moisture content of the first layer according to formula 2.30. The modelled soil moisture content of the first store, the precipitation and actual evaporation are also shown. The modelled soil moisture content follows the ERS-satellite measurement quite nicely. Similar results are obtained for the Niger and Senegal basin, see figure H.1 and H.2 of appendix H. Note that the mean of the ERS-satellite measurements may not always be representative for the entire river basin, because the mean is only based on the locations (pixels) where measurements are available. In case large spatial differences in soil moisture content exist in the longitudinal direction, the mean is particularly not representative, because measurements lie in a single strip (the satellite path) stretching from south to north or vice versa and not from west to east.

It is assumed for this particular thesis that the true soil moisture state lies somewhere between the modelled state and the observation. In case it is assumed that the hydrological model presents the truth, the rmse of the ERS-satellite measurements can be calculated. The rmse is calculated for each watershed, see table 3.3. In Wagner et al. (1999) it was stated that in case the total water capacity was divided by two times the rmse, the result was about five. Therefore they stated that the ERS-satellite estimated soil moisture data from TU Wien
allowed one to distinguish at least five soil moisture classes. In case of this thesis, the maximum storage capacity of the first layer varies in West Africa from ~0.04 m to ~0.15 m. Dividing these values by two times the rmse results in numbers varying from 1.82 to 8.33, with a mean is of 5.08. The result resembles the findings in Wagner et al. (1999).

Table 3.3. Root mean squared error (RMSE) of the mean observation compared to the mean of the model. The mean of the observation and model are given as well.

<table>
<thead>
<tr>
<th>Watershed</th>
<th>RMSE (m)</th>
<th>Mean Model (m)</th>
<th>Mean ERS (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senegal</td>
<td>0.010</td>
<td>0.016</td>
<td>0.023</td>
</tr>
<tr>
<td>Volta</td>
<td>0.011</td>
<td>0.034</td>
<td>0.031</td>
</tr>
<tr>
<td>Niger</td>
<td>0.009</td>
<td>0.026</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Some more information can be extracted from the figures 3.4, H.1 and H.2. It appears that the onset of the increase in soil moisture content in the Volta basin occurs several days earlier in the model compared to the ERS-satellite measurements. At the Niger river basin the same result is visible, but to a lesser extent. For the Senegal basin the onset is at almost the same moment as the onset according to the ERS-satellite measurements.

In summer, the ERS-satellite measures wetter soils in the Volta- and Niger basin than the model simulates. The ERS-satellite measures slightly drier soil conditions in the winter for both basins compared to the modelled soil moisture. At the Senegal basin, the model shows drier soil moisture states compared to the ERS-satellite measurements throughout the years.

### 3.2 Data assimilation

This paragraph presents the results of the sensitivity analysis (3.2.1) and the results obtained from the data assimilation (3.2.1).

#### 3.2.1 Sensitivity analysis

In table 2.2 it was shown with what values four variables are increased and decreased in order to determine the sensitivity of the model to the variables. The four variables that have been changes are actual evaporation ($E_a$), precipitation ($P$), saturated hydraulic conductivity ($K_s$) and thickness of the first layer ($Z_1$).

In appendix I the difference between the original and final state is shown for each change of variable value. Table 3.4 states what is generally the result of the change of the variable value.
Table 3.4. The effect the change of variable value has on the model state.

<table>
<thead>
<tr>
<th>Variable change</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_a$ increase</td>
<td>The soil moisture content of the first store is much lower for a small part in the North and the soil moisture content of the second store is much lower in small parts of the Senegal basin and the Niger basin.</td>
</tr>
<tr>
<td>$E_a$ decrease</td>
<td>The effect is almost the inverse of the effect when $E_a$ increases.</td>
</tr>
<tr>
<td>$P$ increase</td>
<td>The soil moisture content of the first and second store is much higher for a large area in the North. The soil moisture content of the second store is also much higher at the mountainous regions in the Southwest and at large parts of the Niger basin.</td>
</tr>
<tr>
<td>$P$ decrease</td>
<td>In the North the soil moisture content of the first store is much lower. The soil moisture content of the second store is much lower for the mountainous region in the Southwest and for some small parts of the Niger basin.</td>
</tr>
<tr>
<td>$K_s$ increase</td>
<td>The soil moisture content of the first store is a little bit lower for most of the study area, except for the mountainous region in the Southwest where it remains unchanged compared to the original state. In this region the soil moisture content of the second store, however, is much lower.</td>
</tr>
<tr>
<td>$K_s$ decrease</td>
<td>The soil moisture content of the first layer is much lower for most of the study area with exception of the region north of Lake Volta. In the South, the soil moisture content of the second store is much lower.</td>
</tr>
<tr>
<td>$Z_1$ increase</td>
<td>The soil moisture content has not changed. The degree of saturation is lower for the whole region.</td>
</tr>
<tr>
<td>$Z_1$ decrease</td>
<td>The soil moisture content has not changed. The degree of saturation is higher for the whole region.</td>
</tr>
</tbody>
</table>

3.2.2. Data assimilation results

Whenever a measurement is available, in most cases in minimally one of the eight areas updates are made. The updates lie in between the initial (non-updated) model states and the observations. Figure 3.5 shows the average differences between the initial model state and the observations for 2003-2005 for the entire study area. The satellite observes a much wetter soil in the South compared to the model and a drier soil in the North. The wetland in the Niger river basin is also clearly visible. According to the satellite observations, this area is much wetter than the model simulates. Figure 3.6 shows the difference between the original non-assimilated model result and the ERS-satellite measurements. Clearly, the updated model deviates a lot more from the satellite measurements than the original model. This was not expected because the data assimilation should lead to smaller differences.

The average values of the updates, which are calculated as the average differences between the updated model states of the first store and the initial model states, are shown in figure 3.7. The eight different update areas can be distinguished. The bottom right area is never updated because the ERS-satellite has no coverage in this area. The area north and west of that have high differences between the updates and the model states. This may be due to the low frequency of observations (see figure 2.5). In addition the area in the Southwest has little
updates, due possibly to the few gridcells containing land. The requirements for when an update is made may not be reached that frequent in this area. The updates in the remaining areas are relatively small, with exception of the South of the upper left area.

Figure 3.5. The average difference between the ERS-satellite observations and the modelled soil moisture state of the first store for 2003-2005 (m).

Figure 3.6. The average difference between the ERS-satellite observations and the original modelled soil moisture state of the first store for 2003-2005 (m).

Figure 3.7. The average difference between the updated model soil moisture state and the initial model soil moisture state state of the first store for 2003-2005 (m).

After an update is made, the variance between model states is reduced. Figure 3.8 shows the variance of the model state before and after an update for the Senegal basin for two different days at which a satellite measurement is available for that region. At day 231, the variance
decreases for the whole watershed, but at day 234 the variance is only decrease in the West. The variance remains unchanged in the East because this region lies in a different update area for which no update is made because it does not fulfil the requirements stated at the end of paragraph 2.4.

The variance over time at certain locations are given in figure 3.9 and figure K.1 of appendix K. The moments at which an observation is available for that location are also shown in the figure. The variance of the model state is reduced when an update is made. Note that the initial variance of the model state is already very small.

![Variance of model states at 231](image1)

![Variance of adjusted model states at 231](image2)

**Figure 3.8.** The variance of the model state before and after an update at different days at which measurements are available.

![Variance at 10N, 2.5W](image3)

**Figure 3.9.** The variance of the model state at 10°N, 2.5°W. The observations at this location shown with dots, the continuous line represents the variance before the updates are made and the dotted line the variance after the update is made.
Figure 3.10. Mean soil moisture content of the updated model at the Volta basin between 13/08/2003 - 31/12/2003 compared to the mean original soil moisture content and the mean soil moisture content according to the observations. The precipitation and actual evaporation are shown in the second plot.

Figure 3.11. The runoff generation in the adjusted model at day 250 on the left and the runoff generation of the original model at the same day. The colorbar ranges from low runoff (purple) to high runoff (red).

The mean of the updated model states of the Volta basin has been plotted in figure 3.10 together with the mean of the original non assimilated model state and the mean ERS-satellite measurements after these have been translated to soil moisture contents. The difference between the adjusted model state and the observations is larger than the difference between the observations and the original model state. Figures K.2 and K.3 of appendix K show similar results for the Senegal - and Niger river basin respectively.

The adjusted model results of the runoff show why the results are larger. Figure 3.11 shows the generated runoff of day 250 for West Africa. At many locations, much more runoff is generated. As more runoff is generated, less water will infiltrate into the soil. The soil becomes drier and therefore the model predicts lower soil moisture contents. It is checked if the perturbed ensembles of precipitation, saturated hydraulic conductivity and the thickness of the first store cause these errors. One can expect that if the soil layer becomes thinner, its storage capacity decreases and thus saturated conditions occur more frequent. When soil is saturated, there is more direct runoff and less water can infiltrate into the soil. It is found however that the thickness of the first store was not linked to the maximum storage capacity. Although the maximum storage capacity, used by the model, is based on the thickness of the first store, its value is not linked inside the model. As a consequence, the perturbations of the thickness of the first store do not lead to changes in model states. It only leads to larger
variance in the ERS-satellite measurements because these are translated to a model state using formula 2.30, which includes the thickness of the first store.

The method used to create the perturbed precipitation ensembles results in negative rainfall for large areas in West Africa between 2003-2006, because no boundaries were set to the ensemble creation. The negative precipitation does not lead to larger runoff, but may lead to complications inside the model. It is therefore preferred to prevent the creation of negative precipitation.

The method used to create the perturbed ensembles for the saturated hydraulic conductivity resulted in unrealistic values and consequently in extremely large runoff. It is found that no boundaries are set to the creation of the perturbed ensemble members. Since the values are stored as natural log-numbers, a boundary must be set to prevent negative values.

Next it is checked if adjusting the ensemble creation of the saturated hydraulic conductivity and precipitation leads to better results. The minimum and maximum conductivity is set at respectively the minimum and maximum value present in the study area and the precipitation cannot become negative. It is found that the results improved, but still deviate more with the ERS-satellite measurements than the original model. In the wet season, the results are quite similar to the original model result, but in the winter, the soil is much wetter, see figure 3.12. It is assumed that the adjusted method to create the perturbed precipitation ensembles results in generally more rainfall in the dry season. All perturbations leading to negative precipitation are set to zero, while those leading to more positive precipitation are kept. In general more water is added to the system and thus the soil becomes wetter. This becomes clearer when the precipitation from day 320 to 365 shown in figure 3.10 is compared to figure 3.12. In figure 3.12 the mean of the precipitation over the Volta basin is minimally 1 mm/d, while in figure 3.10 there are many days without rain.

Figure 3.12. Mean soil moisture content of the updated model ensembles at the Volta basin between 13/08/2003 - 31/12/2003 compared to the mean original soil moisture content and the mean soil moisture content according to the observation ensembles. The precipitation and actual evaporation are shown in the second plot.
4. Discussion

The TU Wien data is based on the assumption that the lowest backscatter value measured over a long timeseries represent dry soil and the highest backscatter value saturated soil. Such conditions may occur more frequent for the top layer with a maximum thickness of five centimeter than for thicker layers. In case the ERS-satellite actually measures the degree of saturation of a layer which is thicker than five centimeter, this assumption may not be correct. At very dry locations in the North of the study area saturated conditions may almost never occur, whereas in the South completely dry situations are very rare. The range between $\sigma_{\text{dry}}$ and $\sigma_{\text{wet}}$ may thus not represent the true range. In the North this can lead to overestimation - and in the South to underestimation of the degree of saturation. It can be expected that for the region between 8°N and 14°N completely dry and saturated situations occur at least once within a long timeserie, due to the large seasonal variations.

The correlation between the ERS-satellite estimated moisture states and the soil moisture content of the upper two stores is generally high. The satellite observations correlate particularly well for the top layer of the model. The correlation between the satellite observations and the second store could imply that the satellite observes the soil moisture content of the second store as well, but could also be explained by the dependence of the moisture content of the second store to the soil moisture content of the first store.

This research shows that it is possible to update soil moisture states with ERS-satellite measurements using the Ensemble Kalman Filter, but there are many issues that need to be addressed before proper results are obtained. The sensitivity analysis presented in this particular thesis shows how the soil moisture state of the first layer changes in case the actual evaporation, precipitation, saturated hydraulic conductivity or thickness of the first layer are changed. It is not checked what result these changes have on the runoff generation. Checking the effect the change has on the runoff generation would have shown that the creation of perturbed ensembles of saturated hydraulic conductivity lead to unrealistic runoff values. As explained at the end of chapter 3, this discovery was made only after the data assimilation was performed.

It is found that the method used to create the ensembles of perturbed saturated hydraulic conductivity and precipitation lead to unrealistic values. Boundaries can be set to prevent the creation of unrealistic values, but may introduce another problem such as the creation of a bias. In 5.2. recommendations are made with respect to this matter.
5. Conclusions and recommendations

5.1 Conclusions

In this report, the application of the Ensemble Kalman Filter (EnKF) to assimilate the remote sensed soil moisture data is discussed. The ERS-satellite estimated degree of saturation are used to update the PCR-GLOBWB model states of West Africa. The application of the EnKF has not led to more insight about what spatial features and hydrological processes influence the accuracy of the ERS-satellite estimated soil moisture states. However, some interesting progress is made.

The ERS-satellite estimated degree of saturation correlates well with the modelled soil moisture states of the upper stores. It correlates especially well with the soil moisture states of the first store. The correlation is higher for wet (summer) periods than for dry (winter) periods.

The modelled soil moisture state of the first store has been compared to the ERS-satellite measurements, after these have been translated to a soil moisture state. The mean of the modelled soil moisture states follow the mean of the satellite measurement very well for all three watersheds. The rmse of the satellite measurements compared to the model states varies from 0.009 m for the Niger river basin to 0.011 m for the Senegal river basin. The maximum storage capacity of the first layer varies in West Africa from ~0.04 m to ~0.15 m. Dividing these values by two times the rmse results in numbers varying from 1.82 to 8.33, with a mean of 5.08. The result resembles the findings in Wagner et al. (1999). Wagner et al. (1999) states that the method used to determine the degree of saturation allows one to distinguish at least five soil moisture classes, because if the total water capacity is divided by two times the rmse, then a number of about 5 is obtained.

The runoff results of the original -non assimilated- model are compared to the measured runoff at Sabari (Oti) and at Bamboi (Black Volta). The model overestimates the runoff at both locations over the period 2003-2006. At Sabari it is caused by a simulated tail, which is not measured, after each period of larger flow. At Bamboi the baseflow is largely overestimated. The routing model is able to capture the peak runoffs at Sabari well, but not at Bamboi. It is also found that there are some significant errors in the routing model that lead to very large runoff generation in the Senegal river in 2005 and in the Niger river every year.

The application of the EnKF does not lead to better results. It is expected that this is mainly due to the method used to create the model ensembles. The creation of the perturbed ensembles for saturated hydraulic conductivity leads to unrealistic values at different places. This causes very large runoff and drier soil moisture states. For the creation of the perturbed ensembles, no boundaries were set to the saturated hydraulic conductivity values. Neither were boundaries set to the creation of the perturbed precipitation ensembles, which resulted in negative rainfall at different locations at different moments. Lastly, it is found that the changes made to the thickness of the first soil layer have no effect on the soil moisture content of the first layer, because the thickness of the layer is not linked to the maximum storage capacity of this layer.
A second run is performed to check if the model results are improved in case boundaries are set to the creation of precipitation and saturated hydraulic conductivity. The results improve, but still deviate more with the ERS-satellite measurements than the original model. In the wet season the updates are very small and in the dry season the soil moisture content is too high. It is expected that the adjusted method used to create the ensemble members for precipitation introduces a bias, because the perturbations leading to negative precipitation are set to zero and thus only those perturbations leading to more positive precipitation are kept, which lead to wetter soil conditions.

5.2 Recommendations

In 5.2.1 it is described what steps must be taken to improve the results from the data assimilation presented in this report. Some recommendations concerning future studies are given in 5.2.3.

5.2.1 Recommendations for this study

The following steps need to be taken to improve the results from the data assimilation presented in this report:

- Start with simple methods to create ensembles;
- Link thickness of the first store to its maximum storage capacity;
- Perform an enhanced sensitivity analysis;
- Update whenever measurements are available;

This paragraph concludes with a recommendation concerning the analysis of the runoff results.

**Start with simple methods to create ensembles**

To get proper results from the data assimilation, it is recommended to start simple and increase complexity step by step. This way, more insight is obtained about the consequence of certain decisions. One of the most important decisions to be made is to determine how to create the model ensembles. It is therefore recommended to simplify the method used to create the model ensembles for this particular thesis.

As explained earlier, there is literature about the creation of model ensembles. However, it is found that more attention to this topic is of importance for this research and other data assimilation studies. In general, to create model ensembles, it is important to determine the distribution of the uncertainty of the model variable (e.g. normal distribution, uniform distribution, exponential distribution), the uncertainty bounds and to determine whether the uncertainty is spatially and temporally dependent or not. The distribution of the uncertainty determines within what range the variable is required to be perturbed to represent this distribution. If the uncertainties are spatially or temporally dependent, the perturbations are preferred to be spatially or temporally dependent as well.

Information on the distribution of the uncertainty of a model variable is often not available or is very limited. Whether the uncertainties are spatially or temporally dependent is usually unknown as well. In this particular thesis, none of the above is known for the actual evaporation data and the thickness of the first store. Only limited information is available on the distributions of the uncertainty of precipitation data, saturated hydraulic conductivity data and remote sensed soil moisture data.
To summarize, too little is known about all the uncertainties of all perturbed model variables. As little is known, it is suggested to simplify the method to create the model ensembles for this particular thesis. The method can be simplified by perturbing the different model variables, which are used to create the different model ensembles, by a single value or factor according to formula 5.1 and 5.2 respectively, instead of perturbing each gridcell independently. The latter is done in this thesis because it is assumed that the uncertainties are spatially independent and that the results from the data assimilation contain more information as it results in more spatial variation.

\[
Y_i^* = Y_i + \sigma_i \cdot U(-1,1), \quad (5.1)
\]

\[
Y_i^* = Y_i \cdot (1 + \sigma_i \cdot U(-1,1)), \quad (5.2)
\]

where \(Y^*\) is the perturbed model variable, \(Y\) is the initial model variable, \(\sigma\) is the maximum deviation and \(i\) stands for the model variable (e.g. precipitation). \(U(-1,1)\) stands for a uniform distribution between -1 and 1 from which a random number is drawn.

Using simpler methods to create the model ensembles, one must be careful not to create unrealistic values for the model variables. To prevent the creation of unrealistic values, boundaries can be set to the creation (e.g. precipitation cannot become negative). Note, however, that the boundaries may introduce a bias. For example, if precipitation is perturbed through adding or subtracting a single value and a boundary is set such that precipitation cannot become negative, a bias is created. As there are many days without rain, the perturbations that lead to negative precipitation are reset to zero and thus only those perturbations leading to higher precipitation are kept. To overcome such bias, the negative precipitation could be added to the actual evaporation data (van Beek, personal communication, 2008), but then there is still a possibility that the actual evaporation could become unrealistically high. An alternative way of overcoming the bias is possibly by subtracting the absolute of the negative precipitation from the precipitation on other days. It remains yet unknown what consequence that could have.

Instead of perturbing with a single value, the perturbations can also be created by multiplication with a single factor so that this boundary is not required (assuming the factor is higher than 0). However, the perturbations are limited to those areas and moments at which a signal is present. In dry periods the such perturbed precipitation ensembles lead to smaller variations in the model ensembles and thus updates are expected to become smaller. The model is “penalized” in case it rains, and less when it does not. This is not justifiable in case of the TRMM rainfall product, because this product performs bad for drizzle or light rain events (NLR Monterey, 2008). Nevertheless, creating the perturbed precipitation ensembles through multiplication with a single factor is less complicated than through perturbing with a single value.

The creation of the perturbed ensembles for actual evaporation deals with similar issues as for precipitation. The actual evaporation cannot become negative on a daily timescale. Therefore it is recommended to perturb the actual evaporation with a single factor as well. For the creation of the perturbations of the saturated hydraulic conductivity and the thickness of the first soil layer, using a factor may be more suitable, because the values cannot become negative.

To conclude, it is recommended to first create the perturbed ensembles for all four variables (precipitation, actual evaporation, saturated hydraulic conductivity and the thickness of the first soil layer)
through multiplication with randomly picked factors. How the factors are chosen is arbitrary as little information is available on the uncertainty of the data. It is assumed here that the precipitation is the most accurate data product and that the thickness of the first soil layer and the ERS-satellite measurements are the least accurate. According to De Weirdt et al. (2006) the accuracy of the actual evaporation data is about 10%. In Wagner et al. (1999) it is concluded that the ERS-satellite data allows one to distinguish at least five soil moisture classes. This is interpreted as an accuracy of 20%. The accuracy of the saturated hydraulic conductivity is assumed to be in range of the actual evaporation. To start, the factors can be chosen within the following range:

- Precipitation: ±5.0% ;
- Actual evaporation: ±10.0% ;
- Saturated hydraulic conductivity: ±10.0% ;
- Thickness of the first soil layer: ±20.0% ;
- ERS-satellite estimated degree of saturation: ±20.0% .

Formula 5.2 can thus be used to create the perturbations. In case of precipitation, $\sigma$ is set to 0.05. Later it can be investigated what results the different distributions have on the model ensembles, and what are the consequences of creating perturbations with a single value.

**Link thickness of the first store to its maximum storage capacity**

It is found in this research that the thickness of the first store was not linked to its maximum storage capacity. The maximum storage capacity was based on the initial thickness of the first store, but did not change in case the thickness was changed. As a result, the variability of the model remained unchanged while the variability of the (translated) ERS-satellite soil moisture estimates increased. As a consequence, the updates remain small.

It is recommended to link the maximum storage capacity of the first layer to its thickness, because then a more comprehensive comparison between layer thickness and satellite measurement can be made. In the latest version of the model, this link is made, but it has not yet been implemented into the data assimilation scheme.

**Perform an enhanced sensitivity analysis**

In order to create realistic ensembles and to analyze the results from the data assimilation, it is important to have a clear understanding of the influence of the different perturbed variables on the model outcomes. It is therefore suggested to perform a enhanced sensitivity analysis. This analysis:

- must be performed for minimally one full year, so that the seasonal changes are included in the analysis;
- should be performed at the same time step of the model so that effects of variations between different timesteps can be addressed;
- should include the sensitivity of all model outcomes (soil moisture content of the first, second and third store, the drainage components from the first, second and third store and the total discharge) to the different variable values and the combinations of the different variable values.

**Update whenever measurements are available**

In this thesis updates are made in case $H \cdot A \cdot Y' = 0$ is true (see 2.4.2), because it solves the problems with singular matrices. However, according to Evensen (2003), this is only
applicable if the number of observations \((m)\) is large compared to the number of ensembles \((N)\), \(m\gg N\). It is recommended to apply this restriction into the data assimilation scheme. To make optimal use of the observations, it is also recommended to investigate how to deal if \(m\gg N\) is not true. In case of \(m\ll N\), the solution presented by Evensen (2003), see formula 67 of that paper, may be applied. For all other cases, a covariance matrix \(R\) can be used instead of \(\Upsilon \Upsilon^{-1}\) (see formula 2.17). Matrix \(R\) consists of zeros and has the uncertainties, used to create the ensembles of the observations, on the diagonal. However, one must be careful when implementing this solution, because it was found in this thesis (although not presented in this report) that this solution results in much more noisy updates than the method eventually used in this thesis.

**Compare non-routed runoff with measured runoff**

Due to some significant errors found in the routing model, it is recommended to compare non-routed runoff instead of routed runoff with the measured runoff. Since the runoff is not routed, it is advised to compare the cumulative non-routed runoff to the cumulative measured runoff. Note that the non-routed runoff of all gridcells upstream of the measurement location must be added to this calculation, because the runoff is not handed down to the next gridcell. It is advised to consult Rens van Beek how to do this in a convenient way with PC-Raster.

**5.2.2 Recommendations for future studies**

In this paragraph some recommendations are given with respect to future studies on this study. These recommendations are:

- Investigate how perturbations in vegetation fractions can be included;
- Assimilate more model states;
- Increase number of ensembles and/or add spatial dependency;
- Use a more simplified model and/or focus on comparison.

![Figure 5.1 Short stack ratio](image)

**Investigate how perturbations in vegetation fractions can be included**

In the model, the vegetation cover changes each month. The ratio between short- and tall stack vegetation, however, remains the same. The ratio of short stack vegetation is shown in figures 5.1. The ratio of short- and tall stack vegetation are based on the NDVI (from MODIS) of 1995-1996 (personal communication: van Beek, 2008). The type of vegetation determines how much evaporation can occur from the different stores.

It is recommended to include perturbations in the vegetation fractions in order to investigate whether the ERS-satellite data is influenced by vegetation. As the fractions are based on a NDVI map of a single year, it is expected that the uncertainty may be quite significant (>10%). This is because one must be cautious when making decisions based on NDVI measurements, because these measurements are highly sensitive to other factors than
vegetation as well. The measurements are influenced by the atmospheric vapor, by clouds and by the wetness of the soil itself.

**Assimilate more model states**

In this particular thesis, only one model variable is assimilated, namely the soil moisture content of the first store. If an update is made to the soil moisture content of the first store, no changes are made to the soil moisture content of the second and third store and the local drainage components. There are, however, some relations to be expected between the different stores and local drainage components. It is therefore recommended to include other model variables in the assimilation process so that these variables are also updated. Note that including more model states requires more disk space and memory. One can for example start by updating the soil moisture content of the second store and the total drainage as well.

**Increase number of ensembles and/or add spatial dependency**

It was found in this thesis (although not presented in this report) that performing updates with the usage of a covariance matrix $R$, as is explained in fourth recommendation of 5.2.1, leads to more noisy results. One expects that the updates are larger nearby the measurement location and less significant at locations further away from the measurement. Increasing the number of ensembles often leads to better results (Reichle et al., 2002), and may therefore result in less noisy results. However, using more ensembles requires a lot more disk space, memory and computation time.

One can also include spatial dependency to the updates, which state that the updates become less significant if they are further away from the measurement location. In this way, it is not necessary to increase the number of ensembles to overcome the noisy results.

If sufficient disk space and memory is available, it is suggestive to increase the number of ensembles. Otherwise, it is recommended to investigate how the spatial dependency can be included.

**Use a more simplified model and/or focus more on comparison**

It is found that assimilating remote sensed soil moisture data into the PCR-GLOBWB model is quite laborious, because many files need to be updated, copied and renamed at each timestep. This is needed because many processes are simulated in the hydrological model. It is expected that a simpler model will simplify the application of the data assimilation. This is in line with the first recommendation presented in 5.2.1; to start simple and thereafter increase the complexity.

The goal of this thesis was to determine what spatial features and hydrological processes influence the accuracy of the ERS-satellite estimated degree of saturation. The benefit of the current hydrological model is that many hydrological processes are simulated and that it contains several spatial features (e.g. mountains, open water, different soils). Conducting a more enhanced comparison between the model and the satellite measurements, could possibly lead to quite some interesting information already. The comparison should not only include the correlations and the differences between the model results and the ERS-satellite measurements, but also check whether there is a relationship between the model variables (e.g. hydraulic conductivity, elevation, vegetation cover) and the differences between the model results and satellite measurements.
Reference

Akker, van den, C., Boomgaard, M.E., 2001, Hydrologie, Dictaat CThec3010, Delft University of Technology, the Netherlands


Bastiaansen, W.G.M., Molden, D.J., Makin, I.W., 2000, Remote sensing for irrigated agriculture: examples from research and possible applications, Agricultural Water Management 46, 137-155

Batjes, N.H., 2005, Isric-Wise Global Dataset Of Derived Soil Properties On a 0.5 By 0.5 Degree Grid (Ver. 3.0) Isric Report 2005/08.

Beek, van, R., 2007, The macro-scale hydrological model PCR-GLOBWB, unpublished, University of Utrecht, the Netherlands


Beek, van, R., 2008, personal communication, meeting in April 2008.


De Weerdt, M., Rosema, A., Weimin, Z., Foppes, S., Chunqin, W., Dong, D., 2007, Evapotranspiration monitoring and validation in the Wei River basin and upper Yellow River Basin, 3rd International Yellow River forum


Assimilation of remote sensed soil moisture data into a hydrological model using the EnKF

NIVR, Vrije Universiteit van Amsterdam, 2007, Soil moisture updating for Flood Early Warning System of the River Rhine Basin
Moens, M., 1989, Using Neutron and Gamma probe to measure soil physical properties, MSc Thesis, Agricultural University of Wageningen, the Netherlands
Scipal, K., Scheffler, C., Wagner, W., 2005, Soil moisture runoff relation at the catchment scale as observed with coarse resolution microwave remote sensing, Hydr. and Earth Sys. Sci., 9, 173-183


Assimilation of remote sensed soil moisture data into a hydrological model using the EnKF
Appendix
Assimilation of remote sensed soil moisture data into a hydrological model using the EnKF
Appendix A: West Africa

Figure A.1. Soil map of West Africa. From North to South: yellow areas are desert areas, consisting of sand. The grey areas in the North are rock and rock debris areas combined with sand. The (light) brown areas below the desert region represent brown soils. The large dark blue-grey area in the west represents rocks and lithosols. The West coast is a mixture of fersiallitic tropical soils. The river beds and Niger wetland contain mostly young mineral hydromorphic soils. The patches of brown areas in the South represent lithic- and eutrophic brown soils. The area between the brown soils and the South coast consists of mainly fersiallitic soils. The orange and red areas in the Southwest represent yellowish-brown ferrallitic soil. In the South, these areas consists of ferrallitic red soils.
Figure A.2. Average precipitation in West Africa [mm $\cdot$ d$^{-1}$].

Figure A.3. Average actual evaporation in West Africa [mm $\cdot$ d$^{-1}$].
Figure A.4. The difference between the average precipitation and actual evaporation per day in West Africa [mm·d⁻¹].
Assimilation of remote sensed soil moisture data into a hydrological model using the EnKF
Appendix B: Soil moisture theory

Soil moisture content is the amount of water held in the ground. As explained in the introduction, soil moisture plays a very important role in environmental processes. This paragraph describes the underlying theory on soil moisture in more detail and is mainly based on Ward (1975).

A part of the precipitation, that reaches the ground surface, flows over the surface as overland flow (runoff) and a larger part is absorbed by the surface layers and may percolate to the groundwater table. Water moves out of the soil, either by evaporation to the atmosphere or by percolating to the lower layers and eventually to the groundwater table or by adding to the surface runoff. The rate of change of the soil moisture content in the soil can be described by the following formula:

\[ \Delta s = f + c - d - e \gamma \Delta v, \]  

where \( \Delta s \) is the rate of change of soil moisture content, \( f \) is the rate of infiltration, \( c \) is the rate of addition of moisture through capillary rise, \( d \) is drainage or percolation, \( e \) is evaporation and \( \Delta v \) is the rate of addition or loss of water vapour.

The rate of change of water vapour is normally very small. However, in very dry soils with a considerable temperature gradient, the rate of change of water vapour can possibly have a significant effect on soil moisture contents. Even though the North of the study area is very dry, this term is not included in this thesis for estimating the soil moisture content. It is, however, kept in mind that this term may play a significant role in the dry areas of the study area.

![Moisture zones during infiltration.](image)

Movement of soil moisture can occur through the soil. In order for soil moisture to move, energy is needed. Soil water contains kinetic and potential energy, just like any other body. The
kinetic energy can be neglected because soil water moves very slowly in general. The potential energy of the soil water (the soil moisture potential) determines how much energy is needed for the soil water to move. The potential is dependent on the gravitational potential, the osmotic potential and most important the pressure potential. In case the pressure potential is negative, capillary rise (suction) to higher layers takes place. Soil moisture may move because of gravity- and suction forces, and because of the vapour pressure and temperature gradients. The unsaturated movement may be in any direction because gravity is not necessarily the dominant force. However, only vertical movement is regarded for this particular thesis.

During heavy rainfall (or irrigation!), the top layer of the soil will be saturated; the saturated zone, see figure B.1. Below the saturated zone is the transition zone where the moisture gradient remains fairly constant. The moisture content remains constant in the transmission zone, which is below the transition zone. As infiltration (due to rainfall) continues, the transmission zone becomes longer and the wetting zone and wetting front (below the transmission zone) move further into the soil.

When infiltration stops, redistribution of the moisture in the soil begins. The wetting front may still continue downwards, but the moisture content in the top layers will decrease as there is no longer infiltration while percolation and evaporation are continuing. The movement of moisture will decrease until the soil is at field capacity. Field capacity is the amount of water held in the soil after excess water has drained away and percolation to lower layers has nearly stopped. While the percolation may stop, evaporation may still continue, resulting in the drying out of the land. Eventually the soil could reach the wilting point. The wilting point is the moment from which permanent wilting of plants occur. When the soil is at wilting level, the groundwater table can drop because of capillary rise in the capillary fringe (the area just above the groundwater table). Due to the capillary rise, the moisture content in the layers above the groundwater table increases. In dry areas where evaporation exceeds the rainfall, it is likely that there is a predominant upward movement of moisture.
Appendix C: The PCR-GLOBWB model

Chapter 2.2 briefly described the PCR-GLOBWB model. In this appendix it is described in more detail how water flows between stores and how runoff is generated. The second part of this appendix presents some maps of some parameters of the model and explains how the local drainage map is created.

Waterflows and runoff

The soil compartment is divided into three stores in the hydrological model. From each store there is a drainage component adding to the total runoff of the river (open water surface). Any drainage from these stores is delivered to the open water surface within one time step. Routing of water through the drainage network is prescribed by the local drainage direction map (LDD) which allows flow in 8 cardinal directions over 360 degrees.

The direct runoff is determined via the Improved Arno Scheme. Runoff can only occur when precipitation reaches the surface and if the net precipitation and the stored moisture combined exceed the minimum storage capacity. The net precipitation is converted into direct runoff once the pervious area is completely saturated. The infiltration is equal to the difference between the net precipitation and direct runoff. When the infiltration rate exceeds the saturated hydraulic conductivity of the first layer, the infiltration excess is passed on to the direct runoff. If the total infiltration exceeds the storage capacity of the first layer, it is handed down to the second store.

Water moves through the soil proportional to the unsaturated hydraulic conductivity:

\[ k(\theta_e)_{i,j} = k_r(\theta_e)_{i,j} \cdot k_s; \]

where the indices \( i \) and \( j \) stand respectively for the land cover and soil layers, \( \theta_e \) is the relative degree of saturation and \( k_r(\theta_e) \) is the relative unsaturated hydraulic conductivity, and \( k_s \) and \( k \) are respectively the saturated and unsaturated conductivity. The relative degree of saturation is calculated as the ratio between actual storage in layer \( S_{i,j} \) and its maximum storage capacity \( S_{C,i,j} \):

\[ \theta_e = \frac{S_{i,j}}{S_{C,i,j}} \]

The dependence of the relative unsaturated hydraulic conductivity on the relative degree of saturation is described by means of the relationship of Clapp and Hornberger (1978):

\[ k_r(\theta_e) = \theta_e^{\beta + 1} \]

where \( \beta \) is a dimensionless empirical component that varies on average between 4.05 and 11.4 over the range from sand to clay.

In the model percolation is driven by gravitational flow only and is equal to:

\[ Q_{per,i,j} = k(\theta_e)_{i,j} \cdot \Delta t \]
Capillary rise occurs when an upward gradient exists and the flux rate is balanced by the available storage in the underlying groundwater reservoir. If the relative degree of saturation of the top layer is smaller than that of the underlying second store, an upward capillary rise $Q_{CRj}$ will occur. It is driven by the soil moisture deficit in the top layer and proportional to the unsaturated hydraulic conductivity of the second layer:

$$Q_{CRj} = (1 - \theta_{E;1,j}) \cdot k(\theta_{E;2}) \cdot \Delta t, \text{ if } \theta_{E;1,j} < \theta_{E;2},$$  \hspace{1cm} (C.5)

or,

$$Q_{CRj} = 0, \text{ if } \theta_{E;1,j} \geq \theta_{E;2}. \hspace{1cm} (C.6)$$

For the second layer, the capillary rise is described in a similar way, except that the rate of the upward flux is given by $\sqrt{k(\theta_{E;1,j}) \cdot k_{s;1,j}}$, given the proximity of the water table, and that the resulting moisture content cannot rise above field capacity (with $\psi=1.0 \text{ m}$). The soil matric suction is given by: $\Psi = \Psi_s \cdot \theta_{E}^{-\beta}$, where $\psi_s$ is the air entry value [m].

Two land cover types are distinguished in the model: short and tall. They concern the canopy, land surface and the first two soil compartments. The land cover types influence the evaporation rate from the first two layers. This data is derived from the GLCC classification scheme.

**Model parameter maps**

Perturbations were created for the thickness of the first store of the model and for the saturated hydraulic conductivity. The thickness of the first store without perturbations is shown in figure C.1. The thickness of the second store is shown in figure C.2. As is explained earlier, the third store has an unlimited thickness. The saturated hydraulic conductivity of store 1 and 2 are shown in respectively figures C.3 and C.4. The maximum storage of the first store is dependent on the thickness of the first store. The original maximum storage map is shown in figure C.5. The mean maximum storage of the ensembles is shown in figure C.6 and the difference with the original maximum storage map is shown in figure C.7. In case more ensembles were created, it is expected that the overall difference would be less.

The local drainage direction map is derived from the digital elevation map of West Africa. From the LDD map the watersheds map can be computed. It was found that the Volta and the Senegal basin were largely overestimated. Therefore some drainage directions were changed manually so that some parts no longer belonged to the Senegal or Volta watersheds. The watershed map is shown in figure C.8. The area of the different watersheds is summed up in table C.1. Apparently, a too large area has been disconnected from the Senegal basin. The difference between the true area and the modelled area is relatively small for the other basins. The error in the Niger river basin could be explained by the relatively small part of the Niger river basin that is not covered by the model, because it is outside the study area (North of 20°N). From this region, little runoff can be expected and thus it is assumed that the effect on the runoff error is little.
Table C.1. Surface area (km$^2$) of the Senegal-, Volta-, and the Niger watershed. True area versus the modelled area and the relative difference.

<table>
<thead>
<tr>
<th>Watershed</th>
<th>True Area (km$^2$)</th>
<th>Modelled Area (km$^2$)</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senegal</td>
<td>419575</td>
<td>361400</td>
<td>-13.9</td>
</tr>
<tr>
<td>Volta</td>
<td>407093</td>
<td>425520</td>
<td>4.5</td>
</tr>
<tr>
<td>Niger</td>
<td>2261741</td>
<td>2203800</td>
<td>-2.6</td>
</tr>
</tbody>
</table>

Figure C.1. Thickness the first store. Values range from 0.02 m (white) to 0.35 m (black).

Figure C.2. Thickness the second store. Values range from 0.37 m (white) to 1.21 m (black).
Figure C.3. Saturated hydraulic conductivity of the first store. Values range from $0.05 \ln(\text{cm} \cdot \text{d}^{-1})$ (black) to $3.25 \ln(\text{cm} \cdot \text{d}^{-1})$ (white).

Figure C.4. Saturated hydraulic conductivity of the second store. Values range from $0.05 \ln(\text{cm} \cdot \text{d}^{-1})$ (black) to $3.85 \ln(\text{cm} \cdot \text{d}^{-1})$ (white).
Figure C.5. Maximum storage of the first store (m).

Figure C.6. Mean maximum storage of the first store of the ensembles (m).
Figure C.7. The difference between the mean and the original maximum storage (m).

Figure C.8. Watershed map of West Africa, showing the Senegal-, Volta- and Niger Basin.
Appendix D: Satellite details

This appendix describes the details of the satellites of which data is used for this research. The orbital details of the ERS-satellite is shown in table D.1. Table D.2 Shows the details of the TRMM-satellite and table D.3 of the Meteosat Second Generation satellite.

The ERS-satellite

The scatterometers on board of the ERS-satellite consists of three antennae producing beams looking to 45 degrees backward, 90 degrees sideways and 45 degrees forward with respect to moving direction of the satellite along its orbit, see figure D.1. Each beam measures 19 nodes, which are 25 km spaced apart. The swath of the satellite is thus 500 km. The nodes contain measured $\sigma^0$-values, which represent over 50 km diameter integrated backscatter (Woodhouse and Hoekman, 2000).

![Figure D.1. Operation of the ERS-satellite](image)

Table D.1. ERS-2 satellite orbital details

<table>
<thead>
<tr>
<th>Orbit</th>
<th>Sun-synchronous polar orbit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude</td>
<td>780km</td>
</tr>
<tr>
<td>Inclination</td>
<td>98.5°</td>
</tr>
<tr>
<td>Orbital period</td>
<td>100 minutes</td>
</tr>
<tr>
<td>Repeat cycle</td>
<td>3-35 days</td>
</tr>
</tbody>
</table>
Table D.2. TRMM satellite orbital details

<table>
<thead>
<tr>
<th>Orbit</th>
<th>Circular, non sun-synchronous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude</td>
<td>403 km</td>
</tr>
<tr>
<td>Inclination</td>
<td>35°</td>
</tr>
<tr>
<td>Orbital period</td>
<td>92.5 minutes</td>
</tr>
<tr>
<td>Repeat cycle</td>
<td>16 days</td>
</tr>
</tbody>
</table>

Table D.3. Meteosat Second Generation satellite orbital details

<table>
<thead>
<tr>
<th>Orbit</th>
<th>Geostationary, non sun-synchronous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude</td>
<td>35800 km</td>
</tr>
<tr>
<td>Inclination</td>
<td>35°</td>
</tr>
<tr>
<td>Orbital period</td>
<td>24 hours</td>
</tr>
<tr>
<td>Repeat cycle</td>
<td>not applicable</td>
</tr>
</tbody>
</table>
Appendix E: Actual evaporation

Evaporation is the change from liquid to gas. In hydrologic processes a distinction is made between evaporation from vegetated areas, non-vegetated areas and from plants only. They are identified as evapotranspiration, evaporation and transpiration respectively. Evapotranspiration is in fact a combination of evaporation and transpiration. The actual evaporation is a combination of all the evaporation processes; interception, transpiration, soil evaporation and open water evaporation (TU Delft, 2007).

In this thesis, actual evaporation data from EARS is used. The data is largely derived from Meteosat I and II observations. In this appendix the derivation of this product is briefly described. This appendix is based on De Weirdt et al. (2007)

Evaporation is directly related to the earth's surface energy balance. The balance is given by:

\[ R_n = H + \lambda E + E_{lu}, \]  

(E.1)

where \( R_n \) is the net radiation, \( G_0 \) the soil heat flux, \( H \) the turbulent sensible heat flux, \( \lambda E \) the turbulent latent heat flux and \( E_{lu} \) is the photosynthetic light use (all in W·m\(^{-2}\)). The difference between (turbulent) sensible heat and (turbulent) latent heat is that sensible heat, as the name declares, can be sensed (e.g. by satellite), but latent heat cannot. Latent heat is the energy in the form of heat released or absorbed by a substance during a change of phase. In the energy balance, latent heat thus describes the amount of evaporation. Unfortunately, the sensible heat cannot be measured on cloudy days. This term is estimated using the information of the last cloud-free day available.

The soil heat flux \( G_0 \) is often omitted in the energy balance equation, because on timescales equal to or larger than a day it is relatively small compared to the other parameters (van den Akker and Boomgaard, 2001). The energy balance formula then becomes:

\[ R_n = H + \lambda E + E_{lu}, \]  

(E.2)

where \( \lambda \) is the heat of evaporation, which is the energy needed to transform a given quantity of a substance into a gas. For water this is \( \lambda = 40.65 \text{ kJ·mol}^{-1} \).

The net radiation is calculated via

\[ R_n = (1 - \alpha)R_g + R_{lw}, \]  

(E.3)

where \( \alpha \) is the albedo coefficient, \( R_g \) the daily global solar radiation [W·m\(^{-2}\)], \( R_{lw} \) the net long wave radiation [W·m\(^{-2}\)]. The net long wave radiation is equal to the long wave radiation entering the earth's atmosphere minus the long wave radiation emitted by the earth itself (De Weirdt, personal communication 2007). The surface albedo is derived from Meteosat MSG. The incoming global solar radiation \( R_g \) at noon is obtained by:

\[ R_{noon} = S \cdot t \cdot cos(i), \]  

(E.4)

with \( S \) the solar constant (1355 W·m\(^{-2}\)), \( t \) the transmission coefficient and \( i \) the solar zenith angle. The daily global radiation \( R_g \) is obtained by integration of the daily solar cycle and is a function of day number and latitude. For cloudy pixels, the transmission coefficient through the cloud \( t \) is derived from the cloud albedo according to a relationship derived from the Kubelka-Munk theory.
The net long wave radiation is calculated from the surface and atmospheric temperatures and emissivities:

\[ R_{lw} = \epsilon_0 \epsilon_a \sigma T_a^4 - \epsilon_0 \sigma T_0^4, \]  

(E.5)

where the land surface emissivity \( \epsilon_a \) on average varies from 0.85 (desert) to 0.95 (vegetation). An average value of 0.9 is assumed here. The atmospheric emissivity \( \epsilon_0 \) is derived with the Brunt equation, based on relative air humidity values. The surface temperature is represented by \( T_0 \) [K], the air temperature at the top of the planetary boundary layer by \( T_a \) [K] and \( \sigma \) is the Stefan-Boltzmann constant \( \sigma = 5.67 \cdot 10^{-8} \text{ W} \cdot \text{m}^{-2} \cdot \text{K}^{-4} \). The planetary boundary layer is the air layer near the ground affected by diurnal heat, moisture or momentum transfer to or from the surface (De Weirdt, personal communication 2007). The surface temperature is derived with the thermal infrared satellite images. From the surface temperature the air temperature is derived. It is out of the scope of this thesis to explain how this is done. Below clouds it is assumed that the net long wave radiation is almost zero.

The sensible heat flux \( H \) is calculated via:

\[ H = (\alpha_c + \alpha_r) \cdot (T_0 - T_a) = C \cdot v_a \cdot (T_0 - T_a) + 4 \cdot \epsilon_0 \cdot \sigma \cdot T_3 \cdot (T_0 - T_a), \]  

(E.6)

with \( \alpha_c \) the convective sensible heat transfer coefficient and \( \alpha_r \) the radiative sensible heat transfer coefficient. \( C \) is the drag coefficient, \( v_a \) the average wind speed and \( T \) is the mean temperature \( (T_0 + T_a) / 2 \). The drag coefficient \( C \) is calculated by a function that quantifies the effect of elevation on aerodynamic roughness of the area and also incorporates the influence of decreasing air density with elevation. For more information on the derivation of the drag coefficient \( C \), refer to De Weirdt et al. (2007). The difference between the surface temperature and the air temperature at the boundary layer determines the magnitude of the sensible heat flux \( H \).

When vegetation is present, a part of the solar radiation (~5% of the absorbed global radiation) is used for photosynthetic electron transport \( (E_{ib}) \) (De Weirdt, personal communication 2007):

\[ E_{ib} = \epsilon (1 - \alpha_r) \cdot R_g \cdot C_v, \]  

(E.7)

with \( \epsilon \) is the photosynthetic light use efficiency on a daily basis (-) and \( C_v \) is the fraction of the surface covered by vegetation (-).

The photosynthetic light use efficiency is estimated on the basis of the Photosystem Deactivation Model (Rosema et al. 1998). The vegetation cover \( C_v \) is not known independently. It is clear however that presence of vegetation is characterized by high evapotranspiration values. Therefore the relative evapotranspiration \( \lambda E / \lambda E_p \), where \( E_p \) is the potential evaporation) is used as a proxy of crop coverage. This value is determined for every day.

\[ C_v = \lambda E / \lambda E_p \times \lambda E / (0.8 \times R_{ib}) \]  

(E.8)

The actual evaporation \( E_a \) [mm \cdot d^{-1}] is then calculated as:

\[ E_a = \frac{R_{ib} - H - E_{ib}}{\lambda} \]  

(E.9)
The standard deviation of the EARS actual evaporation data

In this section the standard deviation of the EARS actual evaporation data is derived from the information presented in the paper of De Weirdt et al. (2006).

The latent heat is derived from the energy balance, that is simplified to formula E.10.

To calculate the standard deviation, formula E.11 can be used. The sensible heat $H$, is derived by formula E.12, where $C$ is the drag coefficient and $v_a$ the average windspeed in m⋅s$^{-1}$. $T_0$ is the surface temperature measured by the satellite and $T_a$ is the temperature at the boundary layer. The net radiation is derived from formula E.13, where $S$ is the solar constant (1355 W⋅m$^{-2}$), $t$ is the atmospheric transmission coefficient, $i_s$ the solar zenith angle, $\varepsilon_0$ the land surface emissivity (0.9), $\varepsilon_a$ the atmospheric emissivity and $\sigma$ the Stefan-Boltzmann constant (5.67*10$^{-8}$ W⋅m$^{-2}$⋅K$^{-1}$).

In the paper of De Weirdt et al. (2006), the average temperature over a year at 1.5 m height has been validated for 9 different sites and for two years. They are compared to the measured temperatures. The temperature at 1.5m is derived by formula E.14. If a normal distribution of the differences is assumed, then the standard deviation of the temperature at 1.5 m height, derived by formula E.14 is equal to 1.823 K. The temperature at the boundary layer can be derived by some relation between day and night temperature profiles. Figure E.1 shows an example of a possible relationship. If it is assumed that the squared standard deviation of $T_{1.5}$ is equal to half the standard deviation of $T_0$, then the standard deviation for $T_a$ can be computed by formula E.15. The standard deviation for $T_0$ is then $\sigma_{T0} = 1.29$ K and for $T_a$, $\sigma_{Ta} = 2.07$ K.

Knowing the standard deviation for $T_0$ and $T_a$, the standard deviation for $H$ can be derived by formula E.16. In this case $C$ is set at 1 and $v_a$ is chosen at 10 m⋅s$^{-1}$. The standard deviation of $H$ is $\sigma_H = 9.29$ W⋅m$^{-2}$.

To derive the standard deviation for the net radiation some assumptions must be made. It is for example assumed that the temperature at 1.5 m is 25 degrees Celsius and at surface 30 degrees. This leads to a temperature at the boundary layer of 7.7 degrees Celsius. The $\varepsilon_a$ is estimated at 0.9. It is assumed that the standard deviation of $S*t*\cos(i_s)$ is zero. Therefore the standard deviation of $I_e$ can be computed by formula E.17. The standard deviation of the net radiation is $\sigma_{I_e} = 5.18$ W⋅m$^{-2}$.
The standard deviation of the latent heat becomes then, $\sigma_{LE} = 9.29 \text{ W} \cdot \text{m}^{-2}$.

The latent heat of vaporization for water is 2260 KJ $\cdot$ kg$^{-1}$. This is the same as 2260 J $\cdot$ m$^{-3}$. The standard deviation for evaporation $(E)$ is then $(9.29/2260 =) 0.0041 \text{ m} \cdot \text{day}^{-1}$, which is equal to 4.11 mm $\cdot$ day$^{-1}$.

This is a very large deviation. The calculations have also been performed by assuming the standard deviation for $T_a$ and $T_0$ to be equal to $T_{1.5}$, but this yielded an even higher deviation. Also the standard deviation for $T_a$ was set at half the deviation of $T_{1.5}$, but again the deviation for the evaporation was higher than 4.11 mm $\cdot$ day$^{-1}$.

In case the temperature at 1.5 m is guessed at 20 degrees and at surface at 25 degrees, then the standard deviation for evaporation is 4.08 mm $\cdot$ day$^{-1}$. If the drag coefficient is set to 0.9 and the mean windspeed to 8 m $\cdot$ s$^{-1}$, then the standard deviation becomes 3.66 mm $\cdot$ day$^{-1}$. In case the average windspeed is set to 4 m/s and the drag coefficient to 0.3, the standard deviation is still 2.53 mm $\cdot$ day$^{-1}$.

The standard deviation remains quite large compared to the yearly sum of evaporation for the catchment of the study area from which the data is taken (+/- 336 mm $\cdot$ day$^{-1}$). The average evaporation per day is thus 0.92 mm $\cdot$ day$^{-1}$.

**Formulae:**

\[ \text{In} = H + LE \quad (E.10) \]

\[ \sigma_{In}^2 = \sigma_H^2 + \sigma_{LE}^2 \quad (E.11) \]

\[ H = C \times v_a \times (T_0 - T_a) \quad (E.12) \]

\[ \text{In} = S \times \cos(i) + e_0^* \times e_0^* \sigma \times T_1^4 - e_0^* \times e_0^* \sigma T_0^4 \quad (E.13) \]

\[ T_{1.5} = 0.48 \times T_0 + 0.59 \times T_a - 12.9 \quad (E.14) \]

\[ \sigma_{T1.5}^2 = 0.48 \times \sigma_{T0}^2 + 0.59 \times \sigma_{T_a}^2 \quad (E.15) \]

\[ \sigma_H^2 = C \times v_a \times (\sigma_{T0}^2 - \sigma_{T_a}^2) \quad (E.16) \]

\[ \sigma_{In}^2 = \sigma_{T_a}^2 \times e_0^* \times e_0^* \sigma \times T_1^3 - \sigma_{T0}^2 \times e_0^* \times e_0^* \sigma T_0^3 \quad (E.17) \]

The mean windspeed over West Africa is about 2 m $\cdot$ s$^{-1}$. The drag coefficient could vary between 2.7 and 4.2 if the average elevation for West Africa is set to 100m. Higher elevations will increase the coefficient; lower mean elevation will decrease the coefficient. For $C=2.7$, the deviation is 3.3 mm $\cdot$ day$^{-1}$, for $C=4.2$ the deviation is 3.84. An average value of 3.45 yields a standard deviation of 3.6 mm $\cdot$ day$^{-1}$.
Appendix F: Model initialization

This appendix describes how the hydrological model is warmed up in order to create an initial condition for the final model run. In the EnKF, all perturbed models are warmed up using the perturbed variables.

The initial conditions were first set to zero for the soil moisture contents of all stores. The model is then run in blocks (runs) of four years, using the input data from 2003-2006. The soil moisture content of the first store reaches an equilibrium state after the first run. The equilibrium is reached much later in the second and third store of the model. To determine how long the model must be warmed up, it is chosen to check when the equilibrium is reached for the second store, because the third store might never be at equilibrium and store 1 is influenced by the soil moisture content of store 2.

Figure F.2 shows the soil moisture content of the second store for six different locations, which locations are shown in figure F.1. It is assumed that the chosen locations are representative for the entire study area. The soil moisture content is not at equilibrium for most locations after the first run. The final state after the this run is stored and used as initial condition for the second run and so forth. Figure F.3 to F.6 show the results for the next runs. In the second run, the initial soil moisture content for location 2 is for example 0.2 m and after four years it is 0.23 m and for location 5 it is respectively 0.06 m and 0.15 m. For the other locations, the difference between the initial state and the final state is much smaller. In the third run, location 2 reaches an equilibrium as well, but the initial and final state for location 5 still differs significantly. Location five reaches an equilibrium in the fifth run, see figure F.6. Figure F.7 shows the sixth run. The result is very similar to figure F.5. This confirms that the equilibrium is reached for all locations.

One can argue that the model is sufficiently warmed up after the first run already, because the equilibrium is only not reached at just one location which lies in a very dry region anyway. An equilibrium is generally reached earlier in wet areas than in dry areas. For this thesis it is chosen to choose the moment at which all locations are at an equilibrium, because it was not checked if the locations are in fact representative for the whole study area. If more ensemble models are created, it is advisable to reduce the warming up period because it significantly saves computation time.

Figure F.1. Six locations for which the soil moisture content time series of the second store are stored.
Figure F.2. Soil moisture content (m) of run 1 (year 0 – 4) for store 2 at the six different locations.

Figure F.3. Soil moisture content (m) of run 2 (year 5 – 8) for store 2 at the six different locations.

Figure F.4. Soil moisture content (m) of run 3 (year 9 – 12) for store 2 at the six different locations.
Figure F.5. Soil moisture content (m) of run 4 (year 13 – 16) for store 2 at the six different locations.

Figure F.6. Soil moisture content (m) of run 5 (year 17 – 20) for store 2 at the six different locations.

Figure F.7. Soil moisture content (m) of run 6 (year 21 – 24) for store 2 at the six different locations.
Assimilation of remote sensed soil moisture data into a hydrological model using the EnKF
Appendix G: Waterbalance and runoff results

This appendix shows the waterbalance- and runoff results of the model run without data assimilation. The cumulative difference between precipitation and actual evaporation is shown in figure G.1. Figure G.2 shows the average runoff over 2003-2006 in $\text{m}^3 \cdot \text{s}^{-1}$. The discharges at the outlets of the Senegal -, the Volta and Niger river are shown in figure G.3.

Figure G.1. The cumulative difference between $P$ and $E_a$ (m) for Senegal-, Volta- and Niger river basin.

Figure G.2. The average runoff from 2003-2006 ($\text{m}^3 \cdot \text{s}^{-1}$).
Assimilation of remote sensed soil moisture data into a hydrological model using the EnKF
Appendix H: Model- and observation results

Figure H.1. Mean soil moisture content at the Senegal basin compared to the mean soil moisture content according to the observations, after translated to model variable. The precipitation and actual evaporation are shown in the second plot.

Figure H.2. Mean soil moisture content at the Niger basin compared to the mean soil moisture content according to the observations, after translated to model variable. The precipitation and actual evaporation are shown in the second plot.
Assimilation of remote sensed soil moisture data into a hydrological model using the EnKF
Appendix I: Correlations

In this appendix the correlation maps are shown for each watershed. Table 2.1 shows which correlations are computed.

**Correlations for Senegal basin**

![Figure I.1](image1.png)

*Figure I.1. Significant correlations found for the Senegal basin from 24/08/2003 to 30/10/2003.*

![Figure I.2](image2.png)

*Figure I.2. Significant correlations found for the Senegal basin from 31/10/2003 to 30/04/2004.*
Figure I.3. Significant correlations found for the Senegal basin from 01/05/2004 to 30/10/2004.

Figure I.4. Significant correlations found for the Senegal basin from 31/10/2004 to 30/04/2005.

Figure I.5. Significant correlations found for the Senegal basin from 01/05/2005 to 30/10/2005.
Figure I.6. Significant correlations found for the Senegal basin from 31/10/2005 to 30/04/2006.

Figure I.7. Significant correlations found for the Senegal basin from 01/05/2006 to 30/10/2006.
Correlations for Volta basin

Figure I.8. Significant correlations found for the Volta basin from 24/08/2003 to 30/10/2003.

Figure I.9. Significant correlations found for the Volta basin from 31/10/2003 to 30/04/2004.
Figure I.10. Significant correlations found for the Volta basin from 01/05/2004 to 30/10/2004.

Figure I.11. Significant correlations found for the Volta basin from 31/10/2004 to 30/04/2005.
Figure I.12. Significant correlations found for the Volta basin from 01/05/2005 to 30/10/2005.

Figure I.13. Significant correlations found for the Volta basin from 31/10/2005 to 30/04/2006.
Figure I.14. Significant correlations found for the Volta basin from 01/05/2006 to 30/10/2006.

Correlations for Niger basin

Figure I.15. Significant correlations found for the Niger basin from 24/08/2003 to 30/10/2003.
Figure I.16. Significant correlations found for the Niger basin from 31/10/2003 to 30/04/2004.

Figure I.17. Significant correlations found for the Niger basin from 01/05/2004 to 30/10/2004.

Figure I.18. Significant correlations found for the Niger basin from 31/10/2004 to 30/04/2005.
Figure I.19. Significant correlations found for the Niger basin from 01/05/2005 to 30/10/2005.

Figure I.20. Significant correlations found for the Niger basin from 31/10/2005 to 30/04/2006.

Figure I.21. Significant correlations found for the Niger basin from 01/05/2006 to 30/10/2006.
Assimilation of remote sensed soil moisture data into a hydrological model using the EnKF
Appendix J: Sensitivity analysis

Figure J.1. The influence of the change of a variable on the model state of store 1 and 2. The relative difference between the model state based on a changed variable value and the original model state is shown.
Figure J.2. The original surface wetness, the surface wetness in case the thickness of the first layer is increased by 0.10 m and the surface wetness in case the thickness is decreased by 0.10m.
Appendix K: Data assimilation results

Figure K.1. The variance of the model state at 15°N, 12.5°W. The observations at this location shown with dots, the continuous line represents the variance before the updates are made and the dotted line the variance after the update is made.

Figure K.2. Mean soil moisture content of the updated model at the Senegal basin between 13/08/2003 - 31/12/2003 compared to the mean original soil moisture content and the mean soil moisture content according to the observations, after translated to a soil moisture content of the first store. The precipitation and actual evaporation are shown in the second plot.
Figure K.3. Mean soil moisture content of the updated model at the Niger basin between 13/08/2003 - 31/12/2003 compared to the mean original soil moisture content and the mean soil moisture content according to the observations, after translated to a soil moisture content of the first store. The precipitation and actual evaporation are shown in the second plot.