VALIDATION OF SMOS SATELLITE DATA OVER GHANA AND BURKINA FASO
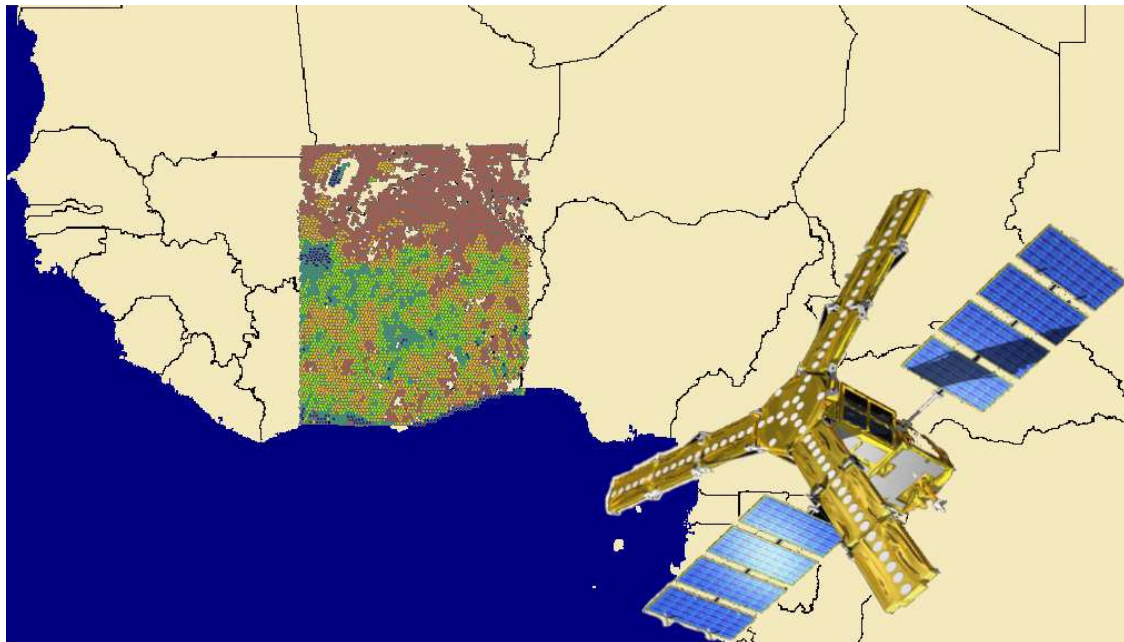
TESTING THE SMOS ALGORITHM OVER GUINEA AND SUDAN SAVANNA VEGETATION

MASTER THESIS
WATER RESOURCES MANAGEMENT
DELFT UNIVERSITY OF TECHNOLOGY
ANKE POELSTRA, BSC
1213202

Universiteitsfonds
Delft

VALIDATION OF SMOS SATELLITE DATA OVER GHANA AND BURKINA FASO

TESTING THE SMOS ALGORITHM OVER GUINEA AND SUDAN SAVANNA VEGETATION

By

Anke Poelstra, BSc
ankepoelstra@gmail.com

Delft University of Technology
Section Water Resources Management

Graduation committee:

Prof. Dr. Ir. N.C. van de Giesen, TU Delft, Water Resource Management
Ir. M.M. Rutten, TU Delft, Water Resource Management
Dr. K. Khoshelham, TU Delft, Optical & Acoustic Remote Sensing
Dr. Dr. S.C. Steele-Dunne, TU Delft, Water Resource Management

29 October 2010

Figures frontpage:

Figure 0.1: SMOS satellite (http://www.euinfrastructure.com/news/smos-water-satellite, 01-09-2010)

Figure 0.2: Universiteitsfonds Delft logo

FOREWORD

This report is the end product of my master thesis research on Soil Moisture and Ocean Salinity (SMOS) satellite soil moisture measurements in West-Africa. Research started in February 2010 and ended in October 2010. Part of this research was a field work in Ghana, executed in April and May 2010.

This research compares soil moisture values of the SMOS satellite to values derived from measurements in the field and other satellites. The purpose of this comparison is to validate the algorithm used to determine soil moisture contents from SMOS satellite measurements. It is written for anybody who is interested in the reliability of the SMOS satellite for West-Africa.

I would like to thank everybody who has helped me with this research. First of all, I would like to thank my daily supervisor, ir. Martine Rutten, for her weekly support and criticism. The rest of my committee, prof. dr. ir. Nick van de Giesen, dr. Kourosh Khoshelham and dr. Susan Steele-Dunne and my team of field work experts, ir. Frank Annor, dr. Jan Friesen, dr. Jens Liebe and William Horner. Also thanks to Natalie Ceperley from Polytechnique Fédérale de Lausanne and Richard de Jeu from Vrije Universiteit Amsterdam for their help in obtaining data. A special thanks goes to dr. Ali Mahmoodi of the Array company, for helping understand Linux and getting the L2PP processor to work. Finally, I would like to thank Universiteitsfonds Delft, for helping enabling my field work.

SUMMARY

Soil moisture (*SM*) is an important parameter in weather, hydrologic, climatologic and atmospheric models because of its influence on evaporation, infiltration, runoff and uptake of water by vegetation. Because of its variability in time and place, however, creating a database of the entire earth based on field measurements would be extremely time- and labor-intensive. A solution to this is satellite imagery. Satellite imagery can give information on *SM* on a global scale, at intervals of at most a couple of days.

For this purpose, the European Space Agency (ESA) has developed the Soil Moisture and Ocean Salinity (SMOS) satellite. This satellite measures brightness temperature (*TB*) at L-band. From these measurements, *SM* in the top five centimeters of the soil can be determined. After resampling, the data resolution is approximately 15 by 15 km. One such cell of 15 by 15 km is called a node. The satellite covers a specific node on average every three days. ESA's objective is to measure *SM* within 4% of its real value.

The SMOS satellite has been in operation since November 2009. Decades of research preceded this launch. Both a suitable measuring device and an algorithm had to be developed. *TB* that is measured by a satellite is not only dependent on *SM*, but also on vegetation, topography, snow conditions et cetera. Different parameters model these conditions in the algorithm.

Many parameters have been tested in research on plot sized areas, in the order of square meters instead of kilometers. Now that the satellite is launched, the algorithm has to be validated and calibrated for different areas in the world. This report focuses on the validation of the algorithm for West-Africa. The first step in this process is a sensitivity analysis of the parameters in the algorithm. This is used to determine which parameters are most influential. This knowledge can be used in further validation and calibration efforts and the design of future field experiments. Validating data was done by comparing SMOS *SM* values to *SM* values from field research and other satellite data.

The sensitivity analysis showed that the parameter that models the scattering of radiation through vegetation (scattering albedo ω), roughness of the area (modeled with the dimensionless parameter *HR*), temperature *T* and litter properties (modeled with its effect on the optical depth of vegetation τ_L) are the most influential parameters. Default values as found in the L2PP processor used for this analysis did not always seem to represent the research area best. E.g. litter was put on a default value of 0, while it is known that at certain periods in time, litter is found in the area. These differences cause the largest changes in *SM*. Looking solely at the influence of parameters when changing them within a range that is thought to represent the research area best, *HR* seems to have the largest impact on *SM*. All results should be met with caution. The functional form of the dependences showed that this is dependent on the measured *TB*, the original value for the parameter and the amount of change the parameter undergoes. Therefore, results cannot simply be generalized for other areas, even when the range of the parameter is the same.

Validation of SMOS satellite data over Ghana and Burkina Faso

The second step was to compare SMOS *SM* data with other *SM* datasets. In order to obtain field data a small field work was executed in northern Ghana, near Tamale. This was done in April and May, at the end of the dry season,. Its objective was to determine the average *SM* content of a SMOS node which could be compared to the satellite measurement of the area in that period. In order to compare field measurements with satellites measurements, an average that represented the *SM* in the complete pixel had to be found. This average is called a well defined average. In order to determine this average value, the area was divided into so-called hydrotopes on the basis of hydrologic features. Hydrotopes have two important characteristics. The *SM* values measured in a hydrotope show a distinct distribution and, in the wet season, the difference between the means of two hydrotopes is larger than the variances of the hydrotopes. The average values of the different hydrotopes should thus be very distinct and the overlap of their distributions is limited. It turned out however that a better distinct distribution for the top five centimeters of soil was obtained by dividing the area based on vegetation type. Comparison of the well defined average of the area with SMOS satellite data showed a relatively low value for the SMOS satellite, that nevertheless falls within 4% of the value measured in the field.

Other comparisons were done with measurements in Burkina Faso, *SM* data from other satellites and precipitation data. The field measurements in Burkina Faso are located near a stream and cover a line of a couple of kilometers, rather than a complete pixel. Moreover, the field measurements are done at a different depth than satellites' measurements. By comparing values for different locations, similarities of and differences between the patterns of the different satellites over time become apparent. The comparison with precipitation data shows whether the *SM* data reacts to rainfall. SMOS showed a very poor correlation with the field measurements in Burkina Faso. It also showed a small range compared to two of the three other satellite datasets and its absolute values were far lower than the field measurements. SMOS did show good correlation with other satellite data for areas with little precipitation. This latter data has in earlier research been found to have an error of 5.4%. SMOS also showed a good correlation with rainfall; its value goes up with precipitation and down in dry periods. The correlation with other satellite datasets for wet areas is however poor. The best correlations are found in the northern part of the research area, with less vegetation and rainfall. Some pixels more to the south, receiving very little rainfall, also showed a good correlation. Remarkable is the difference between *SM* values from SMOS morning and evening overpasses. Comparisons with satellite data did not show which overpass gives the more realistic value.

Concluding, it seems that the SMOS *SM* values are relatively low and although SMOS data responds to precipitation as expected, the amount of change is too small to ensure good correlation for wet areas. The correlation for dry areas is usually acceptable. Ensuring the right values for certain important parameters will probably improve the correlation values. It is expected that then the 4% boundary will be met for dry areas. Whether this will also be the case for wet areas will have to be investigated further.

CONTENTS

Validation of SMOS satellite data over Ghana and Burkina Faso

LIST OF FIGURES

Validation of SMOS satellite data over Ghana and Burkina Faso

LIST OF TABLES

GLOSSARY

APF: apodization function, function applied to visibilities in order to attenuate the effects of the sharp cut-off at the boundaries of the baseline domain.

Complementary class: classes are called complementary when together they cover the whole surface of a pixel.

DFFG: defines an almost equal-area grid system close to the reduced Gaussian ECMWF standard. The description of the grid property is very similar to the way reduced Gaussian grid are described in the WMO GRIB specification, but it has an uniform sampling of latitude. The DFFG samples meridians and parallels with an almost equal distance increment equal to DFFG_STEP_KM. This DFFG is called "flexible" for the two following reasons:

- DFFG_STEP_KM can be set to any resolution
- The DFFG acts as a fixed interface between the processor and the huge diversity of auxiliary data

Default contributions: contribution to the radiometric signal computed with physical parameters obtained from auxiliary data only.

DGG: an equi-surface grid, defined once and for all, on the nodes of which the soil moisture will be retrieved. The average inter-node distance is close to 15 km. For land surfaces only (including large ice covered areas), the grid should include about 6.5×10^5 nodes.

ECOCLIMAP: a global dataset at a 1 km resolution, intended to be used to initialize the Soil-Vegetation-Atmosphere-Transfer schemes (SVATs) in meteorological and climate models (at all horizontal scales). 215 Ecosystems representing areas of homogeneous vegetation are derived by combining existing land-cover maps and climate maps, in addition to using the AVHRR satellite data. Then, surface parameters used in SVATs, (as *LAI*, vegetation fraction, roughness length, minimal stomatal resistance, albedo and emissivity) are computed using look-up tables. The data set is aimed to be used in any type of atmospheric model.

Hydrotope: a landscape unit that shows internally consistent hydrologic behaviour.

ISEA-4-9: gridding system that paves the Earth surface with quasi equal-area cells and minimal distortion all around the globe.

Litter: dead vegetation laying on the ground surface.

Node (DGG node, L1C node, sometimes called a L1C pixel): the list of observed views attached to a given location on the Earth defined by its DGG ident number in the ISEA grid.

Reference value: predetermined values of the surface characteristics.

Validation of SMOS satellite data over Ghana and Burkina Faso

SMOS pixel: set consisting of the brightness temperature and its spatial resolution on the Earth, which can be thought of as the subset of a view made up with the *TB* itself and the 3dB surface characterised by the ellipse semi axes.

Supplementary: classes are called supplementary when in order to introduce them, the previous complementary classes have to be overwritten so that the resulting set again becomes complementary.

View: a collection of information attached with one single SMOS *TB* measurement. It consists of the following information:

- A brightness temperature, which can be a real number in dual polarization mode or a complex number in full polarization mode.
- An observation geometry that consists of the incidence angle and the azimuth angle relative to the observed surface location with respect to the instrument.
- A transformation geometry that consists of two angles; one related to the Faraday rotation, and one related to the transfer from the surface reference frame (H/V polarization) to antenna reference frame (X/Y polarization)
- A radiometric property that provides the radiometric accuracy of the brightness temperature, the semi-major and semi-minor axis of the 3dB ellipse contour of the synthetic antenna footprint related to the observed point.
- An identification which indicates the polarization of the brightness temperature, (either X or Y in dual polarization mode, or XX, YY or XY in full polarization mode) as well as the identification number of the snapshot gave the brightness temperature.

WA$_{DFFG}$: matrix of DFFG cells making the Working Area

NOMENCLATURE

e_{sp} (-)

Surface emissivity coefficient, p stands for polarization

$\bar{M}$

Overall mean value, accuracy depends on the number of samples taken

$\bar{m}$

Mean value of the sample, used in the determination of the minimal number of samples needed to obtain an acceptable range for the average soil moisture value at a given confidence level

n

Number of samples that is necessary to obtain an acceptable range for the average soil moisture value at a given confidence level

p_i

Physical parameters, used in the iterative determination of the soil moisture of a SMOS pixel

p_{i0}

Prior estimates of the free physical parameters, used in the iterative determination of the soil moisture of a SMOS pixel

s

Standard deviation of a sample of soil moisture measurements, used in the determination of the minimal number of samples needed to obtain an acceptable range for the average soil moisture value at a given confidence level

r_{sp} [-]

Integral of the surface scattering coefficient over all scattering directions

T_g, T_c [K]

Weighted sum of soil temperatures at subsurface levels accounting for the penetration depth;canopy physical temperature

TB_{atd}, TB_{atu}, TB_{atu} [K]

Down- and upwelling atmospheric emission (downward emission is reflected (scattered) at the surface and attenuated along the upward path by the atmosphere

TB_p [K]

Brightness temperature, p stands for polarization horizontal (H) and vertical (V)

TB_{sk} [K]

Cosmic background emission attenuated by the atmosphere, reflected /scattered (r_p) at the surface and attenuated again along the upward path by the atmosphere

TB_{sp} [K]

Earth's surface emission, p stands for polarization

TBF [K]

Brightness temperature at antenna level determined with the direct (forward) model, used in the iterative determination of the soil moisture of a SMOS pixel

Validation of SMOS satellite data over Ghana and Burkina Faso

TBM_m (K)	Valid measured values of brightness temperature, used in the iterative determination of the soil moisture of a SMOS pixel
θ (rad)	Incidence angle
μ (m^3/m^3)	True mean value, used in the determination of the minimal number of samples needed to obtain an acceptable range for the average soil moisture value at a given confidence level
σ (m^3/m^3)	True standard deviation, used in the determination of the minimal number of samples needed to obtain an acceptable range for the average soil moisture value at a given confidence level
σ_{i0}^2	Prior variances of the free physical parameters, used in the iterative determination of the soil moisture of a SMOS pixel
τ_{atu}, τ_{atd}, τ_{atm} (-)	(Downward and upward path) atmospheric opacity
τ_c (-)	Opacity of canopy
$\bar{X}$	Standard normally distributed variable, gives a confidence level, used in the determination of the minimal number of samples needed to obtain an acceptable range for the average soil moisture value at a given confidence level
ω_p (-)	Isotropic single scattering albedo, p stands for polarization

1 INTRODUCTION

1.1 MEASURING SOIL MOISTURE

Soil moisture (SM) is an important parameter in weather, hydrologic, climatologic and atmospheric models. It influences evaporation, infiltration, runoff, uptake of water by vegetation and thus indirectly transpiration, which in turn affects precipitation [1.1]. Figure 1.1 shows the different processes in the water cycle affected by SM. This basic water cycle works as follows: Water can reach the ground as precipitation or dew that condensates. Depending on the amount of water that is already in the ground, this water is able to penetrate into the ground. If water does not penetrate, it flows over the soil as run off. Water can of course also fall or condensate on other surfaces, like vegetation. In the ground, part of the water is bound to the soil, until too much water reaches the top layer and gravity takes over. At that point, water seeps through to the ground water. Water can get from the groundwater layer back up, against gravity, by capillary rising. The amount of water that plants can take up also depends on how much SM is in the ground. Water finally gets back in the atmosphere by evaporating from surfaces or transpiration through vegetation. Besides this direct influence on the water cycle, SM also has an effect on the surface energy balance; the division of energy (coming from vegetation, atmosphere and sun) over soil and atmosphere (not in the drawing). One of the ways energy goes back to the atmosphere is as so called sensible heat. This heat is transferred between soil and atmosphere when there is a difference in temperature Heat is transferred by conduction and convection. This process has an effect on the stability of the atmosphere, which is why SM is also important for climatologic and atmospheric models [1.2].
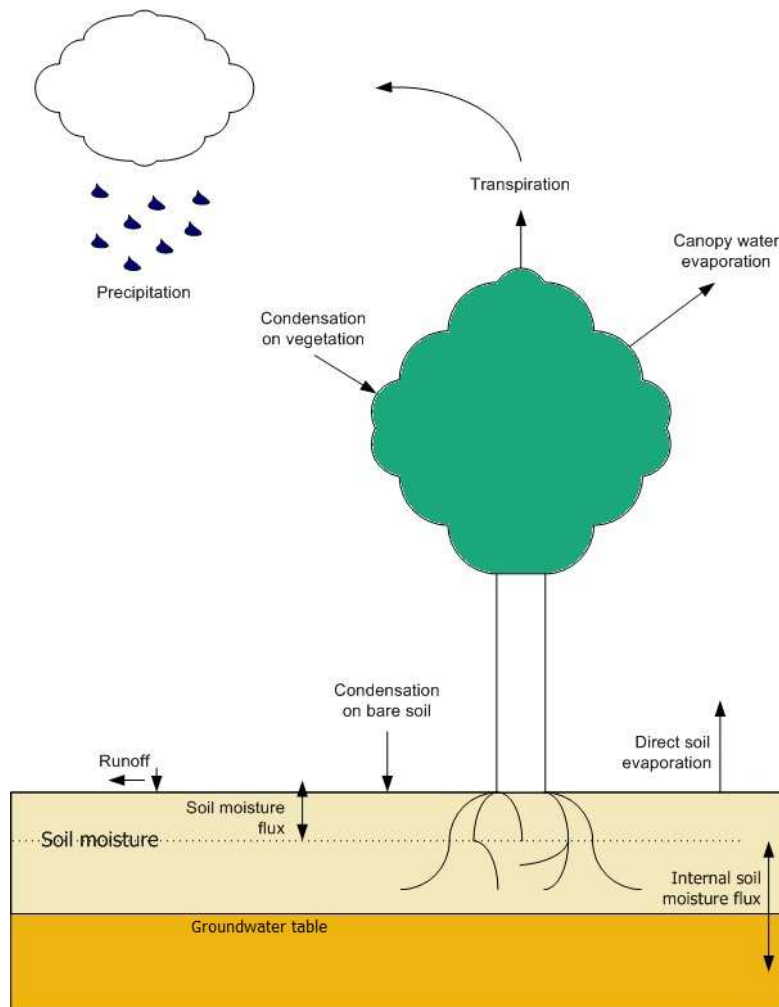
Figure 1.1: Processes of the water cycle influenced by soil moisture

Despite its importance, a limited amount of *SM* data is available. Because *SM* changes over both time and place, field campaigns can never give a complete picture. Nor is it feasible to place a network of meters over the entire earth.

A way to get a more complete overview of *SM* is remote sensing. With remote sensing, it is possible to get data on a global scale. In order to signal important changes in *SM* content, an area should be measured at least every three days [1.1]. The satellite therefore has an orbit that leads to a data resolution in the order of ten kilometers.

At this moment, there are a couple of satellites in orbit that give information on *SM*. These satellites do not measure *SM* directly, but they measure emissions coming from the earth. These emissions are converted into so-called brightness temperature (*TB*), which has a relation with *SM*. The sensitivity of *TB* to *SM* increases for a lower frequency. Most of the satellites that are used for the determination of *SM* content do not measure at the lowest frequency possible. That is because these satellites are not specifically designed for determining *SM*.

To overcome this, the European Space Agency (ESA) developed the Soil Moisture and Ocean Salinity (SMOS) satellite. This satellites measures emissions at 1.4 GHz, within the so-called L-band (ranging from one to two GHz). This frequency is the lowest frequency that is still protected [1.3]. Officially, the frequency is not used by radio communication, leaving it free for SM measurements. The SMOS satellite measures SM to a depth of five centimeter. Its objective is to determine SM content within a range of 4% [1.4].

The SMOS satellite was launched in November 2009, after decades of research. This research focused on the development of the actual measuring device and on that of a correct algorithm, used to determine SM from emissions.

For the algorithm, the so-called τ-ω model is used. This model sees vegetation as a single layer, modeled on the basis of its optical depth (τ) and its single scattering albedo (ω). These model respectively the vegetation attenuation properties and the scattering effects within the canopy layer [1.4]. Figure 1.2 illustrates the model. Soil temperature T_g, canopy temperature T_c, single scattering albedo ω, optical depth τ and emissivity e_g are parameters that have a direct influence on TB. b Is the parameter that links the Leaf Area Index (LAI) to optical depth τ. The complete algorithm is explained in the so-called SMOS level 2 Processor for Soil Moisture Algorithm Theoretical Based Document (ATBD), by the Expert Support Laboratories (ESL) of the ESA[1].
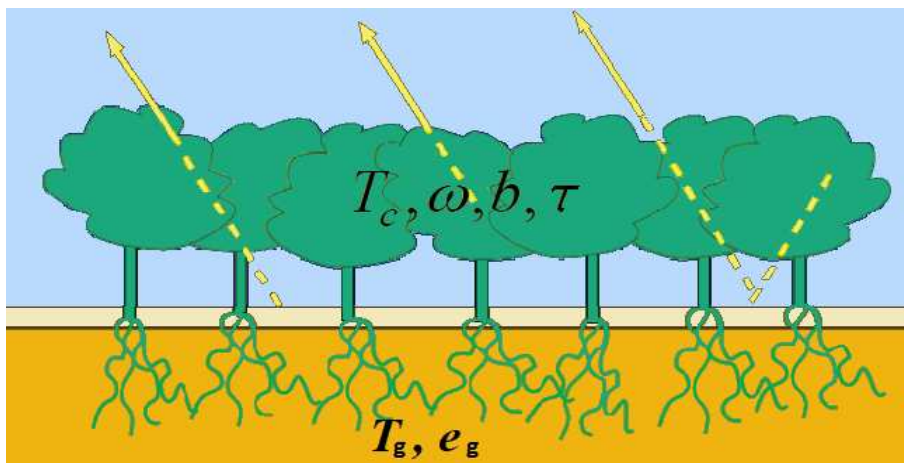
Figure 1.2: Illustration of the τ-ω model (figure created by Susan Steele Dunne). Tc, Tg, ω, τ and eg have a direct influence on SM. b has an indirect influence through LAI

---

[1] These are: Centre d'Etudes Spatiales de la BIOsphère, France (CESBIO), Institut Pierre-Simon Laplace, France (IPSL) - Service d'Aéronomie, Institute National de la Recherche Agronomique, (L'unité de recherche en) Ecologie fonctionnelle et PHYSique de l'Environnement, France (INRA-EPHYSE), Reading University (United Kingdom) and Tor Vergata University (Italy)

Many of the algorithm's parameters have been determined through experiments on a crop field scale. These fields were basically homogeneous, with the same vegetation, of the same age, over the entire field. [1.5-1.9]. The SMOS satellite however, after re-sampling, provides *SM* data for pixels with a size of approximately 15 by 15 km. These pixels almost never have a homogeneous vegetation layer. The values of the parameters determined through experiments therefore have to be adapted to this resolution. For example, the scattering coefficient ω for low vegetation has a default value of 0 [1.4], as it is expected that over a complete SMOS pixel, higher values that were found in experiments cancel each other out.

Now that the satellite is in orbit, the algorithm has to be validated and calibrated. This is done for different areas in the world. One of these areas is West-Africa. This report is about the validation process of the algorithm for Ghana and Burkina Faso.

1.2 VALIDATION FOR WEST-AFRICA

Agriculture in West-Africa depends mainly on rainfall. The closer to the Sahara desert however, the stronger the monomodal regime of precipitation. This means there is a clear dry and wet season, that causes droughts [1.10]. Due to the fact that agriculture in the region is mainly based on rainfall, droughts can have a major effect on communities.

Hydrologic knowledge can help in mitigating the effects of droughts. Knowledge of the area is however limited. A well calibrated data product from the SMOS satellite would be an important step in enhancing this knowledge.

Ghana and Burkina Faso are largely covered with savanna vegetation. Over Ghana, this is mostly Guinea savanna, over Burkina Faso there is Sudan savanna. Natural vegetation is interspersed with (small) agricultural fields [1.11]. Because of the distinct wet and dry seasons and fires in the dry season (either natural or ignited by humans), the vegetation cover changes drastically over the year.

The parameters in the SMOS algorithm, representing vegetation, change with the kind of vegetation. Experiments executed in order to get values for these parameters, however, are mostly done in Europe and the United Stated of America [1.5-1.9].

It remains to be seen whether the default values that are determined in these experiment hold true for West-Africa. Besides that, the difference in vegetation in the wet and dry season can be reason to adapt the parameters over time. Because at the time of this research, only limited SMOS data was available, it was not possible to look into the latter issue. The first, however, leads to the research questions that are answered in this report.

1.3 RESEARCH QUESTIONS

The main research question of this research is:

How well does the algorithm, as presented in the ATBD, perform for northern Ghana and Burkina Faso, West-Africa?

Several sub questions can be formulated in order to get an answer to our main question:

1. How sensitive is the algorithm to changing values of parameters?
2. Which parameters are the most influential for the region of northern Ghana?
3. How do SMOS satellite measurements compare to field measurements for a certain pixel?
4. How do SMOS satellite measurements compare to field measurements of the École Polytechnique Fédérale de Lausanne in Burkina Faso?
5. How do SMOS satellite measurements compare to other satellite data?

These sub-questions will be answered in the consecutive chapters of this report. Together, they form the answer to the main research question.

Part of determining the performance of the algorithm is determining which parameters have the largest impact on the outcome. The performance of the algorithm depends on the values for the different parameters, but also on the sensitivity of the algorithm to them. Every parameter has a range of values that represents a specific region in a certain period best. If changing this parameter within its range leads to a small change in SM, a rough estimate is enough to determine SM within the 4% range set by ESA. If it leads to large changes in SM, the parameter needs to be close to the real value to ensure a good performance of the algorithm.

With the sensitivity of the algorithm known, the outcome of the SMOS algorithm (that is SM data) should be checked. This is done by comparing the SM data to other datasets, such as field measurements, data from other satellites and precipitation data.

The only direct way to determine SM is through field measurements. Comparing SMOS SM data with field measurements shows the SMOS algorithm's error is in a certain time and place. SMOS SM data has a resolution of 15 km by 15 km. For this research, a field work was executed on the pixel with its centre located at 1°07'54" W, lat 9°11'44" N. Sub question 3 thus becomes "How do SMOS satellite measurements compare to field measurements for the pixel with its centre at 1°07'54" W, lat 9°11'44" N". In order to get a representative average value for such an area, a number of measurements is needed[2].

---

[2] The time-stability concept allows to determine the average value of an area by measurements on a specific location, that shows little difference with the actual average value over time. To

Since it would take too much time to measure every pixel in the field (over time), it is not possible to compare all SMOS data with field measurements. Another way to determine the performance of the algorithm is to compare the data with other data obtained through remote sensing. There are several satellites in orbit whose measurements can be used to determine the *SM* content. Finally, the *SM* data can be compared to precipitation data. *SM* should go up after rainfall and down in dry periods. These comparisons show whether the SMOS *SM* data shows the same patterns over time as the other *SM* and precipitation data.

1.4 THESIS OUTLINE

The chapters of this report are, except for chapter 2, 3 and 7, all linked to one or more research questions.

Chapter 2 and 3 give a more thorough introduction to respectively the research area in Ghana and Burkina Faso and the SMOS algorithm that is used for the determination of *SM* content.

Chapter 4 describes the sensitivity analysis that was done on the algorithm. It gives an answer to sub-questions 1 and 2. The effect of a change in value of a certain parameter is an important tool for calibrating the algorithm.

Chapter 5 and 6 cover the validation of SMOS data. In chapter 5, field measurements from a field work in Ghana at the end of the dry season are compared to SMOS data. The chapter gives an answer to sub-question 3. In chapter 6, SMOS data is compared with other satellite data and continuous field data near the town of Madjoari in Burkina Faso. These measurements are executed by the École Polytechnique Fédéralde de Lausanne (EPFL, Switserland). This gives answers to sub-questions 4 and 5.

In chapter 7 finally, the conclusions of the different chapters are summarized and an answer to the main research question is given.

---

determine which location represents the average best however, first different locations in the area need to be measured over time. [1.9]

2 WEST-AFRICA, GHANA AND BURKINA FASO

2.1 INTRODUCTION

As mentioned in paragraph 1.2, hydrologic knowledge on West-Africa is limited. What is known, is that the area has a monomodal precipitation regime and that over the last couple of decades, rainfall has often been very low [2.1]. The resulting droughts can become a big problem in the region. Due to the fact that agriculture in the region is mainly based on rainfall, droughts can have a major effect on communities. *SM* data from the SMOS satellite, when well calibrated, can be part of a better understanding of the drought problem.

The monomodal precipitation regime in the region gets stronger northward, in the direction of the Sahel. This research focuses on the area from mid-Ghana upwards. Two important locations are the sites where field measurements were/are done. The field work described in chapter 5 was executed in north Ghana, near Kusawgu junction, close to Tamale. EPFL measurements are done in south-east Burkina Faso, near the town of Madjoari. In the first area, the field work that is described in chapter 5 was executed. The two locations are indicated in Figure 2.1. The area is further described in paragraph 2.2.
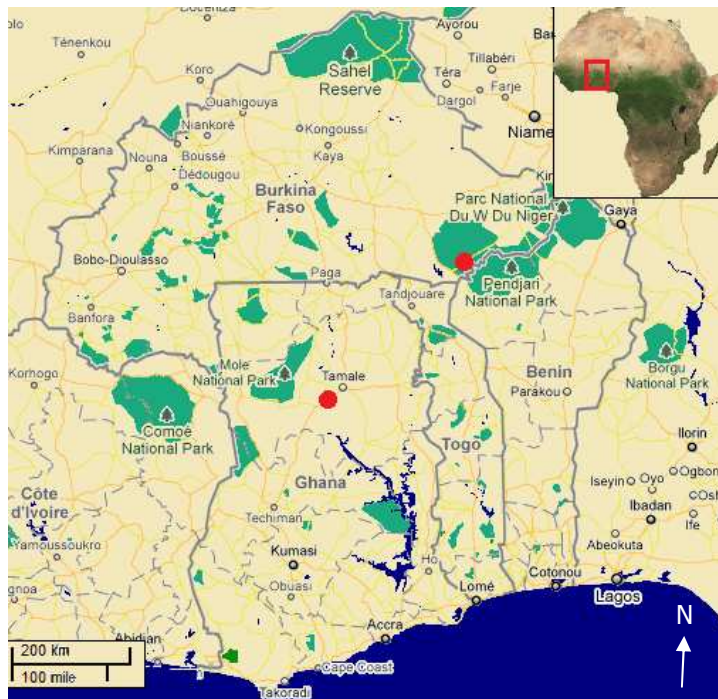
Figure 2.1: Ghana and Burkina Faso, location field work (red dot Ghana) and EPFL measurements (red dot Burkina Faso) (maps.google.com, 31-08-2010)

2.2 CHARACTERISTICS OF THE AREA

2.2.1 PRECIPITATION

Figure 2.2 and Figure 2.3 are illustrations of the difference in rainfall in the area from north to south. Figure 2.2 shows the mean annual rainfall in Burkina Faso. Figure 2.3 shows the distribution of rainfall in Ghana in 1990.

North of lat 9° N, there is one distinct rainfall season with its peak in August or September. Below this latitude, there are two rainy seasons very close to each other. The further north, the shorter the rainy season. Average annual rainfall varies from 1600 mm to the south of the Volta Dam to 380 mm in northern Burkina Faso [2.1].

The location of the field work executed for this research lies on a latitude of 9^o11'44" N. The closest city is Tamale (9^o30' N, 0^o51' W). Rainfall can vary from 500 to 3000 mm per year. Figure 2.4, the top figure, shows rainfall and potential evaporation for a climate station in Tamale.

EPFL measurements are done at a latitude between 11^o26' and 11^o28' N. This is at the edge of the Sudan zone, that goes up to 11^o30' N. Annual rainfall in this area is between 900 and 1200 mm, over on average 74 days [2.1]. Figure 2.4, the bottom figure, shows rainfall and potential evaporation for the city of Fada N'Gourmadiapaga (12^o18' N, 0^o21'36" E). Although this lies on a higher altitude than the EPFL measurements, it illustrates the difference in rainfall for the different locations.
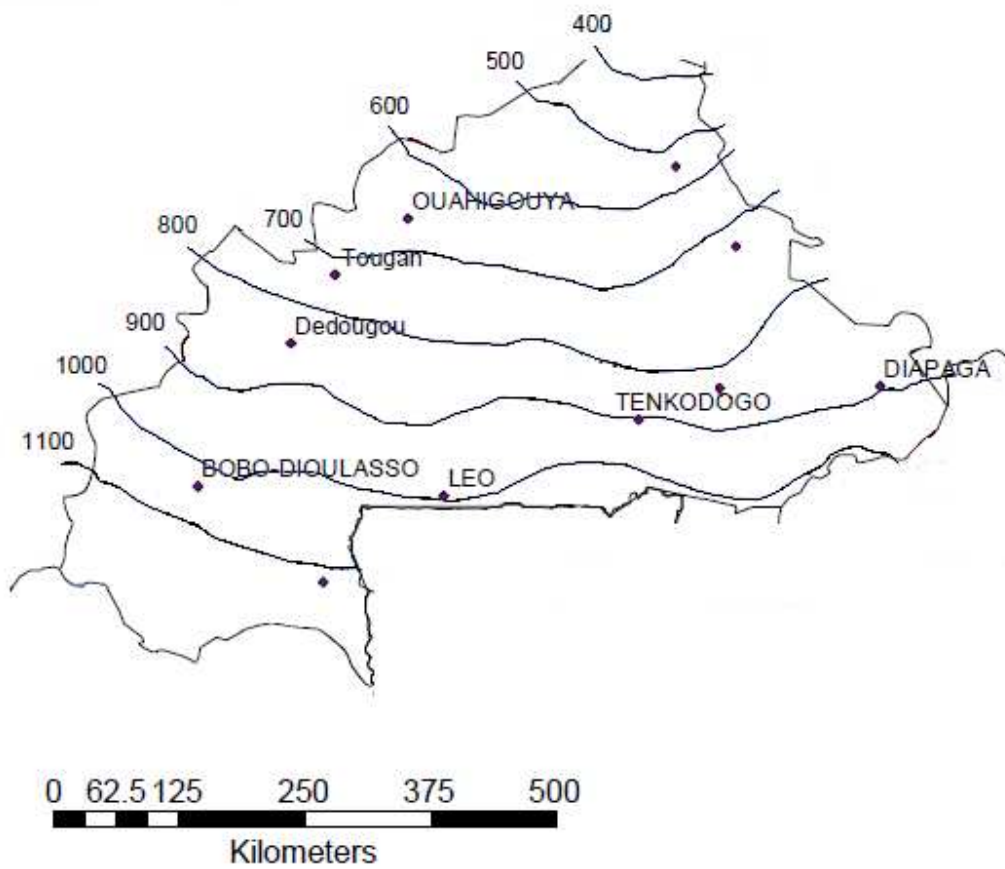
Figure 2.2: Mean annual rainfall (mm) in Burkina Faso. The lines over the area are the isohyets, connecting locations with equal amount of rain to each other [2.1]
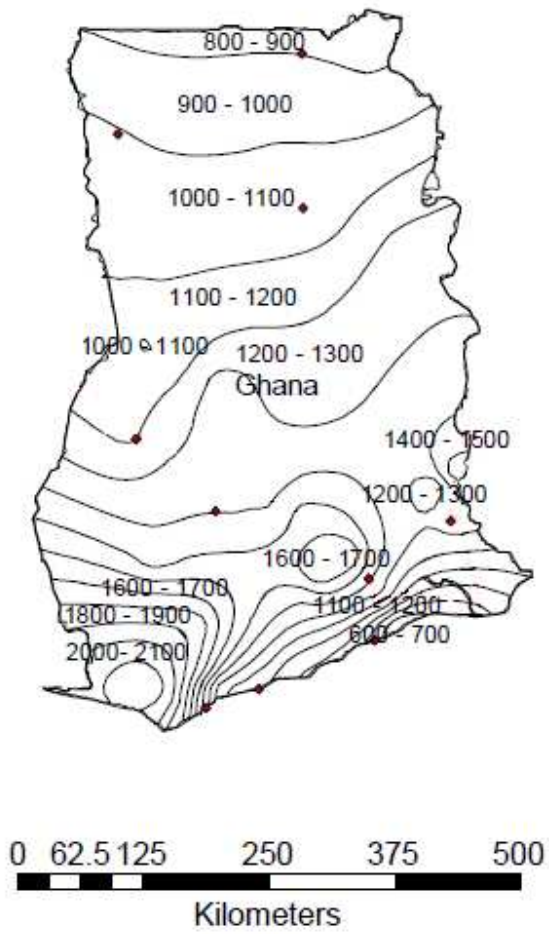
Figure 2.3: Distribution of rainfall (mm) in Ghana in 1990. The lines over the area are the isohyets, connecting locations with equal amount of rain to each other [2.1]
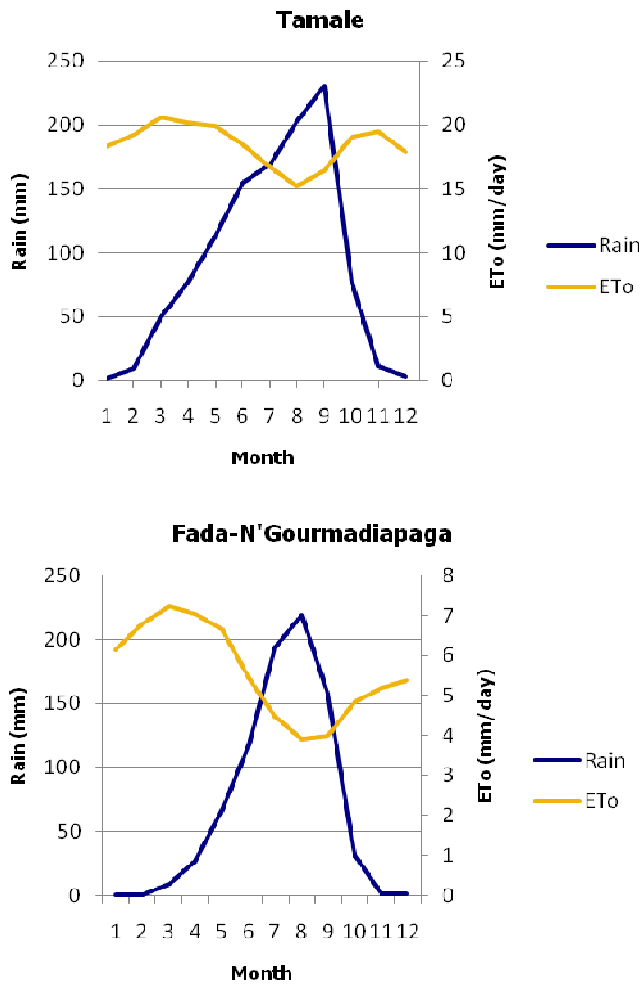
Figure 2.4: Rainfall (Rain) and potential evaporation (ETo) in Tamale (top) and Fada-N'Gourmadipaga (bottom) [2.2].

2.2.2 VEGETATION

Vegetation in Ghana and Burkina Faso changes from north to south, due to the differences in annual precipitation. Figure 2.5 shows the different vegetation types. The area that is under consideration in this research is mainly defined as savanna. Ghana is mostly covered with Guinea savanna, while Burkina Faso's main cover is Sudan savanna.
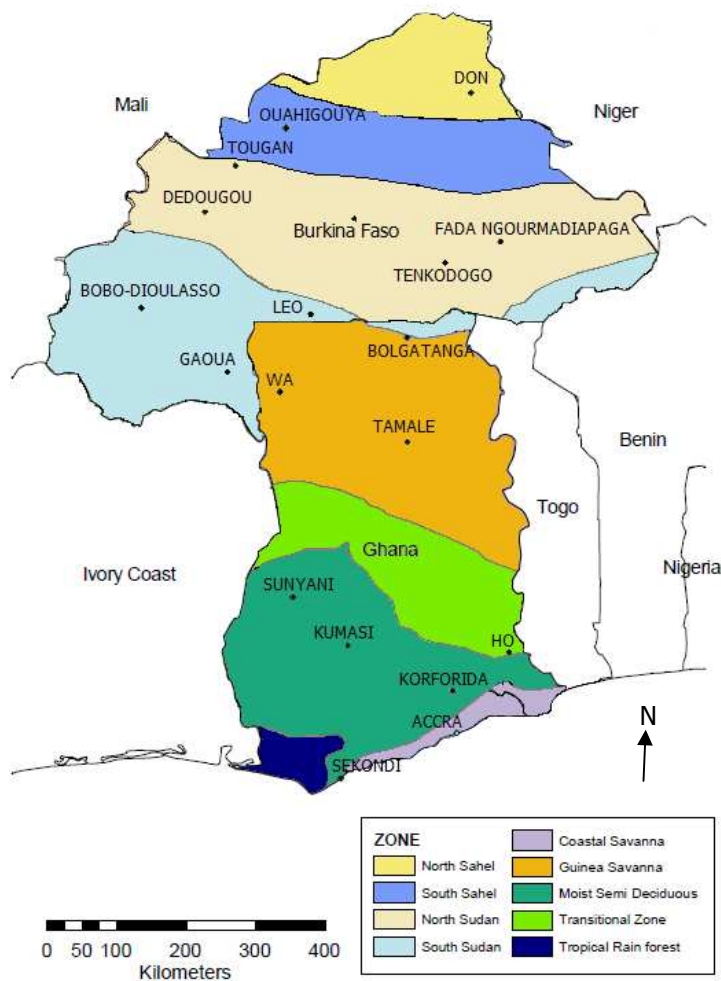
Figure 2.5: Vegetation types in Ghana and Burkina Faso. South Sudan and North Sudan are both savanna types [2.1]

The definitions of the two types of savanna are:

Guinea savanna: "a tree Savannah or a continuous grass cover interspersed with generally fire resistant, deciduous, broad-leaf trees. In the most luxuriant form the trees show varying completeness of canopy. The grasses associated with the Interior Savannah are not uniform but differ according to soil type and moisture regime. The grassy background of the zone is invariably dominated by Andropogon gayanus with Hyparrhenia and Schizachyrium as co-dominants in some areas. The tree cover includes Butyrospermum, Khaya, Ceiba, Pterocarpus, Parkia, Anogeissus, Diospyros and Adansonia" [2.3].

Sudan savanna: The West Sudanian Savanna is a hot, dry, wooded savanna composed mainly of large tree species and long "elephant" grass [2.4]. The areas covered by savanna are used for extensive agriculture and grazing. In the north Sudanian zone the savannas are dominated by

trees which have been retained for mainly economic reasons: Adansonia digitata, Butyrospermum paradoxum, Parkia biglobosa, Tamarindus indica. The most typical grasses are the annuals Andropogon pseudapricus, Loudetia togoensis, Pennisetum pedicellatum, and the perennials Andropogon gayanus and Cymbopogon schoenanthus [2.5]. Figure 2.6 illustrates the differences between vegetation in West-Africa.

In short, this means that both savanna types comprise grasses and scattered trees. However, the Guinea savanna has taller grasses, and taller trees that often grow in clusters. Besides that, Guinea savanna can show thin strips of forest along rivers [2.6].

Figure 2.6: Different vegetation types over West-Africa. Top to bottom: Sub Sudan Savanna, Savanna-Forest Ecotone, Lowland Rainforest (http://www.herpetologie.naturkundemuseum-bw.de/collections/ Africa/htm_docs/hca_feature_2.htm, 24-09-2010)

Validation of SMOS satellite data over Ghana and Burkina Faso

3 THE SMOS ALGORITHM

3.1 INTRODUCTION OF THE ALGORITHM

Satellites cannot measure SM directly. Instead, they measure radiances (emissions) that come from the earth and other sources. These radiances are represented as Brightness Temperatures (TB). TB measured by the satellite is compared to that calculated with the parameters that represent the area that is measured. Since SM is one of the parameters in the equation for TB, SM can be determined by minimizing the difference between the measured and calculated TB.

Radiances coming from the earth have a polarization (the orientation of their oscillations). Horizontal and vertical polarizations can be described by electric fields E_H and E_V. For a full characterization of the polarization state, four parameters are needed. These are described by the four Stokes parameters. The equations for these parameters are:

$$S_0 = I$$
$$S_1 = Ip \cos 2\psi \cos 2\chi$$
$$S_2 = Ip \sin 2\psi \cos 2\chi$$
$$S_3 = Ip \sin 2\chi$$

(eq. 3.1a-d)

I is the energy emitted ($|E_H|^2 + |E_V|^2$). Ip, 2ψ and 2χ are the spherical coordinates of the polarization state, see Figure 3.1.
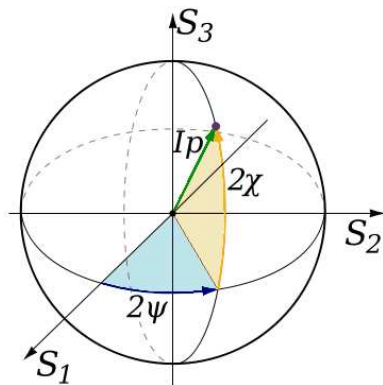
Figure 3.1: The Poincare sphere, the parameterization of the last three Stokes parameters in spherical coordinates (http://en.wikipedia.org/wiki/Stokes_parameters, 31-08-2010)

For low frequencies, like the L-band, radiances are directly proportional (through temperature) with brightness temperature (TB). This means the four Stokes parameters can also be written as four components of TB:

$$TB_H = \gamma_{ST} \left\langle E_H E_H^* \right\rangle$$
$$TB_V = \gamma_{ST} \left\langle E_V E_V^* \right\rangle$$
$$TB_3 = 2\gamma_{ST} \, \mathrm{Re} \left\langle E_V E_H^* \right\rangle$$
$$TB_4 = 2\gamma_{ST} \, \mathrm{Im} \left\langle E_V E_H^* \right\rangle$$

(eq. 3.2a-d)

γ_{ST} is a multiplying factor relating each TB to the electric power density (amount of power per unit volume). The "$\langle\ \rangle$" symbol stands for the ensemble average.

The TB that is measured by satellites can also be calculated by adding up contributions from different sources. These contributions are 1) TB calculated from upwelling atmospheric emission, 2) TB calculated from earth's surface emission, 3) TB calculated from atmospheric down-welling atmospheric emission, reflected (scattered) at the surface and attenuated along the upward path by the atmosphere and 4) TB calculated from cosmic background emission attenuated by the atmosphere, reflected/scattered at the surface and attenuated again along the upward path by the atmosphere.

The contributions of the earth's surface, coming from the ground and vegetation, can be determined with values of parameters that represent these surfaces. For the SMOS satellite, this is done with the so-called τ-ω model, in which vegetation is represented as a single layer. The model is based on two parameters: the optical depth τ and the single scattering albedo ω. These are used to parameterize, respectively, the vegetation attenuation properties and the scattering effects within the canopy layer. The reflection at the top of the canopy (at the vegetation–atmosphere interface) is neglected.

As mentioned before, one of the parameters in the equation for TB is SM. The TB measured by the satellite is compared to that calculated with the parameters that represent the area over which SM is determined. By subtracting the two from each other, a cost-function is created that is minimized in order to find the SM and τ value. This should lead to a SM value that falls with a range of +/- 4% of the 'real' value.

The algorithm that is used to determine the SM content of a SMOS pixel is described in the SMOS level 2 Processor for Soil Moisture Algorithm Theoretical Based Document (ATBD). The τ-ω model cannot be used for all circumstances. Exceptions and how to work around them are mentioned in the ATBD.

The following paragraphs, together with Appendices A and B, give a short summary of the information on the algorithm in the ATBD. Paragraph 3.2 however starts with a short explanation on the measurement aperture in the SMOS satellite.

3.2 SATELLITE APERTURE

The SMOS satellite measures emissions on the L-band. The L-band covers the frequency from 1 to 2 GHz. The L-band is the lowest frequency band that is still protected. Theoretically, that means the band is not used for radio communication that can cause interference with the natural signal (Radio Frequency Interferences (RFI)). The lower the frequency, the greater the sensitivity of TB to SM and the smaller to vegetation and surface roughness.

The measuring aperture in the SMOS satellite is called the Microwave Interferometric Radiometer with Aperture Synthesis (MIRAS) instrument. It has three arms connected to a central core. The angle between the arms is 120°. Every arm has 18 L-band radiometers. On the core, the line is extended with four more radiometers. Besides that, the core contains three noise injection radiometers, used for calibrations of the initial measurements and as an addition to the radiometers. Figure 3.2 gives a diagram of MIRAS.
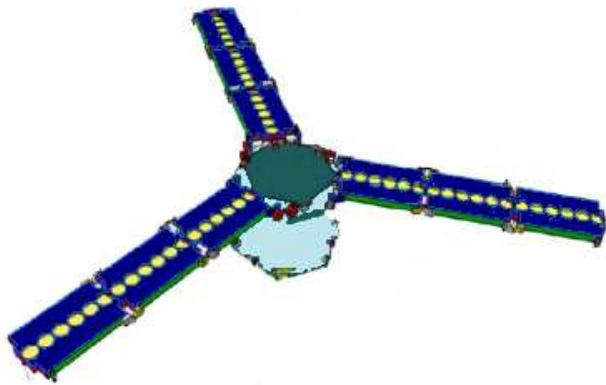
Figure 3.2: MIRAS configuration diagram [1.3]

There are two different observation modes: 1) dual-polarization and 2) full-polarization. For the first, TB of the same polarization (H or V) is measured by all radiometers over an integration period of 1.2 s, after which the other polarization is measured. In the second, TB_3 and TB_4 are also acquired. This is done with 4 integration periods. In the first two, all radiometers measure the same polarization, in the second two, one of the arms measures the opposite polarization from the other two arms [3.1]. Measurements created over an integration period are called a view. Both dual- and full-polarization can be used to determine SM.

In the period that this report was written, there was not much known on the behavior of TB_3 and TB_4. The rest of this chapter therefore explains dual polarization. Most of the explanation also holds for full polarization, since TB_3 and TB_4 are assumed to be very close to zero over land. Because of the way the measurements are done however, for full polarization, measurements sometimes have to be seen as indissociable pairs. Views have to fulfill some requirements in order to be used for the determination of SM. For full polarization, when one of the views of a indissociable pair does not fulfill the requirements and is discarded, the other should be discarded

as well. Elimination criteria for views are of course adapted to imaginary views for full polarization.

There is a difference between *TB* measured by satellites and *TB* at the top of the atmosphere (*TOA*). This is due to geometrical considerations and to the Faraday rotation induced by ionosphere. How the conversion from *TB-TOA* to that measured by the satellite is done depends on what kind of measuring mode is used. The same goes for the determination of *SM*.

It should be noted that for full polarization, the quality of the second integration period is different from the first, which can have an impact on the value of thresholds and coefficients that control the algorithm (see paragraph 3.5 for more information on thresholds and coefficients). Full polarization does however give more data than dual polarization. As an extra bonus, it might be possible to determine RFI with TB_4.

At the end of this master thesis it became known that any further *SM* retrievals by ESA will be done with data from the full polarization mode.

3.3 THEORY

As mentioned before, the SMOS satellite collects directional polarized radiances, directly proportional to brightness temperature. The brightness temperature from a surface is (surface) temperature times (surface) emissivity.[3] The brightness temperature measured at the top of the atmosphere also contains atmospheric emission and consists of four components. It can be calculated with the radiative transfer equation.

$$TB_p = TB_{atu} + TB_{sp} \, exp(\tau_{atu}) + (TB_{atd} + TB_{sk} \, exp(\tau_{atd})) r_{sp} \, exp(-\tau_{atu})$$
(eq. 3.3[4])

With
TB_p: SMOS observed *TB*, "p" subscript indicates polarization
TB_{atu}: *TB* from up-welling atmospheric emission [K]
TB_{sp}: *TB* from earth's surface emission [K]
TB_{atd}: *TB* from down-welling atmospheric emission reflected (scattered) at the surface and attenuated along the upward path by the atmosphere [K]
TB_{sk}: *TB* from cosmic background emission attenuated by the atmosphere, reflected /scattered (r_p) at the surface and attenuated again along the upward path by the atmosphere [K]
τ_{atu} = upward path atmospheric optical depth [-]

---

[3] The process of retrieving brightness temperature from radiances not considered in this report. The ESA publishes among others L1c data, which are already validated temperature brightness values.

[4] the subscript "p" stands for polarization, the subscript "s" for combined (surface + near surface) layers

τ_{atd}= downward path atmospheric optical depth [-]
r_{sp} = integral of the surface scattering coefficient over all scattering directions [-] (with *1-e=r*)

At L-band (1.4 GHz), the frequency at which the SMOS satellite measures, atmospheric effects are small. TB_{atu} and TB_{atd} can therefore be considered equal to TB_{atm}. The same goes for τ_{atu} and τ_{atd}. These can be noted as τ_{atm}.

The earth's surface is often covered by some kind of layer. This layer can be vegetation (e.g. crops or trees), but also litter (dead vegetation laying on the ground) or snow. Any surface variable can be taken into account by looking at its influence on the surface reflectivity r_{sp} and its direct influence (through temperature, scattering and attenuation) on surface brightness temperature TB_{sp}.

If the soil/atmosphere interface is considered as a simple layer, the measured *TB* can be calculated by taking into account the contribution of canopy to the terms of eq. 3.3. Canopy changes the terms as follows:

1. The up-welling atmospheric emission stays the same;
2. The earth's surface emission now consists of three parts:

   - Soil-surface emission attenuated through canopy and the atmosphere: $e_{gp}\ T_g\ exp(-\tau_c)\ exp(-\tau_{atu})$;
   - Upward layer (canopy) emission attenuated through the atmosphere: $T_c(1-\omega)(1-exp(-\tau_c))\ exp(-\tau_{atu})$;
   - Downward layer (canopy) emission scattered at the surface and attenuated through the layer and the atmosphere: $T_c\ (1-\omega)\ (\ 1-exp(-\tau_c))\ exp(-\tau_c)\ exp(-\tau_{atu})\ r_{gp}$;

3. The (reflected) down-welling atmospheric emission and the cosmic background emission are now also attenuated through the canopy layer. The third term therefore becomes r_{gp} $(TB_{atd} + TB_{sk}\ exp(-\tau_{atd})\ )\ exp\ (-2\ \tau_c)\ exp\ (-\tau_{atu})$.

We thus write (assuming downwards and upward atmospheric contributions are equal):

$TB_p = TB_{atm} + exp(-\tau_{atm})(TB_{atm} + TB_{sk}\,exp(-\tau_{atm}))r_{gp}exp(-2\tau_c)$
$+ exp(-\tau_{atm})[e_{gp}T_g\,exp(-\tau_c) + T_c(1-\omega)(1-exp(-\tau_c))(1+r_{gp}exp(-\tau_c))]$ (eq. 3.4)

With:
τ_c = optical depth, defined by layer extinction and scattering coefficients κ_{ext} and κ_{sct} through

$\tau_c = \int_0^{LHsec(\theta)} \kappa_{ext}\,dx$ [-] (eq. 3.4a)

ω = isotropic single scattering albedo defined by layer extinction and scattering coefficients κ_{ext} and κ_{sct} through $\omega = \kappa_{sct}/\kappa_{ext}$ [-] (eq. 3.4b)

T_g = weighted sum of soil temperatures at subsurface levels accounting for the penetration depth [K]

T_c = canopy physical temperature [K]

This model is called the τ-ω model.

It should be noted that often the layer will actually consists of several layers. For example, there might be a layer of litter above the soil's surface, then a layer of snow, a vegetation layer, trees and finally snow on these threes. The τ-ω model models all these layers as if it was one.

The area over which TB is measured, a SMOS pixel, ranges in size from 25 x 25 km to 60 x 60 km. The chance of a certain area of this size being homogeneous is very small. Eq. 3.4 therefore has to be adapted for the case where a variety of surface types contributes to the measured brightness temperature. The area is divided into different parts that can be considered homogeneous. The contribution of the different parts is weighed by its intra-pixel cover fraction and added to the total. If we consider a mixed scene with n = 1 to NF mean fractions FM_n, eq. 3.4 can be rewritten as:

$$TB_p = TB_{atm} + exp(-\tau_{atm})[TB_{atm} + TB_{sk} exp(-\tau_{atm})]R1 + exp(-\tau_{atm})R2 \qquad \text{(eq. 3.5)}$$

With $R1$ (dimensionless) being the sum of all $r_{gp}exp(-2\tau_c)$ and $R2$ (Kelvin) the sum of $e_{gp}T_g exp(-\tau_c) + T_c(1-\omega)(1-exp(-\tau_c))(1+r_{gp}exp(-\tau_c))$ over all fractions.

As mentioned before, over every integration period of 1.2 s of the SMOS satellite, a so-called view is created. These views differ slightly from each other, because of the movement of the satellite. The angle and location of the satellite with the earth change for every view. This means the area of the different fractions changes with every view. Several views are summed up to get to a final TB measurement for a SMOS pixel. In order for eq. 3.5 to approach a view's TB, $R1$ and $R2$ have to be calculated over the fractions of that specific view. These fractions are called incident angle dependant fractions, FV_n and are computed as follows:

$$R1 = SUM_{n=1:NF}\{FV_n R1_n\}$$
$$R2 = SUM_{n=1:NF}\{FV_n R2_n\}$$

(eq. 3.5a and b)

Looking at what $R1$ and $R2$ represent, it becomes clear that the integral of the surface scattering coefficient over all scattering directions (r_{sp}) is the key variable. This means that for any surface, the interpretation of r_{sp} and T_{se} must be further developed in order to calculate the temperature brightness of a pixel.

Different formulas for amongst others reflectivity exist for different kinds of surfaces. These are given in Appendix A. Which formulas are used depends on the so-called forward model that is used. Paragraph 3.4 explains this concept.

3.4 FOUR DIFFERENT FORWARD MODELS

One of the main parameters to determine the reflectivity of a surface is the dielectric constant. The dielectric constant can be determined in several ways. The ATBD distinguishes four different ways, and has formulated models that use them, depending on the kind of surface of a SMOS pixel. The forward models that exist are:

1. The nominal model (MN): This model is used to calculate vegetated soil. The reflectivity is written for a smooth surface as a function of the complex dielectric constant $\varepsilon = \varepsilon' - j\varepsilon''$. The dielectric constant is written as a function of physical parameters, including surface SM. In the future, it might include a variant for sand.
2. The water model (MWP without salinity, MWS with salinity): This model is used for open water. Again, the basis is the dielectric constant. The constant is however calculated in a different way (with among other parameters the salinity of the water). It can also be used for transparent snow.
3. The cardioid model (MD): Sometimes, the dielectric constant cannot be written as a function of physical parameters. An example is iced surfaces. In these cases, the cardioid model can be used. This approach is a regularization of the retrieval problem.
4. The snow model: This model is not validated yet. There is no information on this model in the ATBD.

The choice for a forward model depends on what kind of surface an area has. However, some surface can change over time. An area can for example be flooded or covered with snow. Several steps have to be taken in order to check whether an area's SM content can be calculated and if so, with what model. This process is done by following the so-called decision tree.

Through this decision tree, the so-called retrieval case is determined. The ATBD knows 19 retrieval cases that represent the most important fraction of a SMOS pixel. In turn, the retrieval case determines which model is used for a specific fraction.

As mentioned before, the calculated TB is subtracted from the measured TB. This cost-function is minimized in order to determine SM. This process is called retrieval. To keep the calculation simple, some fractions' contributions are assigned to be default. They are not changed in the minimization.

The decision tree is described in paragraph 3.5. It contains the steps that need to be taken in order to calculate the contributions of fractions to the brightness temperature (including the choice for a model) and the final calculation of the brightness temperature.

3.5 THE DECISION TREE

The decision tree gives an overview of the steps that need to be taken and the values that need to be checked, before any calculations can take place. These steps and the relevant thresholds[5] that control the decision tree are mentioned in the ATBD. The steps are:

1. The weighted mean aggregated fractions (FM0 and FM) are determined. The difference between the two is that the first does not depend on the incidence angle (the angle of the radiance with the line perpendicular to the surface), while the second one does. FM0 are used to drive the decision tree, FM for actual calculations of default or a priori contributions to *TB*. These contributions are determined with so-called reference values, that are in turn determined with auxiliary data.

   FM0 can be computer in advance with auxiliary data as well. However, these fractions can change. River forelands can be flooded, snow can fall, etc. This changes the area of certain fractions, but can also have an effect on reference values like optical depth. To determine the actual magnitude of FM0 and FM, some tests are done based on thresholds. E.g. when temperature drops below a certain value, it is assumed water freezes over. Reference values and fraction values are changed accordingly. In determining the magnitude of fractions and the retrieval case, it is important to keep to the sequence that is given in the Table Generation Requirement Document (TGRD)[6] [3.2], so that the final FM(0) are correct.

2. With the magnitude of the fraction known, the retrieval case is determined, again with the use of thresholds. The retrieval case determines with what model a fraction's contributed is calculated and with what model default contributions are being determined.

   Because a pixel can have a large variety of land use, it is not feasible to execute the *SM* retrieval over the whole area, that would cost too much calculation time. Therefore, pixels are usually divided into an area where retrieval will take place and an area that is considered fixed (the default contribution). If for instance the largest part of a pixel is covered with low vegetation, but there is also a lake, the contribution of the lake is

---

[5] Thresholds are usually noted as TH_**

[6] The TGRD "describes the requirements and generation methods for the tabulated geo-physical parameters required for the Soil Moisture and Ocean Salinity (SMOS) Level 2 (L2) processing, in order to support the retrievals of the Soil Moisture and other user-defined land surface geo-physical properties"

estimated (using so-called reference value[7]) and assumed constant. Adjusting parameters in the retrieval will in this case only be done for the fraction with low vegetation.

3. The next step is to obtain reference values or a priori constraints for every fraction from auxiliary data.

4. The amount of views determines how many equations in the form of eq. 3.5 exist. Not all views can be used in the final retrieval. Whenever a view covers too large an area or shows values that fall outside a certain range, the views are discarded. Appendix A describes which initial views are eliminated for retrieval.

Depending on how many views there are, the parameters that are kept free in the retrieval process are determined. Their so-called a priori standard deviation (APSD) is also determined based on the available amount of views. The larger APSD, the larger the range of values a free parameter can take on. E.g., as long as there are views of a pixel, *SM* is determined through minimization. The parameter that describes roughness, *HR*, is only determined through minimization if there are more than 34 views of an area. Its new value can differ with at maximum 0.1[8] of its old value.[9] A proposal on which parameter can be determined with what APSD, when the number of views is known, can be found in the ATBD.

3.6 MINIMIZING THE COST FUNCTION

With the theoretical and measured *TB* known, *SM* can be determined. As mentioned before, the calculated *TB* is used in a cost function that needs to be minimized in order to get a value for the *SM* content. The cost function subtracts the modelled *TB* (eq. 3.5) from *TB* measured by the satellite. *SM*, being one of the parameters in the equations for *TB*, is kept free in the function in order to determine the best fit of measured and theoretical *TB*. The ATBD suggests using the Levenberg-Marquardt algorihm for the minimization process [3.3].

---

[7] Reference values for space varying quantities are obtained or computed over the DFFG relevant area for aggregated fractions FV.

[8] 34 and 0.1 are values coming from a draft version of a table that is used for selecting the parameters that may be determined through minimization and the range their value is allowed to take

[9] It should be noted that parameters other than SM can only be determined through minimization when their a priori value falls within a certain threshold.

Validation of SMOS satellite data over Ghana and Burkina Faso

The formula that is used is:

$$COST=(TBM_m - TBF(\theta_m, p_i..))^t [COV_T]^{-1} (TBM_m - TBF(\theta_m, p_i..)) + \sum_i \frac{[p_i - p_{i0}]^2}{\sigma_{i0}^2}$$ (eq. 3.6)[10]

With:

TBM_M = valid measured values of brightness temperature [K]

$TBF(\theta, pi...)$ = direct (forward) model for brightness temperature at the antenna level [K]

$[COV_T]$ = is the variance matrix for the observed TBM [K]

θ = incidence angle [rad]

p_i = physical parameters [depends on parameter]

p_{i0} = prior estimates of the free physical parameters [depends on parameter]

σ_{i0}^2 = prior variances of the free physical parameters [depends on parameter]

Appendix B gives a more detailed overview of the decision tree and the COST formula.

3.7 DEFAULT VALUES FOR PARAMETERS

Of the four different models, the Nominal model is the most important. The parameters used in this model are defined in Appendix A. As mentioned before, default values for these parameters are determined in several experiments.

---

[10] Superscript "t" stands for transposition

Table 3.1 gives an overview of the different experiments. The column 'experiment' gives the sites the experiments were executed and its characteristics. It describes for what circumstances (bulk density soil, soil roughness, vegetation) the results are valid. The column 'parameter' shows which parameter was studied during the experiment. The following columns shows values for vegetation parameters found in these experiments (mean, standard deviation, range and number of measurements, if known). These values are compared to the value that is given as a default in the ATBD. The word 'literature' in the sixth to ninth column refers to literature cited in the last column.

For some parameter, the table shows a large difference between default values in the ATBD and values found in experiments. E.g. b0 has a value of 0.3 according to the ATBD, but a mean value found over the SMOSREX site of 0.63. This is because the ATBD tries to give values that are valid in more general cases, and over areas that are larger than the respective test sites. On the other hand, the values from literature are only valid for areas that show the same characteristics as the test sites. The ATBD does however not mention how the default values are obtained.

Furthermore, none of the experiments in the table were done in Africa. The characteristics of the sites do not necessarily conform to those in Ghana and Burkina Faso. For both sources, it is thus unclear how valid the values are for an area like West-Africa. Table 3.1 thus is an illustration of the uncertainty that still exists around the values for the different parameters. This in turn can lead to errors in the *SM* determination.

The following chapter pursues this subject. It describes the sensitivity analysis that is done on the different parameters of the nominal model, in order to see what would be the actual effect of an parameter's erroneous value.

Table 3.1: Values for different parameters; default ATBD values compared to mean value, standard deviation and range from experiments described in literature

Experiment	Parameters	Remarks parameter	No. of measurements	Default value ATBD	(Mean) value literature	Standard deviation literature	Range literature	Reference
SMOSREX, Toulouse, France: bare plot and grass plot, bulk density 1.5-1.7, sand fraction 0.27-0.37, clay fraction 0.16-0.29 (depending on depth), surface height standard deviation 9.2, roughness correlation length 22.4 m	w0	texture parameters to calculate effective soil temperature from temperatures at different depths, values from literature are over different seasons	Data from 1 year, different seasons	0.3	0.36	0.05	0.32-0.47	[3.4]
Avignon Remote Sensing Test Site: 7 different plots (18-40 m, bare fields), bulk density 1.2-1.43, all 62 % silt, 11% sand and 27 % clay. Field were cultivated for different surface roughness conditions	w0	see above, values come from Tsurf taken at different depths	data from 3 months, temperature measured continuously, but problems	0.3	0.62	0.36	0.27-1.08	[3.5]
	b0		with meters not working	0.3	0.31	0.09	0.26-0.5	
	QR	roughness polarization coupling factor, related to the fact that roughness tends to induce polarization mixing		0	0	-	-	
	NRp	roughness power law of cos(θ), used to parameterize the dependence of the roughness effects on incidence angle		0-2	0	-	-	
Combination of Avignon data and data from Beltsville Agricultural Research Center (BARC, bulk density 1.2, clay fraction 15%, sand fraction 67%): soybean, corn, alfalfa, and grass, measured over 1 month at BARC, soybean, wheat (Portos 1991 experiment) and corn (Portos 1993) measured over the whole vegetation period at Avignon.	ωp	scattering albedo for agricultural canopy (low vegetation)	INRA: corn 1144, wheat 400, soybean 320; BARC: corn 392, soybean 220, grass 210, alfalfa 238	0	best fit INRA: corn 0.1-0.13 (depending on configuration), wheat 0, soybean 0; BARC: corn 0.06, soybean 0, grass 0.5, alfalfa 0	-		[3.6]
A model was used to obtain time series of the main variables driving the L-MEB model	ωF	scattering albedo for forest		0.1	0.15 for deciduous, coniferous and rain forest	-		[3.7]
Combination of Portos 1991 and 1993, EMIRAD and REBEX (respective bulk densities 1.29, 1.28, 1.1; sand fractions 11%, 13.2%, 16%; clay fractions 27%, 32.8%, 29%) The EMIRAD cornfield was about 80×120 m. REBEX experiments were conducted during the spring, summer, and autumn of 2001 over an 800×400 m corn field in southeastern Michigan. Measurements were mainly carried out at five times: over bare soil, during vegetation growth (twice), at maximum biomass, and during senescence.	b' and b''	This b gives the relationship of τ with LAI	Data from 2.5 months. PORTOS-91 was obtained over soybean, PORTOS-93 over wheat, EMIRAD-2001 over corn, over corn and soybean .REBEX over corn	b' 0.3, b'' 0	Best fit after determining τ (and soil moisture) b'' is generally very low and the slope b' is strongly dependent on the crop type. Respectively wheat 0.061 and 0.034, soybean -0.020 and 0.090, corn 0 and 0.047	-	corn is measured in two configurations and on two different locations. The values found are: 0.039 (EMIRAD, parallel), 0.047 (EMIRAD perpendicular), 0.055 (REBEX)	[3.8]

4 SENSITIVITY ANALYSIS OF THE SMOS ALGORITHM

4.1 INTRODUCTION

This chapter describes the sensitivity analysis that is done on the SMOS algorithm's parameters and its results. A sensitivity analysis shows the effect of the change of a parameter's value on the outcome of the SMOS algorithm. If a change in value causes a large change in SM value, the value of a parameter should be determined within small error bounds in order to determine the correct *SM* value of an area. If changes are small, it is less important to determine the exact value. This information makes the sensitivity analysis a helpful tool for calibration of the SMOS algorithm and the design of future field experiments.

This chapter tries to give an answer to sub questions 1 and 2 as formulated in paragraph 1.3:

1. How sensitive is the algorithm to changing values of parameters?
2. Which parameters are the most influential for the region of northern Ghana?

The sensitivity analysis is partly executed with the so-called L2PP processor v.3.07, developed by Array, partly with a code in Matlab. The processor minimizes the cost function described in chapter 3. However, not every parameter can be changed in this processor. The sensitivity of the algorithm for these parameters is determined in Matlab. This code is a simplification of the SMOS algorithm. The reason for this simplification is that for many parameters, there was no information available in a format manageable in Matlab. Examples are data on the satellites orbit, temperature of the surface and vegetation, pressure data and *TB* contributions of sky and atmosphere. . In the following, the L2PP processor is denoted as L2PP or processor, the Matlab code as code.

It should be noted that changes of *SM* due to a changed parameter also depend on the extent to which an area is homogeneous, e.g. when an area is partly covered with forest, partly with low vegetation. In that case changing a forest parameter will not change the contribution of the area with low vegetation.

4.2 METHOD

4.2.1 PROCESSOR VERSUS CODE

The calculation of the processor and the code differ from each other. The main reason for this is the fact that there is a lack of data for the code. The most important missing data for the processor and there consequences are the following:

1 Instead of TB data for every view over a pixel, only a single value for TB_H and TB_V were available. These TB values have an incidence angle of 42.5°. Just as with the processor, a limited amount of data limits the amount of parameters that is determined through minimization. Therefore, the only free variable in the Matlab code is SM.

2 The information given in the TB datasets is already resampled, having a resolution of 15 by 15 km. The cells of this resolution are called L1c pixels or nodes. On top of that, too little is known of the area around the pixel over which the sensitivity analysis was done. Any default contributions to TB within an area of 123 by 123 km could therefore not be determined correctly. The L2PP processor looks at an area of this size around a pixel to determine SM. For the Matlab code, it was decided to determine SM by using only information from the pixel itself. It is assumed that the complete area of the pixel can be calculated with the nominal model and no default contributions are added.

3 As mentioned, no correct information was available on contributions from atmosphere and sky. Because of this, in the Matlab code a contribution was added to the formula that accounts for this missing data. More information on the L2PP processor, the Matlab code and how the contribution was determined can be found in paragraph 4.2.3 and Appendices C and D.

4.2.2 LOCATION

Changes in SM content due to changed parameters are compared for the SMOS pixel that is also used in the field work (see chapter 5). The centre of this pixel has coordinates lon 1°07'54" W, lat 9°11'44" N. This is in northern Ghana (Figure 4.1).

Figure 4.1: Northern Ghana with location sensitivity analysis. On the bottom of the picture (to the right), the northern part of the Volta lake is visible (http://maps.google.com/, 31-08-2010)

The vegetation definition comes from auxiliary data files from the L2PP processor. Vegetation was defined in the Discrete Global Grid (DGG, the grid of the L1c pixels) Fraction file with no. 49, NH Africa dry tropical (WG). WG stands for woody grassland. No other vegetation was defined for this node.

For this analysis, only DGG fraction 49 was adapted. As mentioned above, the sensitivity of a specific area for e.g. the optical depth of low vegetation will not only depend on the range of this parameter. It also depends on the percentage of the area that is covered by low vegetation. Different fractions represent different vegetation and therefore have different ranges for parameters. With more than one main fraction, e.g. forest and low vegetation, both fractions have an influence on the *SM* content. Changing the parameter of one of the fractions cause a different change in *SM* than for a homogeneous pixel. The sensitivity analysis done with the code thus does not give the sensitivity of the complete algorithm, but merely of the nominal model.

4.2.3 MISSING CONTRIBUTIONS ADDED TO MATLAB

Although the calculations of both the processor and the code are based on the algorithm described in chapter 3, there are some important differences. Because of lack of data for the code, the Matlab code can only determine *SM*, no other parameters. Furthermore, it uses only information from the node it is determining *SM* for. Finally, contributions that cannot be determined because of missing data are added as a single contribution.

In order to make the outcome of the processor and the code comparable, an artificial contribution was added based on a best fit for several nodes in the processor's data files. First, *SM* was determined for the node used in this analysis and its six surrounding nodes with the processor. No other pixels were taken into account, because these pixels might have different (vegetation) characteristics, making it harder to obtain a good fit with the code. All missing data (e.g. temperature, optical depth) needed for the code were taken from this processor's output file. With this information, *SM* is determined for the same nodes with the code. In order to get comparable *SM* values, a contribution was added to the code. This contribution was added as *TB*, in order to minimize its effect on the outcome of the algorithm. By trial and error, the *TB* contribution that gave the smallest difference with the L2PP processor's outcome over all seven pixels was determined. This was done for each of the 4 dates calculations were done. Because the differences are the same for the calculations on a certain date, this does however not affect the validity of the outcome of the sensitivity analysis.

4.2.4 DATA

For the sensitivity analysis, dual polarization data was used from 14th December 2009, 6th April 2010 and 20th May 2010 (morning and evening overpass). Both morning and evening overpasses are at this point used in the official *SM* determination of ESA. By comparing the outcome of different dates, changes due to other causes than the changed parameter are pinpointed. This is mostly done as an extra check on the Matlab code.

Default auxiliary data used for the calculations with the L2PP processor is mostly test data as is delivered with the processor, all dated January 1st 2005. The auxiliary files used give information on: fractions, *LAI*, flooded areas, roads, roughness data of the soil, optical depth of vegetation, DGG nodes, weighing factors of the different fractions, orbit scenario, soil properties, sky radiation. For the so called European Centre for Medium-Range Weather Forecasts (ECMWF) file, containing weather conditions over the area, a file from the Brockmann site was used [4.1]. The file contains data on soil temperature, wind speed, sea ice conditions, snow conditions, surface pressure and water vapor. This file is dated October 21st 2009.

Although the ECMWF data is from another date than the SMOS *TB* data, this does not affect the sensitivity of the parameters. If more realistic ECMWF values would have been used, these would be the same for every calculation. The same goes for the ECMWF file used in this analysis. The change in *SM* content is therefore only dependent on the changed parameter.

4.2.5 PARAMETERS ANALYZED

Parameters in the algorithm are determined in different ways. Some are determined through satellite imagery, others are kept at default values. Again others are determined and verified in the past in research not linked to the SMOS satellite. In order to keep the amount of parameters manageable, these latter parameters are not analyzed. An overview of these parameters is given in the Table 4.1. It starts with the temperature of both ground and canopy. After that, the parameters that model litter and vegetation (*a_L* till *ω*) are denoted. The next three parameters are used for the representation of the soil. R_{tt} finally models the dependence of the optical depth to the incidence angle of the measurement.

Parameters are first tested by using extreme values. These come from the ATBD. E.g. surface roughness is calculated with an empirical determined formula. The effective roughness of an area is incorporated by the dimensionless parameter HR:

$$HR = (2\,k\,\sigma)^2 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(eq. 4.1)}$$

With:
k = wave number
σ = surface RMS height

In the L2PP processor, when there are enough views (see Appendix B) *HR* is determined in the minimization process. However, the maximum value HR can acquire can be set. For the first step of the sensitivity analysis, this HR_{max} is changed from 0.1 to 0.05 and 5.

If the value of the *SM* content changes because of these extreme changes, values that better represent the research area are used, to see how important it is to determine the value of the parameter exactly. The first step of the analysis is further noted as extreme analysis. The second step is called the location analysis.

Both the influence of some parameters and the Levenberg-Marquardt are not linear. A big difference in *SM* value for the two location analysis values therefore does not necessarily mean that a small change of the parameters value within this range will change the *SM* value

significantly. Therefore, for those parameters that are considered most influential due to the location analysis, graphs are created that show the influence of a consecutive range of values on *SM* of these parameters. Conclusions in this chapters should be read keeping this non linearity in mind and thus be met with caution.

More information on the different parameters and the values used in the two analyses is given below in Table 4.1. There is a difference between *TB* that is measured by the SMOS satellite, and *TB* calculated with the help of the SMOS algorithm. Whenever in the following nine paragraphs a change in *TB* is mentioned due to a changed parameter, the calculated *TB* is implied. It is also assumed that only the discussed parameter changes.

Table 4.1: Parameters that are changed for the sensitivity analysis, their name/description, for what they are used and their default value

Symbol	Name/description	Used to/in	Default value
T_g	Soil temperature	Formula for TB	ECMWF
T_c	Canopy temperature	Formula for TB	ECMWF
a_L	Parameters used to compute litter moisture content (from soil moisture)	Computing moisture content of litter	2.33
b_L			0
B_s_L	Dry biomass of litter	Computing water content of litter	0.3 kg/kg
c_L	attenuation coefficient of litter(c_L = τ_L / LWC)		0
b'	parameter of the relation τ_S / LAI	Determine optical depth standing vegetation	0.3 m2/m2
b"			0 m2/m2
ω	Single scattering albedo	Formula for TB	0
ρ_s	Soil particle density	Determine dielectric constant ε, in turn used to determine reflectivity soil	2.664 g/cm3
μ_s	Magnetic permeability soil	Determination of reflectivity soil	1
HR	Surface roughness	Adaptation of reflectivity of smooth soil to rough soil	0.1
R_{tt}	ratio of angular correction parameters (at H and V polarization) accounting for the dependence of τ_{SP} on incidence angle	Determine optical depth standing vegetation	1

4.2.5.1 CANOPY AND SOIL TEMPERATURE, T_C AND T_G

Every surface has a certain *TB*, that can be calculated (in its most simple form) by multiplying the temperature with the surface emissivity. Temperature thus has a direct influence on *TB*. The higher the temperature, the higher *TB*.

The sensitivity analysis for these parameters was executed with the Matlab code, because it was not possible to change the ECMWF data files, that include temperature data. For the default value of the two *T*'s in the Matlab code the surface temperature from the *SM* data as calculated by the processor is used. This value is 298.9 K (rounded off).

The ATBD does not give information on values for temperature. Temperatures for the analyses were therefore taken from the L2PP processor and measurements during the field work described in chapter 5. The processer gives a minimum and maximum temperature of respectively 230 K and 320 K. These values are used for the extreme analysis.

During the field work (see chapter 0), temperature measurements were done around 6 am and 6 pm, in order to get representative values for the satellite's time of overpass for the location analysis. In the morning, temperature ranged from 25 to 30°C, depending on the kind of surface. High vegetation and shaded areas showed the lowest values, rocky soil the highest. In the evening, temperature ranged from 32 to 38°C. These temperatures are lower than temperatures measured during the day, and the difference between soil and canopy temperature is smaller. The average difference between the temperatures is around three degrees, which is in agreement with the difference in temperature according to the ECMWF data.

Based on field measurements, the values for the location analysis were determined. For overpasses in the morning, for canopy a temperature of 25 and of 27.5^{o}C is tested and for soil 27.5 and 30°C. For overpasses in the evening, the canopy temperatures tested were 32 and 35.5^{o}C and the soil temperatures 35.5 and 38°C.

4.2.5.2   PARAMETERS FOR DETERMINATION OF LITTER WATER CONTENT, A_L, B_L, B$_S$_L

Litter is dead material on the ground. This layer of litter can have a big influence on the radiance of an area. This is because its attenuation properties can be different from that of standing vegetation. A litter layer can intercept water, which will then evaporate slower than with standing vegetation. Besides that, a litter layer has a high volumetric volume compared to standing vegetation. Attenuation properties are strongly dependent on the volumetric volume [4.2]. The Litter Water Content is the amount of water included in the litter layer (kg/m^2). The larger the amount of water intercepted (the larger *LWC*), the larger the optical depth, the lower the brightness temperature. In a way, this is comparable to bare soil, where a larger amount of *SM* causes the brightness temperature to decrease. This leads to a smaller theoretical brightness temperature.

a_L, *b_L* and *B$_s$_L* are all three parameters used in the calculation of *LWC Mg_L = a_L*SM+b_L*. *a_L* and *b_L* are used to calculate the moisture content of the litter layer (*Mg_L*), which in turn is used to compute the water content of litter (*LWC*), together with *B$_s$_L* (*LWC=[Mg_L/(1-Mg_L)]B$_s$_L*). The higher *a_L* and *b_L*, the higher *Mg_L*. The larger *Mg_L*, the larger *LWC*, the larger the optical depth of litter. A larger value of *B$_s$_L* also leads to a larger value of *LWC*.

The values for the parameters are mentioned in the ATBD. For *a_L*, the default value is 2.33, for *b_L* 0 and for *B$_s$_L* it is 0.3 kg/kg. The ranges, used for the extreme analysis, are respectively 0-10, 0-1 and 0-50. Only default values and ranges are mentioned, no research references. It is thus not clear which values are representative for northern Ghana. For *a_L* and *B$_s$_L*, the default value +/- 10% was therefore taken for the location analysis. 10% Is seen as a significant change in value and should therefore, depending on the sensitivity of the algorithm for this parameter, cause a significant change in *SM* value. *b_L* has a default value of 0, making it impossible to do

the same for its location analysis. Instead, the value of 0.1 (default + 10% range) was taken to determine the impact of the parameter. a_L=0 causes Mg_L=0 with a default value of 0 for b_L, thus LWC=0, thus the litter contribution becomes zero. B_s_L=0 leads directly to LWC=0.

The ATBD has set a range for Mg_L between 0 and 0.8 kg/kg. No matter the values for a_L and b_L, Mg_L can never become larger than 0.8 or smaller than 0.

At this point in time, the different parameters have a fixed value. Changes in litter water content (or even the presence or absence of litter) over the year are not taken into account. However, when enough views are available, it is possible to determine the value of other parameters than SM through minimization. One of these parameters is the optical depth of vegetation. The optical depth of litter is a part of the optical depth as determined with the cost-function. For retrieval over time, the idea is to use the value of the optical depth of the previous retrieval as an a priori value for the next retrieval. This is however not done in this sensitivity analysis.

The fact that the optical depth of litter is part of the total optical depth of an area, leads to a complication for the Matlab code. The contribution added to the formula in order to make the outcome of code and processor comparable is based on the optical depth without litter. Adding litter thus might increase the error of the outcome of the code.

4.2.5.3 PARAMETERS FOR DETERMINATION OF OPTICAL DEPTH LITTER AS FUNCTION OF LWC, C_L

This parameters is used to calculate the optical depth as a function of LWC. This coefficient characterizes the attenuation properties of the litter. It depends on litter properties like density or material type.

The default value for c_L is determined with a best fit through SMOSREX data and set on 0.24 m^2/kg. The range for the parameter is set from 0.01 to 1, assuming there is litter. The extreme situation is however no litter, thus (with a fixed LWC) c_L is zero. This value is the default value. For the extreme analysis, only a value of 1 is used. 0.01 and 0.24 are used for the location analysis. The area where the sensitivity analysis was done showed little litter. For the location analysis therefore, only the best fit value was tried.

4.2.5.4 PARAMETERS FOR DETERMINATION OF OPTICAL DEPTH AS A FUNCTION OF LAI, B (B' AND B")

b is used in the calculation of the optical depth of canopy as a function of LAI. The lower the value for b, the lower the value for the optical depth. This gives a higher value for brightness temperature. The parameter is kept constant over time. Changes in optical depth are represented through the LAI value. For the senescence period, LAI can be very low for some vegetation. This can give an underestimation of the optical depth.

The range in the ATBD for b' and b'' respectively is 0.01-1 and 0-3 m^2/m^2. These values are used for the extreme analysis. Experiment show for low vegetation (crops) values range from 0.1-0.3 for b' and 0.03-0.09 for b'' m^2/m^2 [4.3]. These values are used for the location analysis.

4.2.5.5   SINGLE SCATTERING ALBEDO, Ω

The scattering albedo is used to parameterize the scattering of emission caused by vegetation. The larger the scattering, the smaller the temperature brightness, because less radiation is sent straight up (to where a meter can measure it). This reduces the brightness temperature.

Scattering albedos are thought to be small and the default value is therefore 0. This is a value averaged over a large area. The value of the parameter is very much dependent on the kind of vegetation. E.g. vegetation characterized by cylindrical features has a non-isotropic scattering pattern, which causes the scattering albedo to be non-zero. The ATBD gives a range of 0-0.2. However, for grass, values of 0.5 are found [4.4]. Forest on the other hand give a value of 0.15 [4.5,4.6]. Therefore, 0.5 is used for the extreme analysis (as 0 is already used for the default calculation) and for the location analysis the values 0 and 0.2 are used.

4.2.5.6   SOIL PARTICLE DENSITY, P_S

Soil particle density ρ_s is an important parameter in the calculation of the dielectric constant of the soil. The dielectric constant affects the reflectivity of a (smooth) surface. A larger particle density leads to a larger dielectric constant, which in turn leads to a smaller reflectivity. A smaller reflectivity leads to a larger emissivity and thus a higher temperature brightness.

Usually, ρ_s falls between 2 and 3 g/cm^3. These values are therefore used for the extreme analysis. No information was available on ρ_s in the area under consideration. Therefore, for the location analysis, the default as found in the L2PP processor +/- 10% was used, just as it was done for a_L, b_L and B_s_L.

4.2.5.7   MAGNETIC PERMEABILITY, M_S

The magnetic permeability μ_s is the relative increase or decrease in the resultant magnetic field inside a material compared with the magnetizing field in which the given material is located [4.7]. It affects the (smooth) soil reflectivity. Positive values higher than 1 reduce the soil reflectivity and increase the emissivity and the brightness temperature. Positive values lower than 1 do the opposite.

The ATBD only gives a default value for μ_s of 1. The range for μ_s used in the location analysis is deduced from measurements done by Hendrickx *et al.* [4.8]. They have done measurements on the soil magnetic properties in Ghana, including magnetic susceptibility. Magnetic susceptibility is magnetic permeability minus one. Measurements of magnetic susceptibility show a range of 0.17-0.59 e-6 m^3/kg (over all of Ghana). This leads to a range of 1.00000017-1.00000059 H/m for the magnetic permeability. No extreme analysis was done for this parameter.

4.2.5.8   EFFECTIVE SURFACE ROUGHNESS DIMENSIONLESS PARAMETER, HR

HR is the parameter that incorporates surface roughness in the equation. No surface is completely smooth, thus the Fresnel equations, used to calculate reflectivity of a surface, have to

be adapted. The roughness reduces the reflectivity of a surface. A lower reflectivity leads to a higher brightness temperature.

HR as used in the L2PP processor comes from ECOCLIMAP. Ecoclimap is a global database of land surface parameters at one kilometer resolution.

HR is only dependent on surface characteristics, not on vegetation. A way to change *HR* is e.g. by ploughing a field. Because nodes comprise of an area of approximately 15 by 15 km, it is not likely that *HR* will change over a complete area. *HR* is therefore considered constant over time.

The range for *HR* given in the ATBD is 0.05-5. These values are used for the extreme analysis. As mentioned before, in the processor HR_{max} is used. This means that the actual HR of the area depends on the minimization. During the field work described in chapter 5, it was attempted to measure the actual roughness of the area. This unfortunately failed. There are some papers that give information on the value for *HR*. Wigneron, Laguerre and Kerr [4.9] find values (for dry soil) between 1 and 1.8, for different kinds of bare plots. Saleh *et al.* [4.10] define a range between 0.7 and 1.3 for the SMOSREX site, with both bare soil and a fallow site. Because Wigneron, Laguerre and Kerr also look at very deep plowed soil, and the research area in Ghana is a combination of agricultural fields and non-agricultural field, the roughness of Saleh *et al.* is kept as a range for the location analysis. Due to minimization, the actual *HR* values became 0.65 and 1.20.

4.2.5.9   PARAMETER FOR DETERMINATION OF INFLUENCE OF INCIDENCE ANGLE ON OPTICAL DEPTH, R_{TT}

R_{tt} shows the dependence of the optical depth on the incidence angle of canopy. It is defined as tt_V / tt_H, in which tt_V and tt_H are the angular correction parameters for the H and V polarization. For a vertical structure, tt_V can become larger than 1.

Because a SMOS pixel is seldom homogeneous, it is expected that the different values for R_{tt} for the different fractions undo each other's effect. R_{tt} (and tt_H) should therefore be 1.However, research by Wigneron *et al.* [4.11] shows that R_{tt} might get as high as 2.6 for vertical structured vegetation. This value is therefore used to shows the influence of a wrongly assumed R_{tt}.

4.3 RESULTS

Table 4.2 shows the *SM* value as determined with the processor and the code. As mentioned before, the Matlab code used the noted optical depth τ and temperature found in the outcome from the processor.

Table 4.2: *SM* content when calculation is done with all default values, for location where field work is done on different dates. τ mentioned comes from the output of the processor and is used in the code to determine *SM*

	Processor SM [%]	Code SM [%]
14 December, 05:29 h (τ=0.3158)	7.44	6.9
06 April, 05:31 h (τ=0.2013)	8.03	10.0
20 May, 05:15 h (τ=0.1671)	18.78	19.1
20 May, 17:48 h (τ=0.2536)	22.31	23.2

The optical depth in May in the morning and in the evening differs with almost 0.1. This is rather strange, because it is very unlikely that vegetation cover will change so drastically over a time span of 12 hours. This underlines a limitation of the processor as used in this research. There should be some kind of possibility to limit the extend the optical depth can move compared to a value determined for the same area at an earlier date. For the further analysis in this chapter, it is less important, and the optical depth as determined by the processor is considered the correct one.

Table 4.3 shows the values for *SM* for the sensitivity analysis. On every row, the parameter that is changed and the value used is given. Under the different dates, the *SM* content that is derived for the L1c pixel is noted. Whenever the *SM* content falls outside the range for *SM* as set in the ATBD (0-60%), the pixel is rejected and the outcome is Not a Number (NaN).

Whenever there are four values next to a parameter, the first two are the values of the extreme analysis (denoted with ext.), the second two are the values of the location analysis (noted with loc.). The parameters b', b'', ω and c_L only have three values, because their default value is at the same time a value used in one of the analyses. For b'', ω and c_L their default value of 0 is the lowest value possible, thus one of the values for the extreme analysis. For b', its default value of 0.3 is the highest value of the location analysis. For μ_s, no extreme values were known. Only a location analysis was executed. For R_{tt} only one non-default value was found in literature, leaving only one value in Table 4.3. a_L, b_L and B_s_L show five values. This is because the optical depth τ, used for the determination of *SM* of Table 4.2, does not include litter (the default value for c_L is 0). The fifth value is the *SM* value with litter included and is denoted with def.. c_L Has in this case a default value of 0.24. Instead of comparing the outcome of the two analyses with the value from Table 4.2, it should be compared to this fifth value. Looking at the outcomes for c_L and a_L, b_L and B_s_L, one would expect the same values for the default value of the latter three and the location analysis with c_L = 0.24. This is not the case, because the sensitivity of is tested with the processor, while the other three parameters are tested with the code.

Table 4.3: *SM* content for different dates. The calculation is done with default parameters from Table 4.1, except for one parameter, which values are noted in the table in the row of the parameter. Values are (top to bottom) extreme low and high of the extreme analysis, low and high of the location analysis. Rejected pixels get NaN instead of a value. (m) Stands for morning temperature, (e) for evening temperatures, ext. stands for extreme analysis, loc. for location analysis

Parameter/analysis/value			Date/starting time December 14th 2010, 05:29 h SM [%]	April 6th 2010, 05:31 h SM [%]	May 20th 2010, 05:18 h SM [%]	May 20th 2010, 17:48 h SM [%]
T_g [K]	Ext.	230	NaN	NaN	NaN	NaN
		320	25.3	26.7	37.4	43.2
	Loc.	300.5 (m)/ 308.5 (e)	8.2	11.2	20.5	32.0
		303 (m)/ 311 (e)	10.3	13.1	22.6	34.4
T_c [K]	Ext.	230	NaN	NaN	3.7	NaN
		320	18.3	16.2	25.0	33.8
	Loc.	298 (m)/ 305 (e)	6.5	9.8	18.9	26.1
		300.5 (m)/ 308.5 (e)	7.7	10.4	19.6	27.8
a_L [kg/kg]	Def.	2.33	9.7	14.2	20.8	19.1
	Ext.	0	6.9	10.0	19.1	23.2
		10	1.9	1.6	2.0	1.7
	Loc.	2.097	9.2	13.5	23.7	21.5
		2.563	10.6	14.4	18.4	17.0
b_L [kg/kg]	Def.	0	9.7	14.2	20.8	19.1
	Ext.	1	NaN	NaN	NaN	NaN
	Loc.	0.1	14.5	14.6	17.2	16.0
B_s_L [kg/kg]	Def.	0.3	9.7	14.2	20.8	19.1
	Ext.	0	6.9	10.0	19.1	23.2
		50	NaN	NaN	NaN	NaN
	Loc.	0.27	9.3	13.7	21.8	20.0
		0.33	10.2	14.5	19.9	18.2
c_L [m²/kg]	Ext.	1	7.47	8.05	31.17	32.79
	Loc.	0.01	7.44	8.04	18.78	22.31
		0.24	7.45	8.04	18.78	22.31
b' [m²/m²]	Ext.	0.01	1.12	3.36	9.79	8.74
		1	7.47	8.05	31.88	32.80
	Loc.	0.1	7.44	8.04	18.78	22.31
b'' [m²/m²]	Ext.	3	NaN	NaN	NaN	NaN
	Loc.	0.03	7.44	8.04	18.78	22.31
		0.09	7.44	8.04	18.78	22.31
ω [-]	Ext.	0.5	NaN	NaN	6.78	5.00
	Loc.	0.15	1.32	4.17	14.91	15.40
		0.2	1.32	3.38	14.91	15.40
$ρ_s$ [g/cm³]	Ext.	2	7.63	8.23	18.99	22.52
		3	7.34	7.95	18.68	22.21
	Loc.	2.564	7.46	8.06	18.81	22.34
		2.764	7.41	8.01	18.75	22.28
$μ_s$ [H/m]	Loc.	1.00000017	7.44	8.03	18.78	22.31
		1.00000059	7.44	8.03	18.78	22.31
HR [-]	Ext.	0.05	6.92	7.48	18.20	21.70
		5	NaN	NaN	NaN	NaN
	Loc.	0.7	14.26	14.94	23.41	26.25
		1.3	18.61	19.70	25.52	27.61
R_{tt} [-]	Ext.	2.6	NaN	12.63	18.53	26.26

The outcomes for the extreme analysis seem to suggest that every parameter has a major influence on the outcome of the algorithm. Thankfully, the location analysis shows that for most parameters an estimation of the value should be enough to get an accurate determination of *SM*. All results are briefly mentioned in the following sub paragraphs. A discussion of the results follows in paragraph 4.4.

Looking at the results, the following should be kept in mind: the *SM* content of a SMOS pixel is not determined directly. Instead, the brightness temperature, measured by the satellite, is compared with a theoretically calculated brightness temperature, based on location specific parameters. For this sensitivity analysis, these location specific parameters, and thus the theoretical brightness temperature, are changed. The satellite image however stays the same for a certain date. This means that, when a parameter reduces the theoretical brightness temperature of a pixel, it depends on whether the theoretical brightness is higher or lower than the satellite measurements before and after the change of parameters whether the *SM* content increases or decreases. Besides that, as mentioned before, results cannot be simply generalized because of the non-linearity of the algorithm used. This is shown in the graphs below.
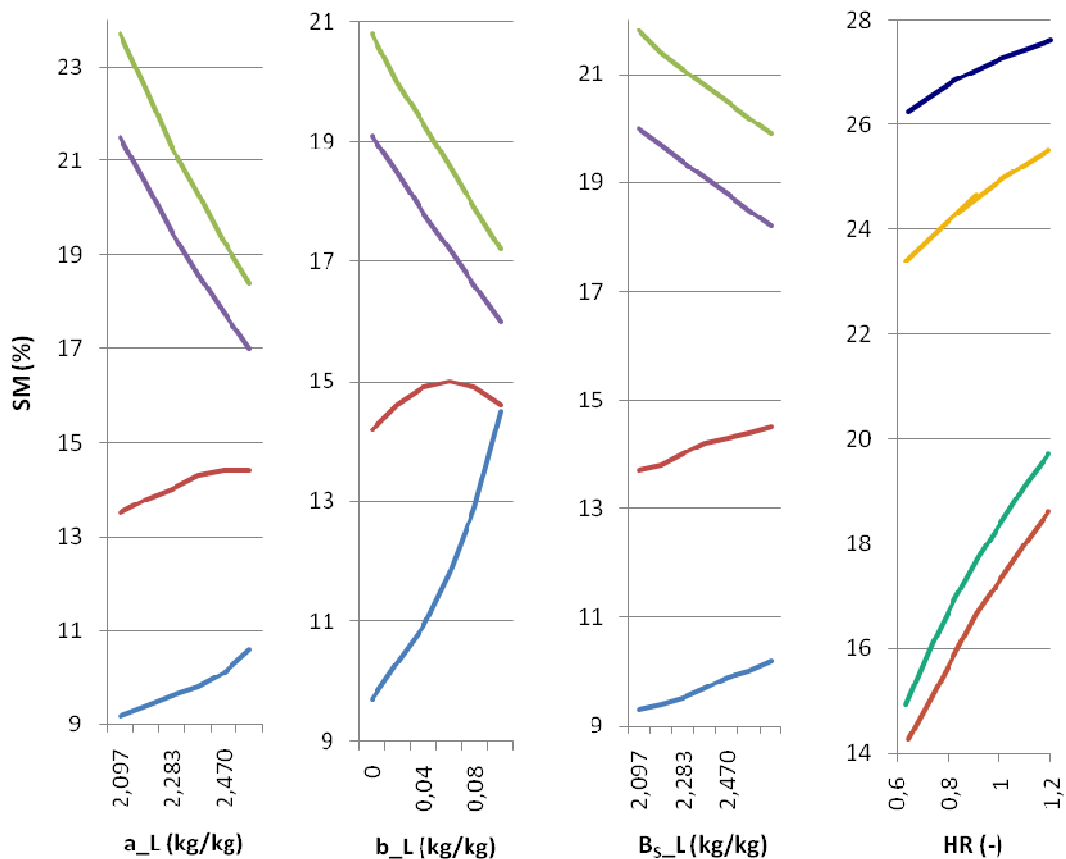
Figure 4.2a-d: Effect of respectively the parameters *a_L*, *b_L*, *B$_S$_L* and *HR* on soil moisture for the values falling within the location analysis. The red line stands for December 14th 2009, the green line for April 6th 2010, the orange line for May 20th 2010, morning overpass, and the blue line for May 20th 2010, evening overpass
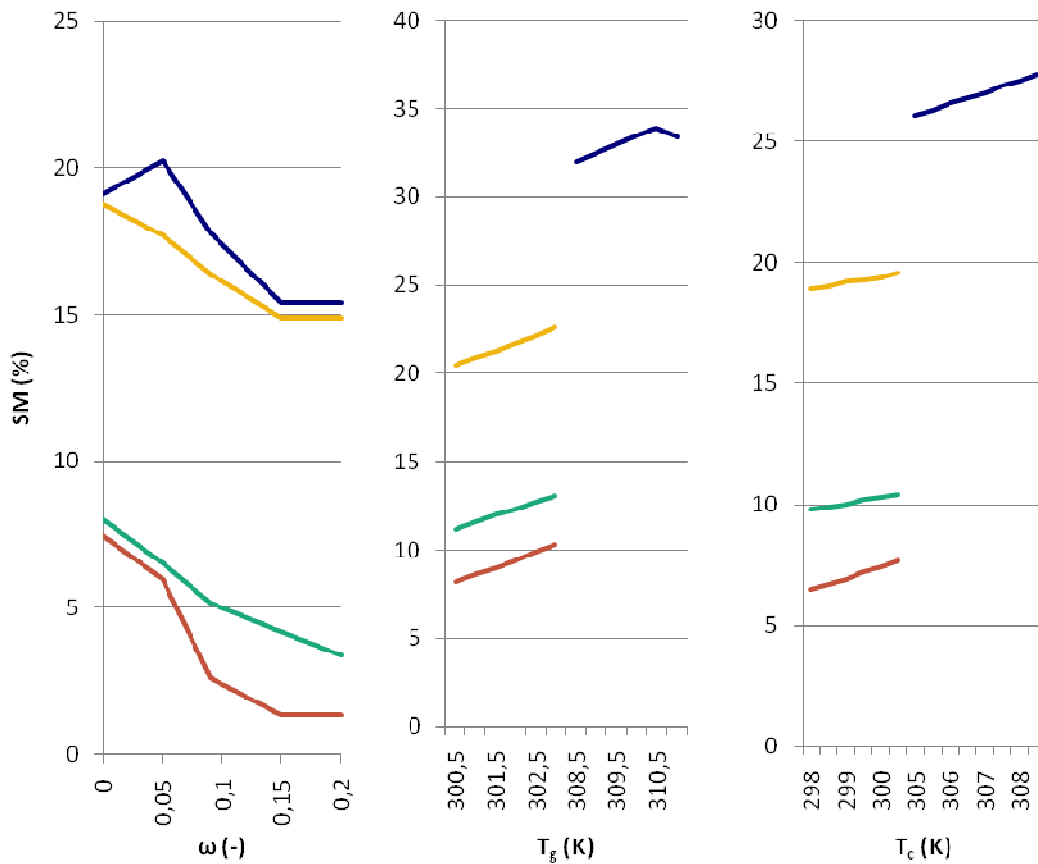
Figure 4.3: Effect of respectively the parameters ω, T_g and T_c on soil moisture for the values 0-0.2 (-) and 300.5-310.5 K and 298-308 K. The red line stands for December 14th 2009, the green line for April 6th 2010, the orange line for May 20th 2010, morning overpass, and the blue line for May 20th 2010, evening overpass

The figures clearly show the non linearity of the reaction of *SM* to the changes in parameters. Besides that, it also shows that the reaction depends on the original value of the parameter and the date, and thus the measured *TB*. It is thus not possible to generalize the results of this chapter to other areas.

4.3.1 CANOPY AND SOIL TEMPERATURE, T_C AND T_G

For this analysis, Table 4.3 shows that for higher temperatures, *SM* increases. The influence of the two temperatures is comparable. Extreme low temperatures lowers the *SM* content until it becomes negative, and is thus given NaN. Both temperatures show a change in *SM* of 1 to 2% for the location analysis. The influence of soil temperature is slightly larger than the influence of canopy temperature.

The reaction of *SM* to the values falling in between the two values of the location analysis show an almost linear shape. It does however depend on the measured *TB* and/or the original value

whether SM goes up or down with the change in parameter, as can be seen in the graph of May 20th, evening overpass for T_g.

4.3.2 PARAMETERS FOR DETERMINATION OF LITTER WATER CONTENT, A_L, B_L, B$_S$_L AND PARAMETER FOR DETERMINATION OF OPTICAL DEPTH LITTER AS FUNCTION OF LWC, C$_L$

The sensitivity analysis of a_L, b_L and B_s_L is done with the Matlab code, while that of c_L is done with the L2PP processor. Because for the Matlab calculation, the τ is taken from the Matlab code, the values in Table 4.3 show the added effect of litter to the SM content. The default value is therefore the same for a_L, b_L and B_s_L.

Comparing this default value with the one found in Table 4.2 shows already the difference litter can make. The addition changes SM content with 2 to 4%. There are some values in the extreme analyses that are exactly the same as the default from Table 4.3. Both $a\_L=0$ $B_s\_L=0$ change the SM values back to the values that are obtained without a litter contribution. Because of their value, Mg_L and/or LWC become(s) zero. The other part of the extreme analysis shows large differences with the default SM value. There is also a big difference between the outcome for the different parameters. Where in the extreme analyses $a\_L=10$ and $B_s\_L=50$ decrease the SM content, $b\_L=1$ increases SM (this does not follow from Table 4.3 for B_s_L and b_L, because the SM values calculated fall outside the 0-60% range). It depends on what the date is how much the in- or decrease of SM is. For $a\_L=10$ it can be seen that more or less the same value is reached for all dates.

The biggest change in SM for the location analysis by changing parameters comes from a_L and b_L (1-4%). B_s_L shows a smaller change with 1-2 %. For a_L and B_s_L, the first two dates show a lower SM value for the lower value of the location analysis than the default value with litter. The higher value of the location analysis shows a higher SM value. For May however, this is the other way around.

Changing c_L shows for its location analysis only minor changes for SM. Looking at the value for the optical thickness as determined with the processor, this has hardly changed compared to the value found in Table 4.2. This is a big difference with the Matlab code, that does show a big difference in optical thickness, because the litter contribution is added to the default optical depth. LWC in the processor must be very small, causing only a change in SM for large values of c_L. The extreme analysis shows the same difference in outcome for December and April compared to May. This has to do with the TB values found in May, that are much lower than the ones in December and April.

The graphs in Figure 4.2 show again that the actual influence of the chance of parameter depends on measured TB, original value of the parameter and the amount of change of the parameter.

4.3.3 PARAMETER FOR DETERMINATION OF OPTICAL DEPTH AS A FUNCTION OF LAI, B (B' AND B")

Decreasing the default value of b' changes the *SM* content in this analysis. Increasing the value however hardly changes the *SM* content. For b", the extreme analysis shows a very large impact on *SM*. The location analyses of the two parameters however show only a small difference in *SM* values with the default value.

4.3.4 SINGLE SCATTERING ALBEDO, Ω

Changing the scattering albedo changes the *SM* content greatly. Already a change from the default value to forest vegetation (0 to 0.1) gives a difference of almost 4%. Changing it to the value found for grass changes the value even more.

4.3.5 SOIL PARTICLE DENSITY, P_S AND MAGNETIC PERMEABILITY, M_S

The extreme range of 2-3 g/cm^3 for the soil particle density gives a maximum change in *SM* of 0.21%. This is less than 10% of the 4% range set for the SMOS satellite and thus not significant.

Changing the value of the magnetic permeability within the ranges to be expected in Ghana does not change *SM* at all.

4.3.6 EFFECTIVE SURFACE ROUGHNESS DIMENSIONLESS PARAMETER, HR AND PARAMETER FOR DETERMINATION OF INFLUENCE OF INCIDENCE ANGLE ON OPTICAL DEPTH, R_{TT}

A change in the roughness factor changes the outcome of the *SM* content drastically. The location analysis shows a range for *SM* of which minimum and maximum value can differ up to 4%. The extreme analysis shows that a small error in roughness value can already change *SM* with more than 0.5%. This large change however is not see in Figure 4.2d. This graphs also show that the change in *SM* is not linear with the change in *HR*. It does however seem to come close for the range of the location analysis.

Changing R_{tt} from 1 to 2.6 also shows that this can lead to a change in *SM* of up to 4%. Although very little is known about this value for an pixel sized area, it does show, just as the parameters *a_L*, *b_L* and *B$_S$_L*, the influence of a wrong value for the optical depth.

4.4 DISCUSSION

4.4.1 CANOPY AND SOIL TEMPERATURE, T_C AND T_G

Changing the temperature with three degrees Kelvin changes the temperature up to 2% in this analysis. It seems that this relationship is rather linear, although whether or not SM goes up or down depends on the measured TB as well. Two percent is half of the allowed range set by ESA. A good estimate of the temperature is thus necessary to determine SM correctly. On top of that, it is important to distinguish soil temperature from canopy temperature. Field measurements show a distinct difference between the two temperatures, that cannot be ignored according to the outcome of the sensitivity analysis.

4.4.2 PARAMETERS FOR DETERMINATION OF LITTER WATER CONTENT, A_L, B_L, B$_S$_L AND PARAMETER FOR DETERMINATION OF OPTICAL DEPTH LITTER AS FUNCTION OF LWC, C$_L$

Calculating with the nominal model, the addition of the litter component to the default calculation already gives a change in SM up to 4%. On top of that, a wrong estimate of the different parameters, in the range used in this analysis, can lead to a comparable change in SM. For West-Africa, where much is still unknown about litter, this can lead to large errors in the determination of SM. For a good estimate, more information on when litter can be found and which parameters would fit with the circumstances is necessary.

The fact that a_L=0 and B_S_L=0 change SM values back to the values that are obtained without a litter contribution can be explained by the fact that both values change Mg_L and/or LWC to zero. This in turn leads to a litter contribution of zero.

The outcome for the sensitivity analysis of c_L is not comparable to that of a_L, b_L and B_S_L. Its location analysis shows only small changes in litter. The optical depth for the location analysis and the extreme analysis does hardly change compared to the default values found in Table 4.2. The LWC value in the processor is apparently so small, that c_L only has an effect when it is very large. Possibly, there is a mitigating effect of other areas outside the pixel that is not taken into account in the Matlab code, but is in the processor. For May in the extreme analysis however, the SM values do change. These files have far lower values for TB. Apparently, the minimization process here uses a larger value for the litter in order to get a better fit of the calculated TB with the measured value. Because in this analysis, the optical depth changes, this is really the only place where you can see the impact of litter on SM. It also underlines an apparent weakness of the processor as used in this research.

The outcome of a_L=10 and b_L=1 is opposite. This is probably because b_L has a direct influence on Mg_L, while a_L's influence is also dependent on SM. Where b_L=1 always causes Mg_L to be 0.8, a_L=10 only causes Mg_L to be 0.8 when SM >=0.08. Since the value of SM

changes with every iteration, so does the litter contribution, and thus the final outcome of the minimization.

The fact that for some dates, SM increases with a certain location analysis value, while for others it decreases, has to be accounted to the fact that SM is determined indirectly by comparing two TBs.

4.4.3 PARAMETER FOR DETERMINATION OF OPTICAL DEPTH AS A FUNCTION OF LAI, B (B' AND B")

The difference in reaction to higher or lower values of b' and b" can be explained by the fact that the optical depth occurs in the SMOS algorithm as an exponential function. Because the location analysis parameters show higher values than the default value, this range has very little impact on the outcome of the optimization. A exact determination of the parameters is thus unnecessary, as long as this range represents the area well.

The difference in effect for extreme values of b' and b" can again be attributed to the fact that b" influence on the optical depth is direct, while that of b' also depends on LAI, just as the influence of a_L depends on SM.

4.4.4 SINGLE SCATTERING ALBEDO, Ω

The sensitivity analysis shows that SM can change drastically by changing the single scattering albedo. During the field work, the area showed both forest areas and areas covered with grass, and not solely with grassland, as the default value used in the L2PP processor suggest. On top of that, the vegetation also changes over time with the wet and dry season. Since forest have a default value for ω of 0.1 and grass of 0, it seems that in order to get the right value for SM in this area, the scattering albedo will have to be determined more accurately. It will depend on the final range of the area how accurate exactly the parameter has to be determined. Because of the change in vegetation in the dry and the wet season, it is could well be that ω changes over time.

4.4.5 SOIL PARTICLE DENSITY, P_S, AND MAGNETIC PERMEABILITY, M_S

The analysis shows that the sensitivity of the algorithm for soil particle density and magnetic permeability is small to non-existent (assuming the location analysis values are a good representation of the characteristics of West-Africa). It is therefore not necessary to determine their values exact. An estimation will already lead to a minimal SM error.

4.4.6 EFFECTIVE SURFACE ROUGHNESS DIMENSIONLESS PARAMETER, HR

The difference in SM for HR=0.7 and 1.3 already shows a change in SM value. If these values do indeed represent West-Africa better than the default value of 0.1, it is necessary to determine a good estimate for HR in order to determine SM within the 4% error bounds. Changing HR from from 0.1 to 0.05 already gives a change larger than 0.5%. This seems to suggest that it is

important to determine HR within 0.05 and thus that it is important to determine the roughness correctly.

4.4.7 PARAMETER FOR DETERMINATION OF INFLUENCE OF INCIDENCE ANGLE ON OPTICAL DEPTH, R_{TT}

Changing R_{tt} from 1 to 2.6 shows big differences between the change in SM content for the different dates. Possibly, this has to do with the horizontal and vertical value of the optical depth. Just like with the b' and b" parameters, not every change in R_{tt} might change the optical depth in such a way that there is actually a (large) change in SM content. The fact that there is a change however does warn for assuming the default value of 1 without hesitation.

4.5 CONCLUSIONS

Looking at the different parameters, it seems that the most important parameters are the scattering albedo, the roughness of the area, temperature and litter properties. Unfortunately, at least three of these parameters (ω, HR and τ) are hard to determine for a inhomogeneous area. The biggest changes in SM come forth from changing just these parameters from their default value to one of the location analysis values. Regarding the location analysis itself, changes in SM are largest for HR. It should be kept in mind that the exact sensitivity of SM to the parameters within the range of the location analysis is not always linear.

The influence of the erroneous value for the scattering albedo can be far larger than the permitted range of 4%. The problem with this parameter is however that, although it can be measured over a small homogeneous area, the effect of different kinds of vegetation on its value is not well known. Perhaps in the future, it might be possible to obtain the correct value by calibration. Assuming that all other parameters can be determined correctly, SM content measured in the area in for example a field work should be compared to the theoretical value, and the albedo should be changed until the two values are equal. Another way would be to find a way to measure the albedo over a SMOS node with remote sensing.

Roughness values found on experimental sites differ vastly from the default value of the research area. Determining the roughness of an inhomogeneous area as it is done on small fields is however a very tedious job. The ATBD however gives an equation that uses the field capacity of an area in the calculation. This seems a good method to get an average value, once the formula is validated and calibrated.

A wrong estimation of the temperature of an area can influence the SM content, but the change depends on the accuracy of the temperature's estimate. The ranges measured during the field work show changes of approximately 2% between maximum and minimum temperature (with a difference of approximately 3 K). The different effect of canopy and soil temperature (for the

same temperature change) show that it is important to get a good (separate) estimate for the two parameters. ECMWF data shows however that this is possible.

Litter properties are hard to determine, especially when the litter layer changes over time, which is expected in northern Ghana, where in the dry period areas are often burned. Its effect on *SM* can be large however. More research and local measurements should be done in the field to get a workable, calibrated equation. Perhaps it is possible to obtain a distinction between different areas that show different litter layers. However, since the final optical depth is often also determined through minimization, when using the processor, one should be aware of the fact that minimization is a process that can still produce erroneous values, if boundaries are not set correctly.

Validation of SMOS satellite data over Ghana and Burkina Faso

5 VALIDATION THROUGH FIELD WORK

5.1 INTRODUCTION

As mentioned before, the ESA has set an error bound of +/- 4% for SMOS *SM* data. A first check to see whether this is indeed the case is described in this chapter.

In May 2010, a small field work was executed in northern Ghana. The objective was to arrive at an average *SM* value of a SMOS pixel (approximately 15 by 15 km). The location of the field work is also used in the previous chapter for the sensitivity analysis. The field work was executed at the end of the dry season. The comparison should give an answer to sub question 3:

3. How do SMOS satellite measurements compare to field measurements for the pixel with its centre located at 1°07'54" W, lat 9°11'44" N?

The biggest challenge of the field work was to obtain an average *SM* value that is representative of the entire area. This has been done with the so called landscape-unit approach. This approach divides the area into different classes (also called units) , based on characteristics. Theoretically, landscape units should have a smaller variance than the difference between their samples. Differences between units are however more distinct in wet areas than in dry areas. According to the Central Limit Theorem, when the number of sample is large enough, the distribution of *SM* within a unit can be assumed to be normal. This makes it possible to determine an average *SM* value for the unit by determining the overall sample mean. With that the average value of the complete pixel can be determined and compared to the value measured by the SMOS satellite.

In order to determine the sample locations, originally, the hydrotope method developed by Friesen *et al.* (2008) [5.1] was used. This method divides the area into units with a consistent hydrologic behavior. For this particular method the units are called hydrotopes. Some hydrotopes did not show a normal distribution for the top five centimeter of the soil. Therefore, other landscape methods were tried as well. The measurement were divided in units according to what type of vegetation was growing on the location, the soil type and whether or not the location was shaded.

Samples were taken with the gravimetric method. The errors are not quantified. If the measurement errors are large however, the 4% error bound for SMOS might have to be adapted in order to incorporate this measurement error. Different landscape unit methods lead to different average values of an area. This difference can give an indication whether the error bound does indeed need adaptation.

The following paragraph gives a more detailed explanation on the different landscape units used in this research. It also explains how the ground samples were processed in order to determine

their *SM* content. Paragraph 5.3 gives the results of the fieldwork. These results are discussed in paragraph 5.4. The research question is answered in paragraph 5.5.

5.2 LANDSCAPE METHOD

In order to get a well defined average of the *SM* content, enough samples need to be taken in order to obtain a distribution that approaches the normal distribution. This can lead to several hundred samples [5.2]. A way to limit the variability of a group of samples, and thus the amount of samples necessary, is using a landscape-unit approach. The following sub paragraph explains the hydrotope method, the theory behind it and the division of sample locations according to the hydrotopes defined for the research area. Paragraph 5.2.2 explains the other landscape-units methods used in this research.

5.2.1 HYDROTOPES

One of the landscape-unit approaches is the hydrotope method, developed by Friesen *et al.* (2008) [5.1]. The area that is being researched is divided into different hydrotopes. A hydrotope is defined as 'a unit characterized by the dominance of similar hydrologic processes'[11]. They are therefore different in different areas. A hydrotopes shows a temporal pattern that is distinguishable from other hydrotopes in the area. Comparing two hydrotopes should always show a difference in their mean larger than their variability. The landscape-unit approach comes from pedological communities, were several studies have shown that characteristics seems to be organized in different land types [5.2]. The same goes for *SM* in wet periods [5.3]. *SM* then shows a high degree of correlation with terrain parameters. It should be noted however that this correlation is far less in dry periods.

The amount of samples that should be taken, can be calculated with the help of the Central Limit Theorem and preferably data from previous research. According to the Central Limit Theorem, if enough samples are taken, normality can be assumed. The 'true' value for the mean and standard deviation are than approximately the same as the ones calculated with the samples. Using this theorem, it is then possible to calculate the amount of samples needed to obtain a representative mean value within a certain confidence level. The equation used for this calculation is:

---

[11] Thus not the same as fractions used in the algorithm for determining the soil moisture by the satellite

$$\bar{M} \sim \bar{m} + \frac{s}{\sqrt{n}} \cdot \bar{\chi}$$ (eq.5.1)

With:

$\bar{M}$ = overall mean value

$\bar{m}$ = mean value of a sample (e.g. of a previous research)

s = standard deviation of a sample (e.g. of a previous research)

n = numbers of samples taken

$\bar{\chi}$ = standard normally distributed variable, gives the confidence interval of the mean value

In order to get a representative average over the complete area that is being researched, the amount of samples in a certain hydrotope should be made proportional to the area it covers.

5.2.1.1   HYDROTOPES FOR FIELD WORK

For this field work, sample locations were chosen with the classification for the Volta basin as used in Friesen *et al.* (plateau, slope and wetlands), extended with the hydrotope forest (on a plateau). Plateaus are large upland areas, which are characterized by deep soils. Precipitation is stored in the root zone. This unit mainly contributes to runoff via subsurface flow. Wetlands saturate very early during the wet season and contribute almost solely to the direct runoff or Dunne flow. Slopes connect the plateaus to the wetlands. Both Hortonian surface runoff and shallow-groundwater flow occur on the slope unit. The hydrologic state of the slopes tends to vary relatively quickly over time and space. Forest on plateaus have all the characteristics of plateaus, but have forest as the main vegetation cover. This hydrotope was added because of the big influence of forest on the SMOS algorithm and the relative large area covered with forest (according to satellite images). The hydrotopes in the area were classified by comparing slopes and the upstream area, and for wetlands and forest by looking at land cover satellite images. Figure 5.1a-d give an impression of the different hydrotopes' vegetation.
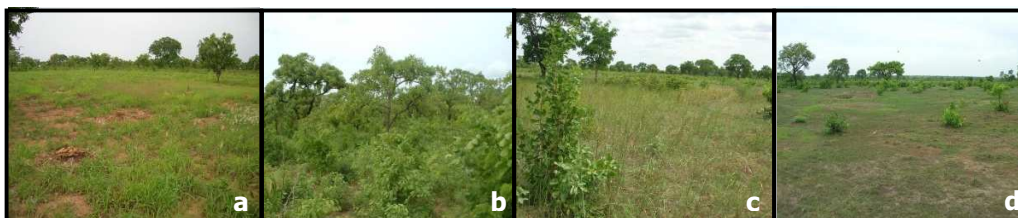
Figure 5.1a-d: Impression of vegetation types of different hydrotopes, a=plateau; b=forest; c=wetland; d=slope

Figure 5.2 and Figure 5.3 show the area chosen for the field work and the location of the sample locations, with their hydrotope-classification. In total, 130 locations were marked to take samples. From the 130 locations, 60 were classified as plateau, 40 as forest, 20 as slopes and 10 as wetlands (see Figure 5.3). On every location (but two, that could not be reached), samples were taken from the top soil (0-5 cm). 89 Locations were also sampled at a depth of 5-10 cm and 21 at a depth of 10-15 cm. This was done in order to be able to examine the influence of these deeper soil layers on the signal the satellite receives. Practical considerations, like time

constraints and tough ground were reasons not to sample every location at three different depths.
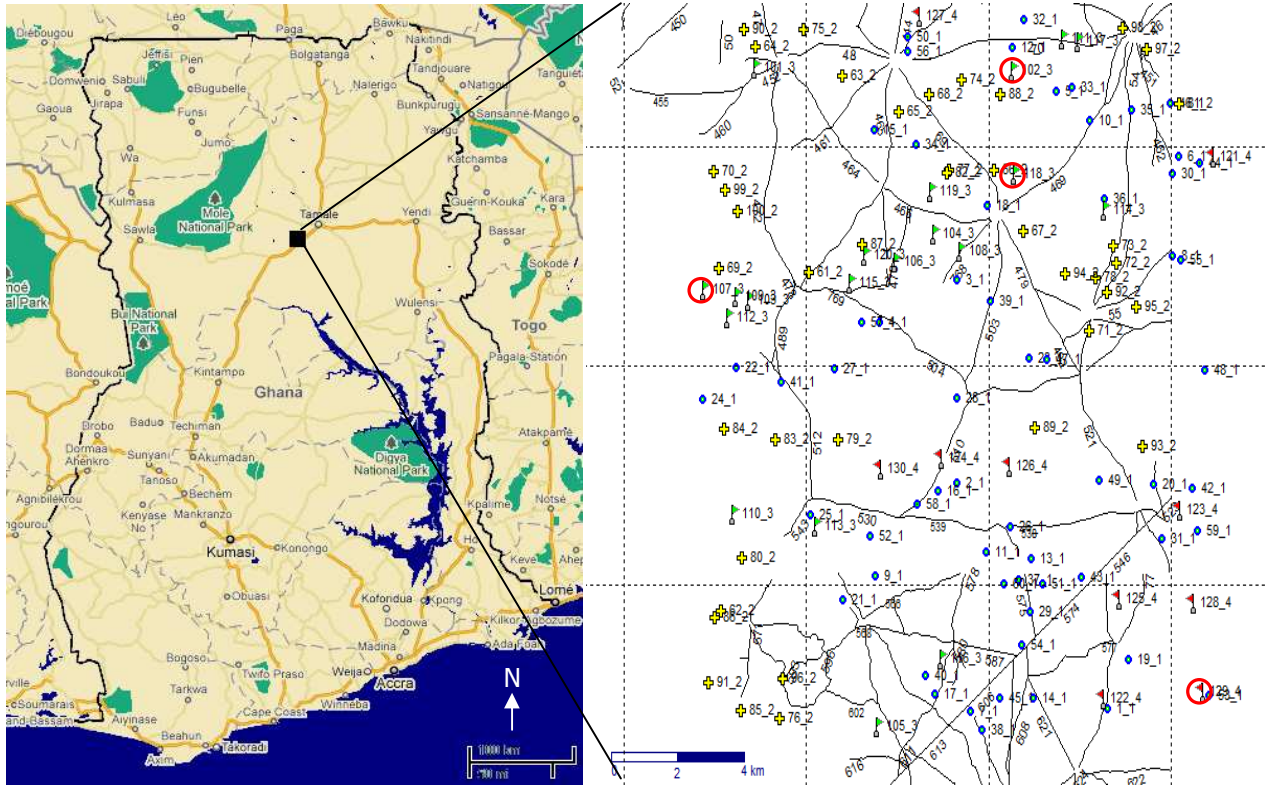
Figure 5.2 (left): Ghana and measurement location (http://maps.google.com/, 31-08-2010)

Figure 5.3 (right): Measurement locations; blue cirkel=plateau, yellow cross=forest, green flag=wetland, red flag=slope

5.2.2 DIVISION BY OTHER LANDSCAPE-UNIT METHODS

SM in the top soil is affected directly by the environment and shows a less nominal distribution than deeper layers [5.4]. To get a normal distribution for this layer, it is necessary to take more local differences in account than can be achieved with hydrotopes. Therefore, for the top soil, several other methods were tried based on local information, namely soil type, shade and vegetation. These methods divide the area in different units, comparable to hydrotopes, but based on different characteristics.

Information on the characteristics of the measuring locations was obtained from different sources. Soil types were obtained from previous research, shade and vegetation classification was done with the help of photos made at the sample locations. 33 Locations were not photographed. This left very little data for deeper layers. These have only been classified with the hydrotope method. Classifications for the different methods can be found in Table 5.1.

74

The new classifications do not have an amount of measurements per unit based on its area. To get a statistical representative average, the mean value of the complete area should therefore be determined by weighing the means of the different units with the inverse of the standard deviation of the area. The latter gives an indication on how well the mean of a unit represent the SM value of the complete area. By averaging the weighed mean values of the units, a representative mean value of the complete area can be obtained.

Table 5.1: Different landscape-unit methods for top soil and their units

Method	units
Hydrotopes	Plateau; forest; wetland; slope
Soil types	Acrisols; planosols; lixisols
Shade	Yes; no
Vegetation	Small patches of dry grass; grass 2-10 cm; grass higher than 10 cm; grass and bushes higher till 50 cm; bushes higher than 50 cm/forest

5.2.3 PROCESSING THE SAMPLES

The gravimetric method is the only direct method to determine *SM* content [5.5] and is therefore the only method that can be used to calibrate other sampling methods [5.6]. Sources of error can follow from: the sort of instrument used in the sampling and the use of the instrument, particularly when it comes to compaction of the ground when one needs to determine bulk density or when there are rocks in the soil; the sample containers; the balance; the oven; the drying time; the cooling of the samples after the drying period [5.6]. For the sampling, metal rings were used with a fixed volume. Small rocks and roots are not pushed out of the samples, ensuring a realistic sample. In the area, no large rocks or other obstacles were found at the sample locations. Samples were put in two plastic bags and were weighed in the field, as soon as possible. If thought necessary these measurements were checked after the field work, because of the sensitivity of the balance to wind and skew surfaces. The accuracy of the balance was 0.1 g. In the laboratory, the samples were weighed once more before drying, to ensure no large mistakes were made. Most of the samples weighed between 100 and 200 grams. In order to dry the samples completely, they were dried in the oven for 24 hours at 105°C. For the drying phase, the samples had to be moved to metal cups. These cups were weighed, including and excluding the soil sample, as were the empty bags. This way, the error of the transfer could be calculated and its effect on the *SM* content. Because of the amount of samples that were processed, the samples had to cool off in the open air before weighing. The samples could thus have taken water in after drying, which causes an unknown error in the determination of the *SM* content.

5.3 RESULTS

5.3.1 RESULTS FIELD WORK

Mean, standard deviation, range and number of samples for bulk density (g/cm^3) and volumetric *SM* content (%) for the different hydrotopes are noted in Table 5.2. Of the 130 selected locations two (one forest location, one plateau) were not sampled, because of difficulty reaching them. Labeling was done on the basis of satellite imagery and 6 locations were labeled incorrectly. For those locations, a second sampling was done nearby, on the right hydrotope. Wrongly labeled locations were added to the right hydrotope dataset. 89 Locations were also sampled at a depth of 5-10 cm, 21 of these also on a depth of 10-15 cm. Bulk density and *SM* content are calculated taken into account the error created by moving samples from plastic bag to metal cup. The loss of soil was calculated in percentage and deducted from the volume in the ring. This changed the bulk density with a maximum of 0.32%. Apart from one sample, this leads to an error in the *SM* no larger than 0.37%.

Measurements were done from May 10th to May 14th. During this period, no rain fell on the measurement area. On Wednesday May 5th, an intensive rainstorm hit Tamale. It is not sure whether this also hit the measurement area.

SM values show a big range. This can be attributed to four distinct sampling locations, namely location 53, 102, 107 and 118 (see Figure 5.3). Any surrounding locations near these outliers show values in the range from 0-10% (0-6% for the top layer). The three highest values were found in the wetland hydrotope, on peaty soil. The fourth value was found on loamy ground on a plateau.

The first characteristic of hydrotopes is that, in the wet period, there is a clear distinction between the *SM* values between them. This distinction becomes less clear in the dry period. The division of the data according to the different methods soil type, vegetation, shade and hydrotopes is shown in Figure 5.4. The thin lines in the graph show the complete range of the unit's data. The rectangles show the mean plus and minus one standard deviation for each unit. The 'b' behind the name of a unit stands for a calculation done without the extreme values from location 53, 102, 107 and 118. Units with an 'a' are calculated with all data available.

Figure 5.4 shows the division of the data according to the different methods soil type, vegetation, shade and hydrotopes. The thin lines in the graph show the complete range of the unit's data. The rectangles show the mean plus and minus one standard deviation for each unit. The 'b' behind the name of a unit stands for a calculation done without the extreme values from location 53, 102, 107 and 118. Units with an 'a' are calculated with all data available.

All methods' *SM* contents show a big overlap, especially when locations 53, 102, 107 and 118 are taken into account. E.g. for the hydrotope method, average values for the top layer range from 2.54% to 3.96%. The Smallest standard deviation of the layer is 1.49%, larger than the biggest difference between the average values. The differences in mean between a and b calculations

range from 0.11% to 1.77%, with an average of 0.71%. The difference in standard deviations range from 0.26% to 2.88% with an average of 1.45%. Big differences in the mean *SM* value between hydrotopes are usually combined with large standard deviations. This means any distinction that could have been made on the basis of the mean is lost because of the large standard deviation.

Looking at the differences between units without taking into account the location 53, 102, 107 and 118, shows that the biggest differences in mean between units can be found in the hydrotope method. At the same time, the hydrotopes' average standard deviation is the largest of all methods. For every method, the landscape units' average standard deviations is at least almost twice as large as their average mean, again prohibiting a clear distinction between units.
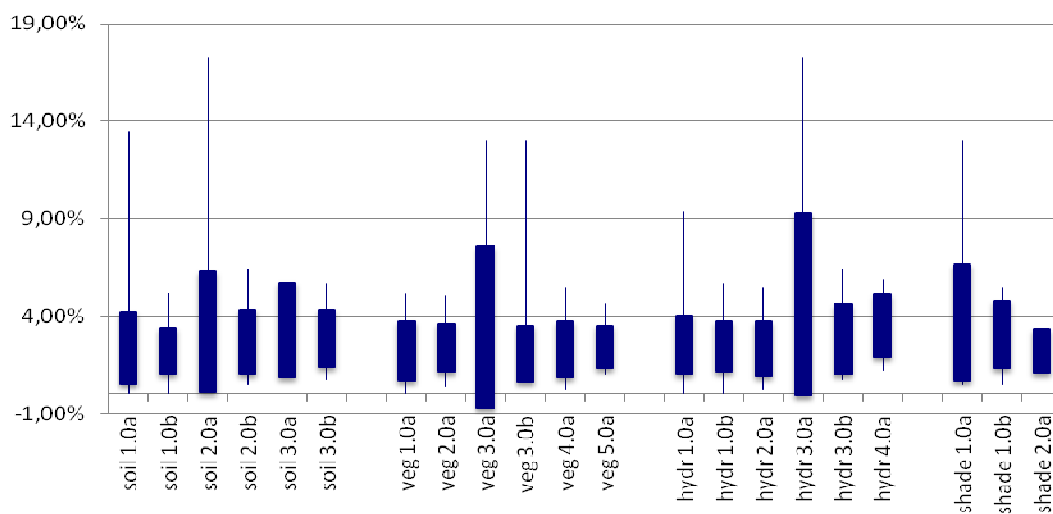
Legend

Soil, veg, hydr, shade = Different methods used;

1.0-5.0 = Landscape units within methods:

      Soil: 1) acrisols; 2) planosols; 3) lixisols

      Vegetation: 1) barren/small patches of dry grass; 2) grass from 2-10 cm; 3) grass higher than 10 cm; 4) grass and bushes higher till 50 cm; 5) bushes higher than 50 cm/forest

      Hydrotopes: 1) plateau; 2) forest; 3) wetland; 4) slope

      Shade: 1) with; 2) without

a, b = Calculation made respectively with and without 4 highest values

Figure 5.4: *SM* minimum, maximum, mean - standard deviation, mean + standard deviation for different landscape unit methods at a depth of 0-5 cm

Figure 5.5 is a histogram of the complete dataset. This histogram shows a distribution that seems to resembles a lognormal distribution, rather than a normal distribution. The only values that fall outside this distribution are the four mentioned above. This clear distribution, together with the overlap shown in Figure 5.4, support the findings of Friesen *et al.* that hydrotopes show less distinction in the dry period. Despite this, it is important to realize that the landscape-unit method is an essential tool to determine a correct average *SM* value for an area in wet periods. This is the reason this method was used in this research.
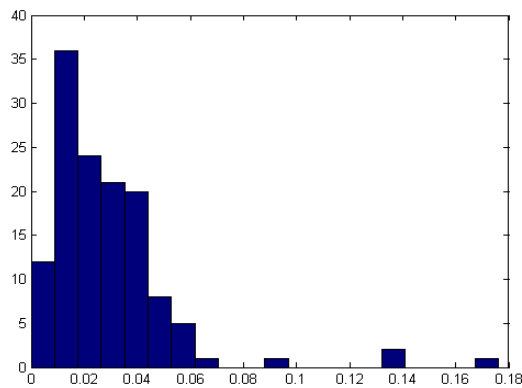
Figure 5.5: Histogram of *SM* data for top soil layer (0-5 cm depth)

The fact that the complete dataset shows a lognormal distribution instead of a normal distribution, can be explained by the fact that the *SM* values are low, but can never become negative. This is not necessarily remarkable, but can have complications. It suggests that instead of a normal distribution, the data of the different hydrotopes might also shows a lognormal distribution. This means that taking the sample mean of the different hydrotopes and averaging that might not give a representative average *SM* value after all. Histograms of the hydrotopes (See Figure 5.6 for an example) however do not give a clear indication that a lognormal distribution describes the values better than a normal distribution. Therefore, in the following, it is still assumed that well defined hydrotopes show a normal distribution for *SM*.

Visually, it is not possible to say in how far the hydrotopes approach a normal distribution. Therefore, the normal distribution hypothesis of the different landscape-unit methods was checked with the chi-square goodness-of-fit method. For the hydrotope and vegetation method, these values can be found in respectively Table 5.3 and Table 5.4. The chi-square goodness-of-fit method works as follows: the probability function (=the probability of the occurrence of any value within a normal distribution with a specific mean and standard variation) is calculated with the mean and standard deviation of a specific dataset. This is then compared to the actual occurrence of values within the data set. In how far the data set is normally distributed is calculated by determining the root mean square difference between the two. Whenever this difference is larger than 5%, the hypothesis that the data set is distributed normally is rejected.

For the hydrotope method, most hydrotopes do indeed show a normal distribution (at all depths). There are however large deviations, especially in the top soil, and the hypothesis is rejected for the plateau and forest hydrotopes in this layer. This means that for these hydrotopes, the sample mean cannot be used as a representative value. It thus seems that, at the very least in the dry period, other characteristics than those used for the classification of hydrotopes have a part in determining the distribution of *SM*.

In order to examine this, other landscape-units were defined. The objective was to see whether it would be possible to get a better approach of the normal distribution than with the hydrotope division. As mentioned above, units are defined based on soil type, vegetation type and shadiness. Of these method, only the vegetation method shows a normal distribution for all units (in the top soil). It does need an extra unit however in comparison to the hydrotope method.

Combining vegetation between 10 and 50 cm, thus creating 4 units, would already lead to a unit that is not normal distributed. Although the vegetation method seems an improved classification for the top soil, it should be noted that 33 locations were not photographed and therefore not classified. The number of samples for the different units can be found in Table 5.2.

Both for the hydrotope method and the vegetation method (although less distinct), plotting data against the normal distribution of the top soil show for almost every unit the same pattern. The probability of very high and very low values of *SM* in reality is lower than for a normal distribution. For deeper layers, there is a higher probability for high *SM* values than the normal distribution would justify. For the forest hydrotope at this depth however, the data shows a sort of wave-movement around the normal distribution (see Figure 5.6a). This is also the case for the plateau hydrotope at a depth of 5-10 cm. The data is bimodal. This is illustrated by the histograph of the data (Figure 5.6b) that shows two peaks.
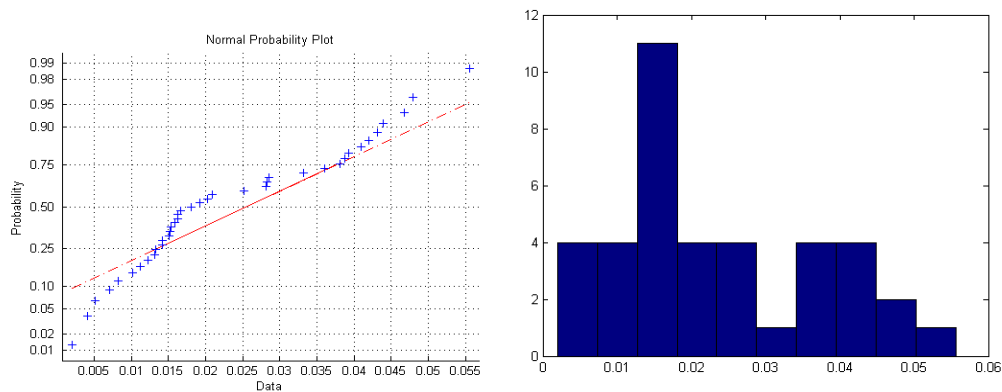
Figure 5.6a-b: Respectively data of forest plateau 0-5 cm plotted against normal distribution and histogram data forest plateau 0-5 cm

The unit that shows the biggest difference in mean with other units of the same method is the slope hydrotope. Its standard deviation is in every case at least 0.37% bigger than the difference in mean between hydrotopes.

Table 5.3 and Table 5.4 shows the values for the chi-square goodness of fit for respectively the hydrotope method and the vegetation method. To see what the influence is of a extra fifth class within the distribution, for the vegetation method the same calculation was done for a combination of class 3 (grass higher than 10 cm) and 4 (grass and bushes higher till 50 cm) together. Vegetation units show a better fit with the normal distribution than the hydrotope method, except for the slope hydrotope. Extreme low vegetation (short grass) and extreme high vegetation (forest and bushes) show the best fit to a normal distribution. Combining grass higher than 10 cm and bushes lower than 50 cm clearly gives a less normal distribution, but its value is still lower than for the plateau and forest hydrotopes.

Table 5.2: Mean, standard deviation, minimum, maximum and number of samples of volumetric soil moisture and soil bulk density over different hydrotopes and depths

Calculation over			Mean	Standard deviation	minimum	maximum	Number of samples
Complete area	SM	0-5 cm	2.84%	2.37%	0.00%	17.27%	131
		5-10 cm	5.46%	3.19%	0.92%	23.06%	89
		10-15 cm	6.59%	2.78%	2.16%	10.40%	21
		0-15 cm	4.96%	2.85%	0.70%	10.40%	20
	ρ_b(g/cm^3)	0-5 cm	1.44	0.18	0.77	1.98	131
		5-10 cm	1.50	0.18	0.88	1.96	89
		10-15 cm	1.52	0.21	1.17	2.03	21
		0-15 cm	1.48	0.18	1.11	2.03	20
Plateau	SM	0-5 cm	2.54%	1.49%	0.00%	9.33%	63
		5-10 cm	4.84%	2.21%	1.52%	8.83%	36
		10-15 cm	6.63%	2.63%	2.16%	10.25%	10
		0-15 cm	5.12%	2.80%	1.08%	10.25%	9
	ρ_b(g/cm^3)	0-5 cm	1.44	0.18	1.03	1.98	63
		5-10 cm	1.50	0.19	1.23	1.89	36
		10-15 cm	1.46	0.17	1.17	1.70	10
		0-15 cm	1.42	0.16	1.11	1.77	9
Forest	SM	0-5 cm	2.64%	2.79%	0.20%	17.27%	49
		5-10 cm	5.37%	3.82%	0.92%	23.06%	32
		10-15 cm	5.81%	2.53%	3.09%	8.96%	5
		0-15 cm	4.27%	2.60%	1.39%	9.02%	5
	ρ_b(g/cm^3)	0-5 cm	1.45	0.16	1.12	1.82	49
		5-10 cm	1.49	0.17	1.15	1.96	32
		10-15 cm	1.50	0.22	1.30	1.77	5
		0-15 cm	1.46	0.16	1.19	1.77	5
Wetland	SM	0-5 cm	3.96%	3.68%	0.70%	13.46%	20
		5-10 cm	6.68%	3.97%	1.43%	13.80%	15
		10-15 cm	2.60%	0.63%	2.16%	3.05%	2
		0-15 cm	3.01%	3.05%	0.70%	9.02%	2
	ρ_b(g/cm^3)	0-5 cm	1.37	0.20	0.77	1.72	20
		5-10 cm	1.45	0.20	0.88	1.61	15
		10-15 cm	1.42	0.07	1.38	1.47	2
		0-15 cm	1.45	0.07	1.38	1.56	2
Slope	SM	0-5 cm	3.75%	3.10%	0.70%	13.46%	9
		5-10 cm	7.50%	1.69%	5.16%	9.90%	6
		10-15 cm	9.27%	0.84%	8.57%	10.40%	4
		0-15 cm	6.24%	2.93%	1.41%	10.40%	4
	ρ_b(g/cm^3)	0-5 cm	1.56	0.20	1.29	1.95	9
		5-10 cm	1.64	0.12	1.47	1.78	6
		10-15 cm	1.78	0.17	1.66	2.03	4
		0-15 cm	1.65	0.16	1.41	2.03	4

Table 5.3: Chi square goodness-of-fit for different hydrotopes and depths[12]

Data over		χ^2 (count)/ hypothesis rejected y/n
0-5 cm	Complete area	9.3649 / y
	Plateau	7.0513 / y
	Forest	6.9155 / y
	Wetland	3.1667 / n
	Slope	0.1331 / n
5-10 cm	Complete area	3.7018 / n
	Plateau	1.2364 / n
	Forest	0.0090 / n
	Wetland	0.6258 / n
	Slope	0.0019 / n
10-15 cm	Complete area	0.4647 / n
	Plateau	0.1455 / n
	Forest	0.0659 / n
	Slope	0.0038 / n

Table 5.4: Chi square goodness-of-fit for vegetation method, top soil (0-5 cm) and amount of samples of different units[13]

Data over (top soil)	χ^2 (count)/ hypothesis rejected y/n	Number of samples
1) Barren/small patches of dry grass	0.6867 / n	12
2) Grass 2-10 cm	0.0234 / n	29
3) Grass higher than 10 cm	2.2470 / n	9
4) Grass and bushes higher till 50 cm	1.0501 / n	23
5) Bushes higher than 50 cm/forest	0.5805 / n	30
3) + 4)[1]	6.5762 / y	32

5.3.2 SATELLITE DATA

At the moment of writing, brightness temperature data was available from May 17th to May 23rd. However, the last field measurements were done on May 14th. It did not rain in the field work area until the evening of May 19th. It is assumed that *SM* content does not change significantly from May 17th to May 19th.

Data was processed with the L2PP processor developed by Array. Auxiliary files used were mostly compiled for a test data set of February 25th 2007. The Weather data (temperature and pressure) temperature and used here came from the European Centre for Medium term Weather Forecast

---

[12] The hydrotope wetland at a depth of 10-15 cm has only 2 measurements and is therefore left out

[13] 3) + 4) stands for a combination of units 3) and 4) of the vegetation method (thus gras higher than 10 cm and grass and bushes till 50 cm)

(ECWMF) and was composed for October 21st 2009. This was the only available file for this research that covered the node where the field work was executed.

Whenever a node's *SM* is determined to be negative, it gets assigned NaN. Unfortunately, from May 17th to May 19th, this happens for the measured node while using default values for the processor as mentioned in chapter 4. The only date that does give any information on *SM* on the area is May 20th. This overpass happens hours after a heavy rain storm hit the area. The *SM* values found on this day are thus far higher than the ones found in the field (11 and 12.6% with default values for parameters).

In order to get a value to compare with the field measurements, the HR value for the area in the processor was adapted. Changing this value from 0.1 to a value that is found in literature, namely 0.7, does give a value for *SM*, namely 0.2%. Although that falls within the +/- 4% margins set by ESA, it is more than 15 times as low.

5.4 DISCUSSION

The data found in the field work are compared to that of Friesen *et al.* (2008) [5.1] and Algyare (2004) [5.7]. Both did research in areas near Tamale, comparable to the area research in this paper. Friesen *et al.* found for *SM* on a depth between 15 and 20 cm a higher mean value (10%), but a similar standard deviation (3%). Measurements were done between May 26 and June 5. The different average values can be explained by a difference in precipitation in the preceding period, but this has not been investigated. Bulk densities concur with values found by Agyare (2004) [5.7] to the north of Tamale, which has the same soil types as the measurement area of this paper. Whenever the hydrotopes show a different bulk density, the standard deviation is always smaller than this difference. For *SM* however, comparing any two hydrotopes shows that their standard deviation is almost always bigger than the difference between their means. This makes a distinction between hydrotopes much harder. This correspond to findings of Friesen *et al.* (2008) [5.1], who finds a much clearer distinction between hydrotopes for *SM* in wet conditions than in dry.

The top soil shows a distinct pattern in comparison to the normal distribution. Very high and very low *SM* values show a lower probability than the normal distribution. This can be explained with the dry season evaporating the water in the areas with originally high moisture contents. At the same time dry soil prohibits the evaporation of the last water, which would cause extreme low moisture contents [5.4]. The high probability for higher *SM* values at deeper layers might be explained by the fact that water that cannot evaporate from the top soil can percolate to deeper layers. The effect is however less for deeper layers, causing a better fit with the normal distribution. For the top soil layer however, more localized factors that can cause differences in evaporation need to be taken into account in order to arrive at a normal distribution. The bimodality in the measured data of the plateau hydrotope (0-5 cm) and the forest hydrotope (5-10 cm), suggests that there is the need for an extra hydrotope. The better fit for the vegetation

hydrotope seems to suggest the same, and that vegetation should be taken into account for a distinction of landscape units for the top soil.

For a correct determination of the average value of an area, future field campaigns should thus focus more on the kind of vegetation and its influence on the area's *SM*. It is however recommended to use the hydrotope method in combination with vegetation, because the hydrotope method has proven its worth in wet periods, when *SM* is more linked to hydrotopes. Furthermore, the determination of vegetation-units as done in this reseach, using photos, is time-consuming and not possible to do beforehand. The hydrotope method on the other hand can be done relatively quickly for large areas by looking at satellite images. Research should be done whether it is possible to define vegetation from remote sensing imagery as well. Also, more research should be done into what kind of vegetation should be distinguished, based on their influence on *SM*. For a better determination of *SM*, a next field campaign should also focus on obtaining representative values for parameters instead of default values as found in the ATBD.

There are several causes why (average) *SM* values differ for field measurements and satellite data. Errors in the SMOS algorithm are an important part of this difference. Another part of the difference comes forth from the field work. First of all, it depends on the landscape-unit chosen what the average value for the area is. This shows the bias of the mean measured in the field compared to the 'real' average value, caused by the non-normal distribution. On top of that, there are measurement inaccuracy, explained in paragraph 5.2.3. The SMOS satellite is designed to provide global maps of *SM* content with an accuracy of +/-4 %. This range has been established assuming that the satellite measures only the top five centimeters of the soil. An error of 4% is accurate enough to give a good estimation of evaporation and soil transfer parameters [5.4]. The 4% error is linked to the SMOS algorithm. The question remains whether the error from the field work can create a difference between the field average and SMOS satellite that is larger than 4%.

Not all errors from the field work are qualified. What has been qualified are the processing errors as described in the beginning of paragraph 5.3.15.2.3. This error was 0.37% on average, which is less than 10% of the 4% accuracy set by ESA. This is thus considered negligible. This leaves inaccuracies due to the chosen landscape-unit method. In order to find whether the bias of the measurements justifies increasing the inaccuracy range, their theoretical probability distribution was determined. The *SM*s mean and standard deviations of the hydrotope and vegetation method, and that of all samples together, were determined for the different depths. With this the probability distributions for the different methods were calculated. These were then plotted in one graph. Next, the average value of the three different methods and the SMOS accuracy range around that average [average-4%, average+4%] was drawn into the graphs, comparing a theoretical 'real' mean value of *SM* and the satellites accuracy range with the different distributions, see Figure 5.7a-c. When over 95% of the distribution of one of the methods falls within these boundaries, it is expected that a method gives a representative average value of the area and it is not necessary to increase the 4% range set by ESA.

Figure 5.7 shows that the difference between the means of the three distributions is smaller than 1% for the top soil layer and up until 1.3% for deeper layers. The difference between the standard deviations is smaller than 0.5% for the two highest layer. It is maximal 1.12% for the

deepest layer. For the top soil, the comparison shows that the distribution according to the vegetation method falls for more than 95% inside the range of the satellite. The average of the area will therefore definitely fall in that range as well. The other distributions also fit relatively well within the satellites range. On a depth of 10-15 cm, the hydrotope method seems to fit best in the satellites range. The other two distributions do not seem to fit as well. On a depth of 5-10 cm none of the distributions fit very well in the range of the satellite. In order to take the bias of the measurements into account, the accuracy for the SMOS satellite has been adjusted for the two deeper layers, around the 'real' average value. This is only necessary when the SMOS satellite would actually measure deeper than five centimeter. This is theoretically only the case in very dry areas [5.8].The accuracy of the satellite for the deeper layers is taken as two times the standard deviation of the all-method, which for this layer has the broadest range of values within a 95 probability. The different values can be found in Table 5.5.

For the above analysis, normal distributions are used, following the assumptions on hydrotopes that their *SM* value distribution is normal. For the top soil, one can see that according to the normal distributions, 10% of all measurements should be negative. This is obviously not possible. The comparison of the data to the normal distribution already showed that for the top soil, extremely low and extremely high *SM* values are less likely to occur than in a normal distribution. This goes for both the hydrotope method and the vegetation method. It therefore seems likely that instead of negative *SM* values, more data would be found within the accuracy bounds. Creating a histogram of the data and comparing it to the normal distribution that fits the data best shows that there are far more locations with a *SM* content between 1 and 2 % than would be expected from the normal distribution determined with the mean and standard deviations of the dataset (Figure 5.8). It is therefore expected that, even when the data would in fact be lognormal distributed, the 4% error bounds can be maintained. Assuming that the SMOS satellite only measures *SM* in the top five centimeters of the research area, this means that the satellite data determined with the L2PP processor should fall within 4% of the value measured in the field.
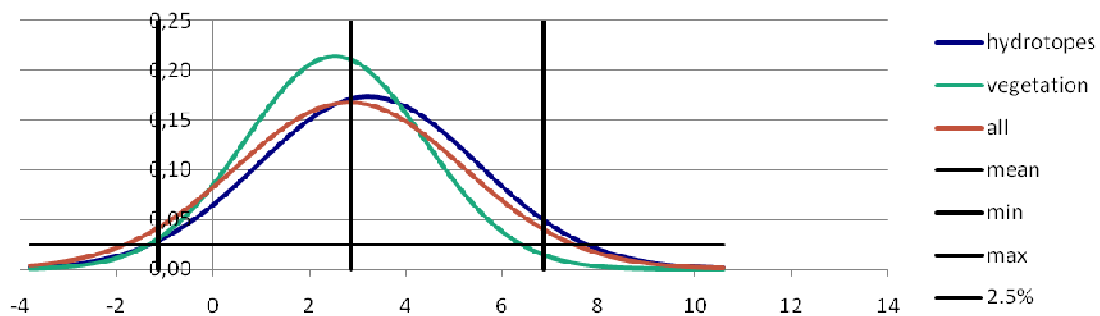
Figure 5.7a: Normal distribution for a depth of 0-5 cm
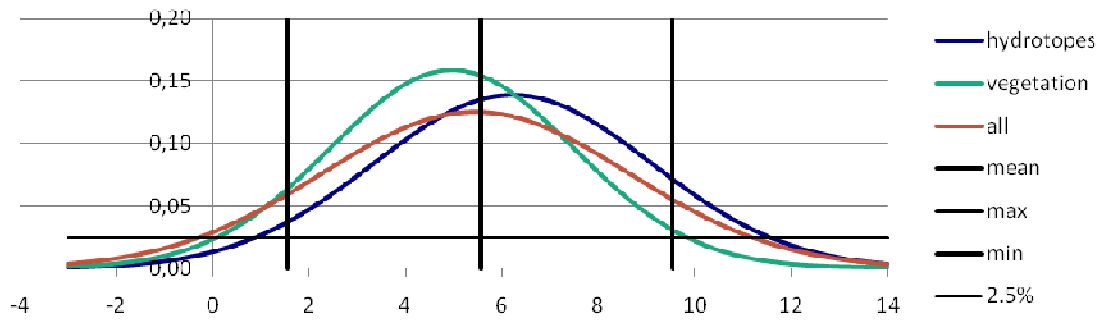
Figure 5.6b: Normal distribution for a depth of 5-10 cm
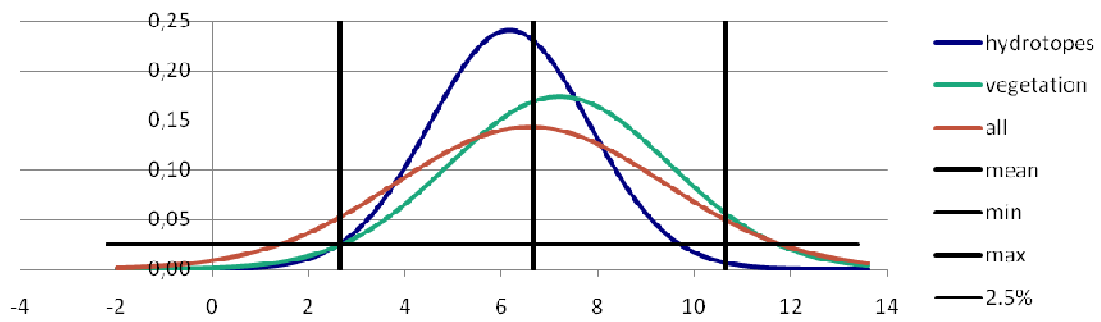
Figure 5.6c: Normal distributions for a depth of 10-15 cm
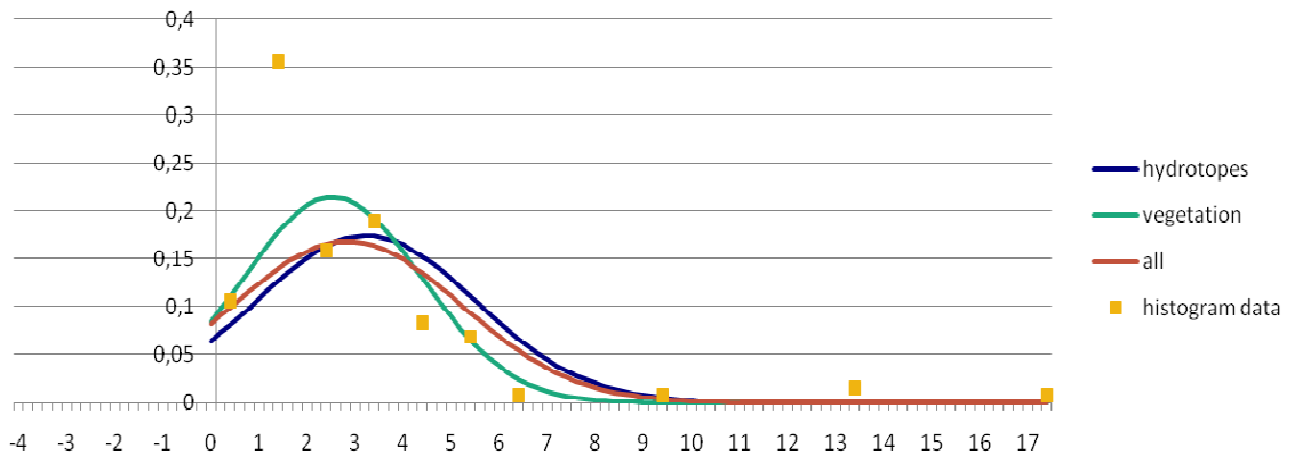
Figure 5.8: Real data top soil compared to theoretical distribution, purple squares are located in between whole *SM* values, giving the number of samples found in this range as a percentage of the total number of samples

Table 5.5: Used mean and accuracy satellite for comparison with field measurements

Layer	Method	Mean/standard deviation	
0-5 cm	Average = average means 3 different methods; Standard deviation = accuracy satellite	2.87%	4%
5-10 cm	Average = average means 3 different methods Standard deviation = two times standard deviation all-method	5.55%	6,38%
10-15 cm	Average = average means 3 different methods Standard deviation = two times standard deviation all-method	6.65%	5.56%

The *SM* as calculated by the L2PP processor with default values for parameters (see Table 4.1) is far lower than the average value determined in the field. In fact, it becomes negative, causing the *SM* content to become NaN. However, the values of the pixels to the north of the measured area 0.01 and 0.07, do fall in the accuracy bound as determined above (2.84% – 4%). Seeing that the surroundings of the research area shows such low *SM* values, and deeper layers show higher *SM* values than the top soil layer, it does not seem the satellite measures any more than the first 5 cm. Changing the roughness value from 0.1 to 0.7 namely 0.2%. This also falls within the +/- 4% margins set by ESA, it is more than 10 times as low.

5.5 CONCLUSIONS

The mean value over the complete research area for the different landscape-unit methods and the one calculated over all samples show a very small difference. Looking only at the outcome of this research, it thus seems unnecessary to use any landscape-unit for the determination of a representative average *SM* value during a field work. However, the field work in this research was executed at the end of the dry season. At that point, there is hardly any water in the soil, limiting the difference in *SM* value of the different units. When the soil is wet however, it depends on local characteristics what happens to the *SM*, and far larger differences can be found. This larger range of *SM* values leads to a very large number of measurements necessary to determine a representative average value of the area. This number can be reduced by defining landscape-units with their specific (smaller) standard deviations. The smaller the standard deviation, the better the mean value represent the complete pixel. For a dryer area, it is possible to get a mean that represents the area statistically better by weighing the means of the units with the inverse of their standard deviation. Landscape units are thus a very important tool to keep a field work feasible (money- and time wise) and to get an representative mean value.

For this research, 130 sample locations were chosen based on the hydrotope method. The landscape-unit that should be used however, depends on the depth one is interested in. For the top soil, important for the comparison of the area with the SMOS satellite data, dividing the area on the basis of vegetation gives a better fit with normal distributions than the hydrotope method as used for this area. For deeper layers the hydrotope method does show a good fit with normal

distributions. In this research, the differences in average value of the different methods were however small.

For this paper, deeper layers showed a relatively large range and large standard deviations. This leads to a bigger uncertainty on the bias of the measurements. The SMOS satellite can determine *SM* for the top 5 cm of the soil. It is suggested in the ATBD that for very dry areas, the satellite can measure deeper soil layers. In that case, it might be necessary to allow the satellite data to fall in a bigger range around the real *SM* content than the 4% set by ESA[14]. It should be noted that for the lower layers, far less measurements were available. This might be the cause for the bigger uncertainty.

Comparing field measurement with satellite data furthermore seems to suggest that the satellite only measures *SM* of the top soil. The SMOS data gives an extremely low value for the research area. Deeper soil layers show a higher average *SM* value than the top layer. Assuming the parameters in the processor represent the area accurately (except for the roughness parameter), it seems the satellite only measures the top layer.

Although the SMOS *SM* value is extremely low, it does fall within the 4% margins set by ESA. This might partly be the case because the field work was executed at the end of the dry season. *SM* content over the node was therefore relatively homogeneous. The SMOS value is however not very realistic, being more than 15 times smaller than the field measurements suggest. For more realistic values however, default values, set for the entire earth, will need to be adapted to the specific area. A future field work should focus on the influence of vegetation and possibly other local differences, on this top layer and adapt the hydrotope method to include these local differences in advance.

---

[14] This range is only set for the top layer

Validation of SMOS satellite data over Ghana and Burkina Faso

6 VALIDATION THROUGH COMPARISON WITH OTHER DATASETS

6.1 INTRODUCTION

In the previous chapter, the SMOS *SM* data is compared with field measurements. This is a single comparison at only one location, at a certain time. Another way to validate the SMOS data is to compare it to other field and satellite data that has been measured over a certain period of time. This makes it possible to look at the SMOS data over a longer time period and, in the case of satellite data, over a larger area. In addition, the different *SM* data can be compared to precipitation data, which gives insight into why the *SM* content changes.

Through these comparisons, the chapter tries to give an answer to sub questions 4 and 5:

4. How do SMOS satellite measurements compare to field measurements of the École Polytechnique Fédérale de Lausanne in Burkina Faso?
5. How do SMOS satellite measurements compare to other satellite data?

There are several datasets that could be used in order to validate the SMOS product. Rudiger *et al.* (2007) have compared three different satellite datasets, a model and field measurements with each other over France. The objective was to validate data over a larger area, other than with field measurements. This research used satellite data from the Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E) and from the ESA Remote Sensing Satellite with the Scatterometer ERS-Scat. There are two algorithms developed to determine *SM* from AMSR-E measurements, namely that of the National Snow and Ice Data Centre (NSIDC) and that of the Vrije Universiteit Amsterdam (VUA) in combination with NASA. [6.1]. Their research showed a lack of dynamics for AMSR-E NSIDC data set and a wet bias for AMSR-E VUA-NASA and a dry bias for ERS-Scat, while both are well correlated. Winter showed a drop in correlation however. ERS-Scat shows a slightly better fit for *SM* in more forested areas than AMSR-E VUA-NASA. The absolute error found for ERS-Scat was 5.6%.

With an indication of the validity of the ERS and AMSR-E algorithms, it is easier to determine the validity of SMOS when comparing it to the first two. It was therefore decided to use a similar approach for validating SMOS data as Rudiger *et al.* (2007) used. In this chapter, SMOS data is compared to AMSR-E and ERS (in this case measurements from the so called advanced scatterometer, Ascat) data and field measurements in Burkina Faso. Because there was no model available that determines *SM* over the entire area, both satellite and field measurements are compared to precipitation. The field measurements in Burkina Faso come from École Polytechnique Fédérale de Lausanne (EPFL, Switserland). These are done at a depth of five centimeter. Precipitation data used comes from the Tropical Rainfall Measurement Mission (TRMM, by the Japan Aerospace Exploration Agency JAXA and NASA). Data was available for only a short period of time, namely July 11th to August 5th.

The results from Rudiger *et al.* give an indication on how SMOS data should perform when it does indeed represent *SM* within a 4% range. Based on this, some hypotheses are formulated in order to evaluate the comparisons of SMOS with the different datasets. Because the EPFL

measurements have not been designed in order to get a well defined average value for the pixel they are in, it is not expected that SMOS data will show comparable absolute values. Correlation for this pixel with normalized values should however be high. Assuming then that EPFL measurements show a high correlation with TRMM data, so should the SMOS data. Since the results from Rudiger *et al.* showed a low correlation for AMSR-E NSIDC data with the field measurements, while AMSR-E (VUA-NASA) and ERS-Scat showed a high correlation, it is expected that SMOS shows the highest correlations for comparisons with the latter two datasets. Possibly, correlation in more forested areas might be higher for the comparisons with ERS-Ascat than with AMSR-E VUA-NASA.

In the following paragraphs, the validation is described. Paragraph 6.2 starts with the research area, the data availability, an explanation on the different datasets used, how they are adapted for the comparisons, an explanation on the statistical properties determined and the hypotheses formulated on the comparisons. Paragraph 6.3 gives the results of the comparisons. In paragraph 6.4, these results are discussed. Paragraph 6.5 finally tests the hypotheses as formulated in paragraph 6.2.5 and answers sub research questions 4 and 5.

6.2 METHOD

In paragraph 6.2.1 the research area for which satellite data was available is shown. Data availability is also dependent on the satellites' orbit. This is explained in paragraph 6.2.1. In paragraph 6.2.2, a detailed description of the different datasets is given. It also mentions for what period data was available for the validation. The changes to the data (resolution), needed to compare the datasets with each other, are explained in paragraph 6.2.3. Paragraph 6.2.4 mentions the different statistical properties used in the validation. Paragraph 6.2.5 finally formulates the hypotheses used for the final conclusions on the validation.

6.2.1 RESEARCH AREA

The area that is being looked at covers Ghana and Burkina Faso completely. The area contains approximately 6340 SMOS pixels. Comparisons were done for the period July 11th to August 5th 2010. Figure 6.1 shows the different pixels in the area.
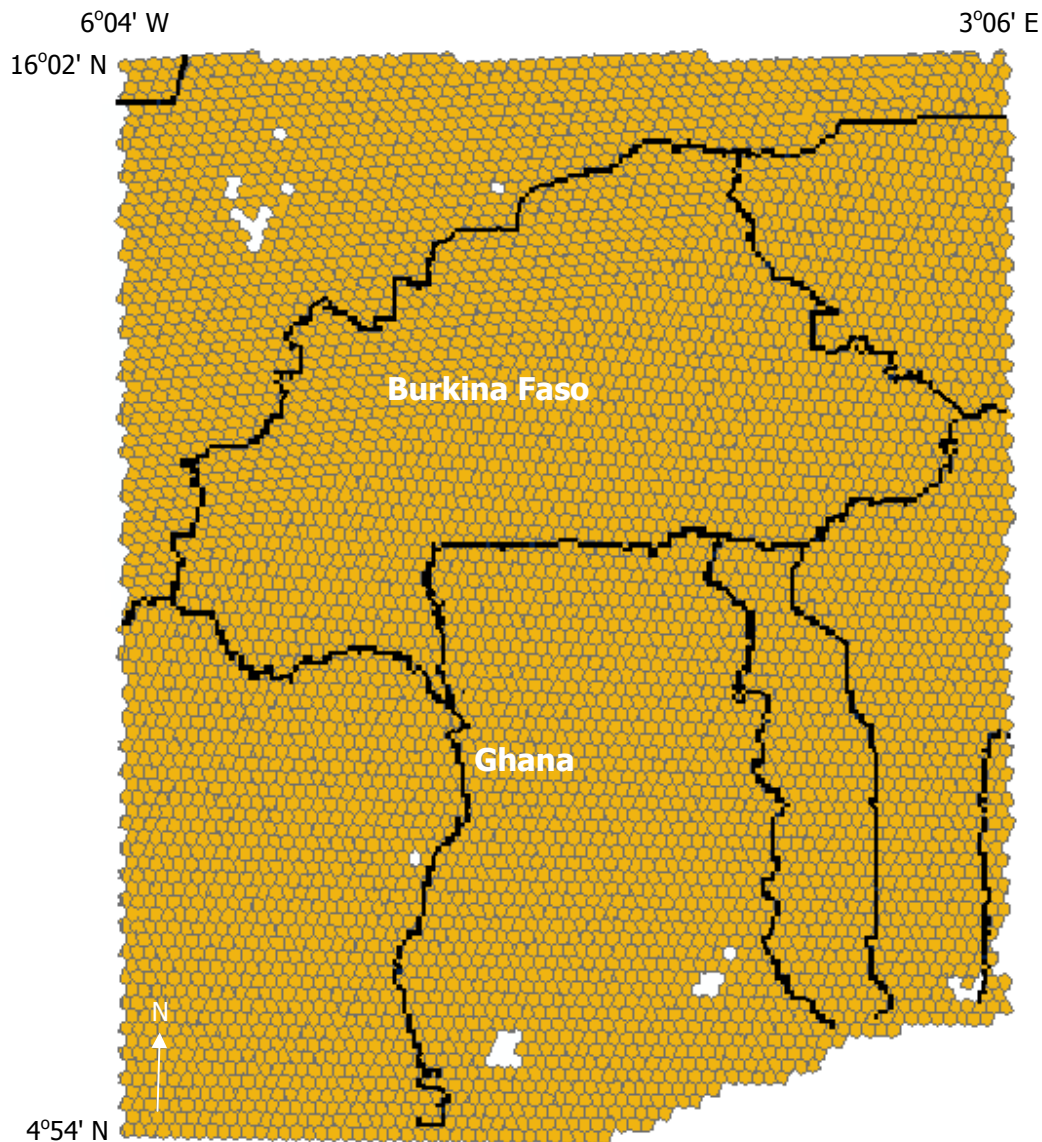
Figure 6.1: Pixels over West-Africa. The coordinates of the upper left corner are lat 16°02' N, lon 6°04' W, the coordinates of the lower right corner are lat 4°54' N, lon 3°6' E.

Because of their orbit, a satellite will cover, on a specific day, only part of the area described above. A comparison between datasets can only be made when both two satellites cover the same pixel on the same day. The number of times a pixel is measured by the SMOS satellite and (at least) one other satellites is called the count.

For the comparisons of the different satellite datasets with each other, the highest value for counts range from 6 to 10. For pixels with the highest count, graphs were created with SMOS data, precipitation data and data from one of the other satellites. The same graphs were also created with EPFL measurements. This in order to compare both *SM* values and patterns of the

different datasets. Several statistical properties were calculated for all pixels, in order to back up the visual analysis of the data. These coefficients become more relevant with a higher count. Their relevance is discussed by looking at the chance of their value occurring with random numbers. *SM* data from the different datasets is given in different unities. In order to compare them, the datasets were normalized. The different coefficients are therefore all dimensionless.

6.2.2 DATASETS

SMOS data is compared to four different datasets. In the comparison of the three satellite datasets with SMOS, it is possible to have a look at the complete research area. The EPFL data on the other hand covers only one node, restricting the comparisons of the satellite data with these field measurements to that area.

6.2.2.1 MEASUREMENTS IN BURKINA FASO

As part of the Info4Dourou project[15], EPFL has placed several *SM* meters in Burkina Faso, amongst others near Madjoari. Eleven of these meters measure the *SM* content at a depth of 5 centimeter. In the period of interest for this research, one of the meters was malfunctioning. The other 10 are used to determine the *SM* content for the SMOS pixel that covers the area in which the meters are located.

The meters can be found in the area between $1^{o}12'58"$ E, $11^{o}27'28"$N and $1^{o}13'24"$ E, $11^{o}26'30"$N (pixel no. 2254, see Figure E.2 appendix E). They form a line of approximately 4 km at the north west of the SMOS node on that location. The meters are close to a stream and therefore show high values compared to the other datasets[16]. Most of the meters are at a height of approximately 225 m. Three meters are at a different altitude, namely 214 m, 309 m and 318 m. These differences in heights do not show in the *SM* content.

To compare the meters with the satellite data, the daily average was taken over all ten meters. Data was available from July 11th to August 5th 2010. This data will be noted in the following as EPFL.

6.2.2.2 AMSR-E

The AMSR-E satellite has a morning (descending) and evening (ascending) overpass. There are two different datasets derived from the AMSR-E satellite, NSIDC and VUA. The difference lies in the overpass and the algorithms used to determine *SM*. *SM* is determined for approximately the first centimeter of the soil. Both datasets have a resolution of 0.25 by 0.25 decimal degrees.

---

[15] http://cooperation.epfl.ch/page-9064-en.html, viewed on 26-10-2010

[16] It should be noted that while the EPFL meters measure the soil moisture content of a single point, the satellites measure soil moisture over an area of minimal 15 by 15 km.

NSIDC *SM* data is determined with X-band (10.65 GHz) and Ku-band (18.7 GHz). The algorithm uses two low frequency dual polarized channels to optimize *SM*, vegetation optical depth and effective soil temperature simultaneously. The input *TB* has a spatial resolution of 56 km. The method uses two low frequency dual polarized channels to optimize *SM*, vegetation optical depth and effective soil temperature simultaneous [6.3].

Data has been derived from a combination of the descending and ascending overpass of the satellite and was available from July 11th to August 5th 2010. AMSR-E NSIDC data is in the following noted as NSIDC.

VUA data is derived according to the Land Surface Parameter Model (LPRM) (Owe *et al.* 2007). The model uses dual polarized channel (either 6.925 or 10.65 GHz) and retrieves both *SM* and vegetation water content. The model is based on a microwave radiative transfer model which has *TB* as input [6.4].

Data comes from the descending overpass and was available from July 11th to July 30th. AMSR-E VUA data is in the following noted as VUA.

6.2.2.3   ERS-ASCAT

The resolution of this dataset is 25-50 km. The satellite measures actively; it sends a signal and measures its backscatter, from which it determines *SM*. The method is called a change detection method. Instantaneous measurements are extrapolated to a reference incidence angle. These values are then compared to dry and wet backscatter references. Vegetation effects on the relationship between backscatter and incidence angle are removed, which gives *SM* variations. The historically lowest and highest values are assigned 0% and 100%, for the rest backscatter is considered to have a linear relationship to backscatter [6.5].

Ascat measures to a depth of <2 cm. Data is sometimes coming from an overpass in the morning, the evening or a combination of both. It was available July 11th to August 5th. ERS-ASCAT data is in the following noted as ERS.

6.2.2.4   TRMM

TRMM data has a resolution of 0.25 by 0.25 decimal degrees. Precipitation is determined through infrared measurements for every three hours.

The first step in determining precipitation is calibrating and combining microwave precipitation estimates. Then, infrared precipitation estimates are created using these estimates. The two estimates are then combined and rescaled to monthly data [6.6].

TRMM data was available from July 11th to August 5th. In the following, it is simply noted as TRMM.

6.2.3 ADAPTING DATASETS

As mentioned before, in order to compare the different datasets with each other, the units have to be adapted. This can be done by normalizing the datasets. The description above however indicates that there is another issue that requires attention: the datasets' various resolutions. Because of this, the cells overlap each other at points. More than one pixel of a different dataset can cover a SMOS pixel. In order to compare the two datasets, the resolution of the second dataset has to be changed to the SMOS resolution. How this is done is explained in paragraph 6.2.2.1. Paragraph 6.2.2.2 explains the normalization of the data.

6.2.3.1   RESOLUTION DATASETS

In order to compare the different datasets with each other, the datasets are adapted using linear interpolation. The data is put on top of each other in ArcGIS. This program shows the different datasets as a map. Initially, there are two maps with different resolutions. The resolution of the second dataset is changed to that of the SMOS dataset. This is done by joining it with the SMOS map on the basis of spatial location. Before that is possible, the resolution of the dataset with the largest resolution (never SMOS) is given a smaller resolution. This does not have an effect on the values, because the datasets cells are square, just as the new resolution. Bigger squares are simple divided into smaller squares. The smaller squares are also smaller than the SMOS resolution. By taking the average value of these smaller squares that fall within the SMOS pixel, a well defined value can be determined. This thus results in a map with the resolution of the SMOS dataset, but the values of the second dataset. The values in this map and the SMOS map can then be compared to each other.

6.2.3.2   NORMALIZATION DATA

In order to compare the different *SM* datasets, the data was normalized. For every pixel, the mean and standard deviation over time is determined and the normalized *SM* content is determined with is

$$\theta = \frac{(SM - \bar{m})}{\sigma}$$

(eq. 6.1)

6.2.4 PROPERTIES

There are many coefficients that can be used to analyze the correlation between different datasets. The ones used in this report are the Root Mean Square Difference, the Mean Difference, the Pearson and the Nash coefficient. These are the same as used by Rudiger *et al.* (2007). Rudiger *et al.* compare different *SM* datasets, amongst those the satellite data (except for the SMOS data) used in this report. The statistics are completed with the correlation between the different *SM* datasets and TRMM.

When more than one overpass per day is available, *SM* values can differ per overpass due to for example different temperatures. This can have an influence on the statistical coefficients. In

order to determine this influence, the coefficients are also done with only morning SMOS data. The objective is to see whether this gives better values for the statistical properties. Since NSIDC data is derived from a combination of morning and evening data, this comparison was not re-evaluated.

6.2.4.1   ROOT MEAN SQUARE DIFFERENCE

The root mean square error (RMSD) as used here is the difference between two datasets. The formula used for the calculation is:

$$RMSD(\theta_1, \theta_2) = \sqrt{\frac{\sum_{i=1}^{n}(x_{1,i} - x_{2,i})^2}{n}}$$

(eq. 6.2)

$x_{1,i}$ and $x_{2,i}$ are the values of the normalized *SM* content (θ_1 and θ_2) at a certain location i. n is the total amount of measurements in this location.

6.2.4.2   STANDARDIZED MEAN DIFFERENCE

The Mean Difference (MD) is the difference between two normalized mean values, divided by the within group standard deviation. The latter is computed by adding the variations of the two datasets that are compared, dividing it by 2 and taking the square root. It expresses the size of the intervention effect in each study relative to the variability observed in that study.

$$MD = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{\sigma_x^2 + \sigma_y^2}{2}}}$$

(eq 6.3)

6.2.4.3   PEARSON SAMPLE CORRELATION COEFFICIENT

The Pearson correlation coefficient (r) is a measure of dependence between two quantities. The value is 1 in the case of a perfect positive (increasing) linear relationship, −1 when there is a perfect decreasing (negative) linear relationship, and some value between −1 and 1 in all other cases. It then indicates the degree of linear dependence between the variables, which is stronger if the value is closer to (-)1. The formula to calculate the coefficient is

$$r_{xy} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

(eq. 6.4)

6.2.4.4   NASH-SUTCLIFFE EFFICIENCY COEFFICIENT

The Nash-Sutcliffe Efficiency Coefficient (E), further called Nash coefficient, was originally developed to assess the predictive power of hydrologic models. The formula used is:

$$E = 1 - \frac{\sum_{t=1}^{T}(Q_o^t - Q_m^t)^2}{\sum_{t=1}^{T}(Q_o^t - Q_o)^2}$$ (eq. 6.5)

Q_o is observed discharge, and Q_m is modeled discharge. The coefficient ranges from $-\infty$ to 1. A value of 1 shows a perfect match, a value of 0 means that the model is as accurate as the mean of the observed data. A value lower than 0 means the observed mean is a better predictor than the model.

For the comparisons with EPFL data, the EPFL data is always used as the observed data. For the other comparisons, the dataset other than SMOS is considered the observed dataset.

6.2.4.5 FALSE/TRUE POSITIVE/NEGATIVE

False/True Positive/Negative (FP/FN/TP/TN) between TRMM and other datasets represents whether SM increases for rainfall and decreases for dry periods. It is in that way comparable to the Pearson coefficient, although it does not look at the size of the in- or decrease. To determine the correlation, change in SM has been compared to rainfall. If for example rain would have fallen on July 13th after the SMOS overpass, one could assume that SMOS SM on July 14th is higher than on July 13th. Change in SM would therefore be positive. If this is indeed the case, this correlation is called a true positive (TP). A false positive (FP) would in turn mean a rise in SM, while there was no rainfall. A true negative (TN) would represent a drop in SM, while no rain has fallen, a false negative (FN) a drop in SM while there was rain.

What TRMM date is used for the calculation depends on what SM data is available. When a change in SMOS SM is determined and the second dataset is a morning overpass, it is compared to TRMM data of the day before the second day. This means that for the change in SM from July 13th to July 14th, TRMM data is used from July 13th. For the change from July 11th to July 13th (there is no SMOS data for July 12th available), TRMM data from July 12th is used. If the second overpass is an evening overpass, it is compared to TRMM data of the second day. VUA data is again compared to rainfall of the day before the second day, AMSR-E data to rainfall of the second day, as is EPFL data. This method is used because only daily rainfall amounts were available. Always using e.g. rainfall of the second day could lead to wrongfully determined correlations. It is hoped that this is minimized by this method.

6.2.5 HYPOTHESES

Several hypotheses can be formulated with the information on datasets and methods in the previous paragraphs.

1. All datasets will react to precipitation by a raise in SM content. This raise will be relatively equal.
2. The middle of Burkina Faso receives a lot of rain from July 20th to July 22th. Because the rest of the area does not receive much rain in the same period, the SM content in the middle of Burkina Faso on 23rd of July will be significantly higher than the rest of the area.

3. Temperatures in an evening overpass are higher than for overpasses in the morning. For the τ-ω model, this means that with all other parameters kept the same, this leads to a higher calculated temperature brightness. This in turn means a lower *SM* content.
4. AMSR-E VUA and SMOS use a same kind of model, the τ-ω model, to determine *SM* content. The values of these datasets will therefore be closest to each other and show the highest coefficients.
5. The different datasets, although they show different absolute values, should react in the same manner to precipitation. Therefore, the Pearson coefficient of the comparisons of SMOS data with other datasets will be positive and closer to 1 than to 0.
6. The Nash coefficient will be closer to 1 than to 0 for the comparison of SMOS and VUA data.
7. For the other comparisons, except for those with EPFL data, the Nash coefficient will still be higher than 0, because they are all estimates over larger areas.
8. Because EPFL data has high values compared to the other datasets, the Nash coefficient of the comparison with this dataset will be lower than the other coefficients.
9. The Mean Difference of the different datasets will be comparable to each other, because the reaction to precipitation is the same.
10. The RMSE will be largest for the comparison with EPFL data and smallest for the comparison of SMOS and VUA data.

6.3 RESULTS

This paragraph and Appendix E give the results of the comparisons described above. Besides that, graphs of the accumulated rainfall over the area from July 11th to July 30th and a figure with the locations of the pixels specifically mentioned in this paragraph is also given in the appendix.

Paragraph 6.3.1 gives a short summary of the most important rainfall events. In paragraph 6.3.2, the results of the comparison of the different satellite datasets with the EPFL data are given. In paragraph 6.3.3, the results of the comparisons of SMOS and another satellite dataset are given. Paragraph 6.3.4 continues with scattering graphs of the different satellite datasets. Scattering plots of the complete research area can be found in Appendix E. In paragraph 6.3.5, the values for the different coefficients are given of the pixels with the highest count. It also shows figures of the values of these coefficients over the complete area for the comparison NSIDC-SMOS. The same figures and those for the comparisons ERS-SMOS and VUA-SMOS are given in Appendix E.

6.3.1 PRECIPITATION EVENTS

Rainfall data was available from July 11th to July 31th. Accumulated rainfall in that period ranges from 0 to 220 mm per day. Most of the time, the largest amount of rain falls in a small area, while the remainder of the area receives 0 to maximum 40 mm. The areas that are most often hit with intensive rain are the western border of Ghana and the middle of Burkina Faso. Figure 6.2

shows as an example the precipitation in the area on July 20th, with maximum precipitation falling in the middle of Burkina Faso.
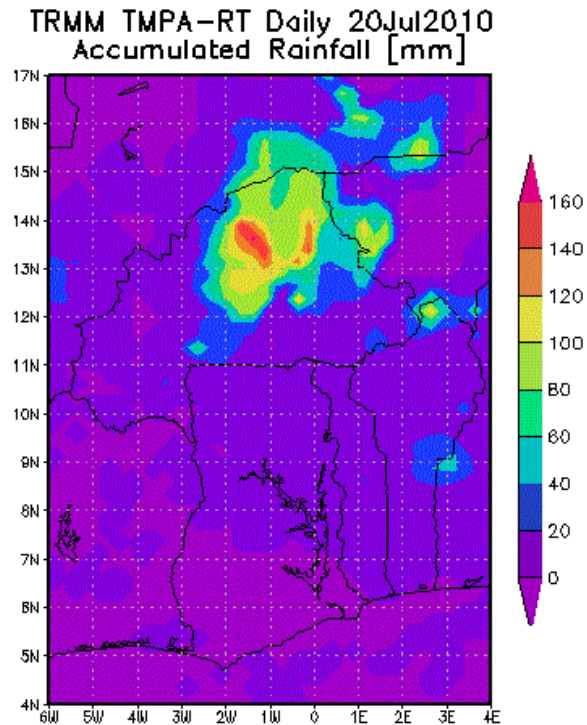
Figure 6.2: TRMM precipitation data for July 20th, 2010, over research area

6.3.2 COMPARING SATELLITE DATA WITH EPFL MEASUREMENTS

Figure 6.3a-d show the graphs of the comparisons of the different satellite datasets with the EPFL data. The values in the graphs are not normalized. On July 13th, 15th and 16th and August 2nd and 5th, SMOS has two values. This is because data was available from two overpasses of the satellite, one in the morning (somewhere between 04:00h and 08:00h) and one in the evening (somewhere between 16:00h and 20:00h). Values for the statistical properties can be found in paragraph 6.3.5.
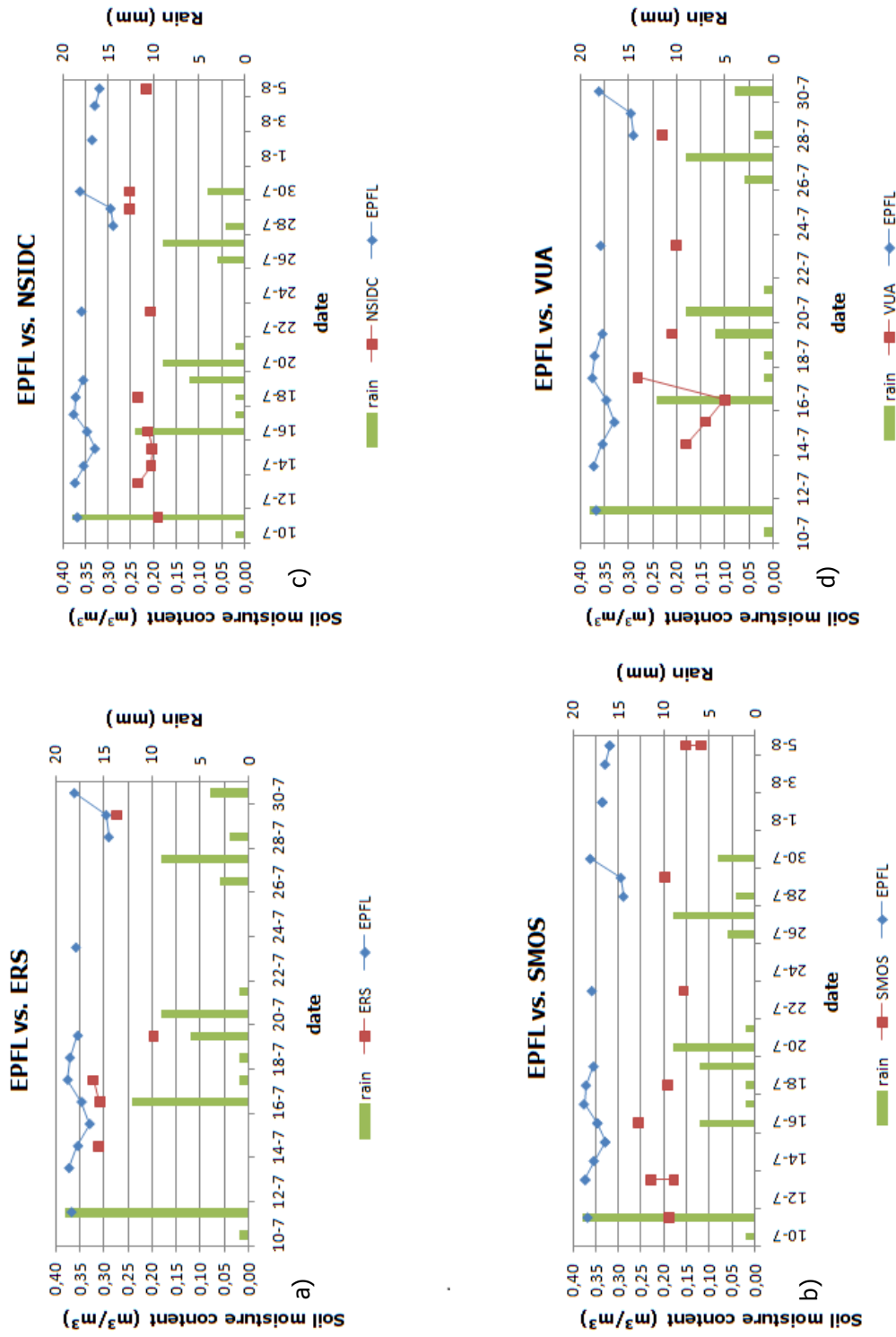
Figure 6.3a-d: Graphs of respectively ERS-ASCAT, AMSR-E NSIDC, AMSR-E VUA and SMOS satellite data compared to EPFL measurements at location of EPFL meters, including accumulated precipitation over the day in the area. Statistics can be found in paragraph 6.3.5

The comparisons show the large difference in absolute values of the satellite data compared to EPFL measurements. The only dataset that does show values in the same range as EPFL is ERS. ERS' temporal pattern shows, when disregarding July 19th, a smaller range than EPFL. NSIDC also shows a small range (only 5%) and, in contrast with ERS, shows relative low absolute values. SMOS data shows a bigger range, but it is very hard to see whether the pattern is comparable to that of EPFL. Due to the large change in value of both ERS (July 19th) and VUA (July 17th), SMOS shows a smaller range than these datasets. The first four data points of VUA show some similarity to the EPFL pattern. The data however react a day later to precipitation. The following raise is far larger than that of EPFL. The reaction of the EPFL datasets seem to follow precipitation data, but do show a shift in time, probably caused by the influence of the stream on the *SM* content.

6.3.3 COMPARING SATELLITE DATA WITH EACH OTHER

Because of the orbit of the satellites, the pixels with the highest count for the comparison of SMOS with NSIDC are to the east of the area, whereas the highest count for the other comparisons are to the west. The location of these pixels can be found in figure E.7 (appendix E). Below, a number of graphs is displayed in which SMOS is compared to other satellite data. The pixels are chosen based on their location. Pixels with the highest count form groups, of which one pixel is chosen to be displayed in this chapter. The exact location of the different pixels can be found in Appendix E. Figure 6.4a-d, Figure 6.5a-d and Figure 6.6a-d show the graphs of the different comparisons. Since it is already clear what bias the different datasets have compared to EPFL measurements, graphs are based on normalized values.
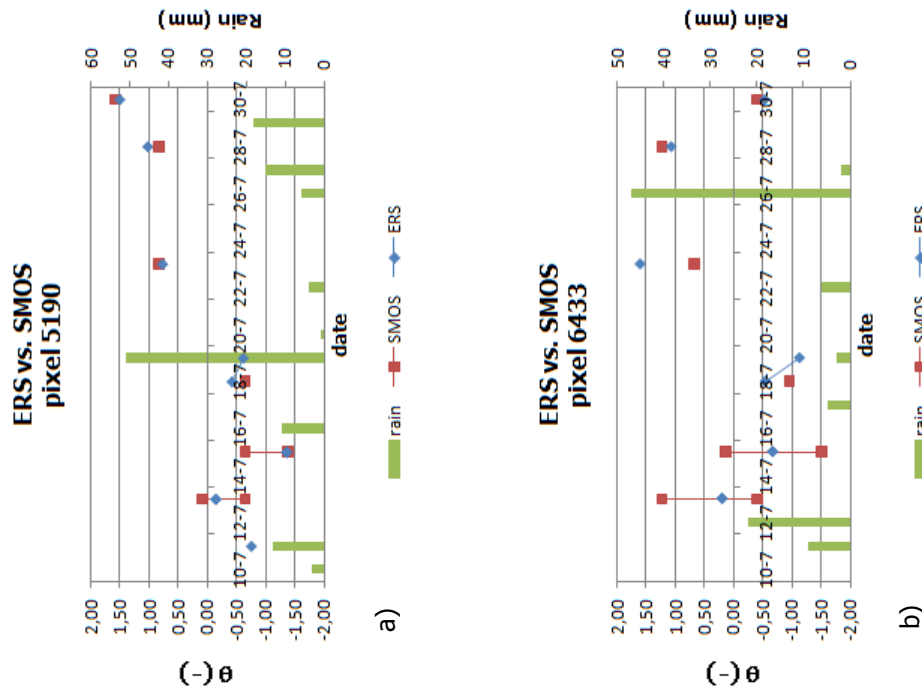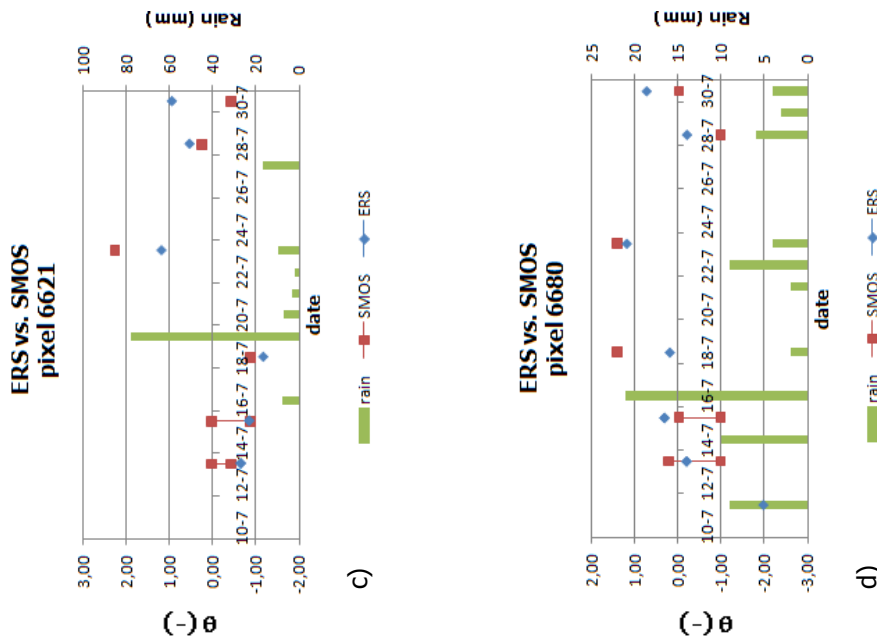
Figure 6.4a-d: Graphs of ERS and SMOS satellite data, including accumulated precipitation over the day in the specific area. Statistics can be found in paragraph 6.3.5

ERS and SMOS have some data points that are almost exactly the same, but at the same time points with a difference of 1 or more. There does not seem to be any correlation between data being similar or not and precipitation events or latitude. Whether ERS has a bigger similarity with morning or evening SMOS data depends on location and date. From the four pixels show here, pixel 5190, the second pixel from the north, shows the best fit of SMOS with ERS.
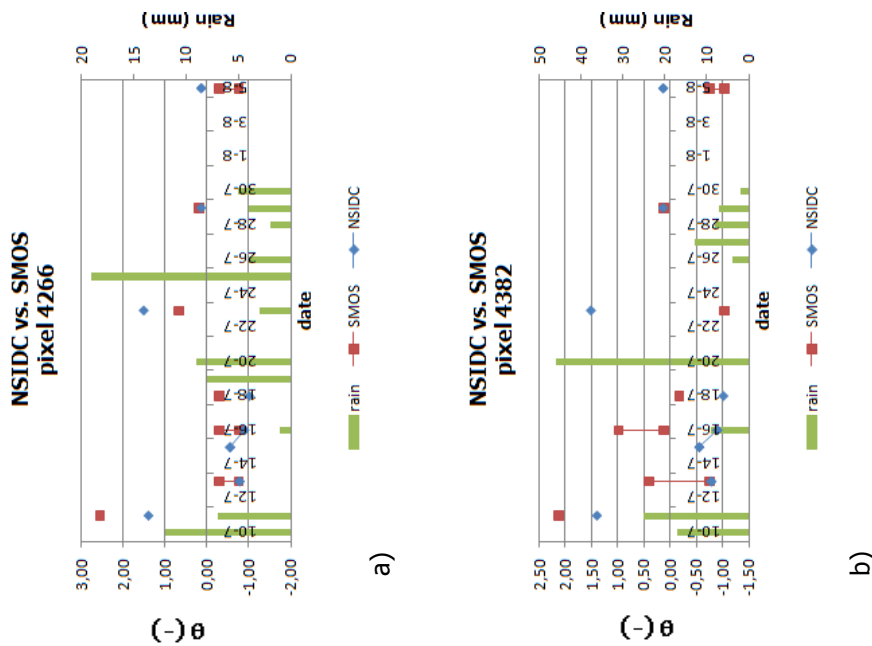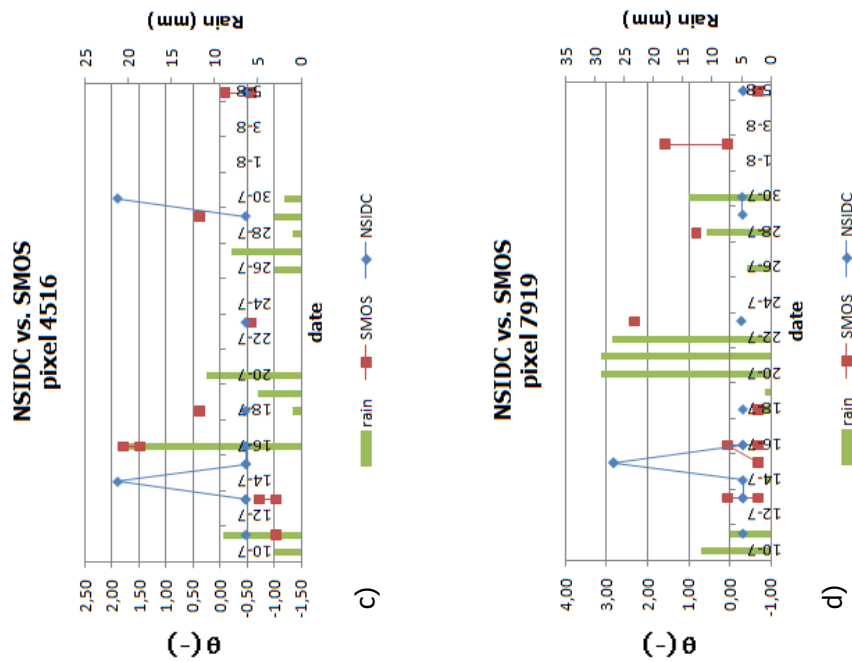
Figure 6.5a-d: Graphs of NSIDC and SMOS satellite data, including accumulated precipitation over the day in the specific area. Statistics can be found in paragraph 6.3.5

NSIDC data fits less well to SMOS data than ERS. Again, no clear link between precipitation or location and similarity can be found. The two pixels at the middle latitudes seem to have the best fit.
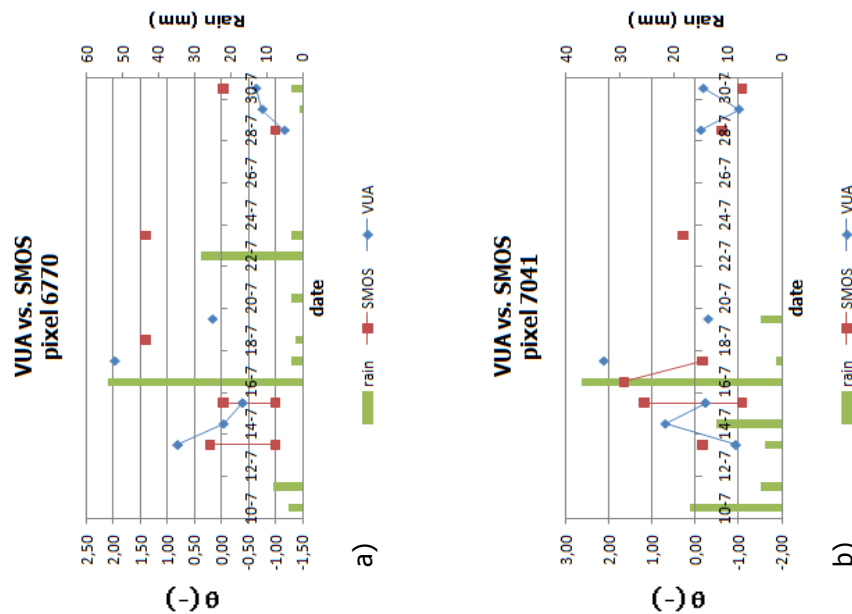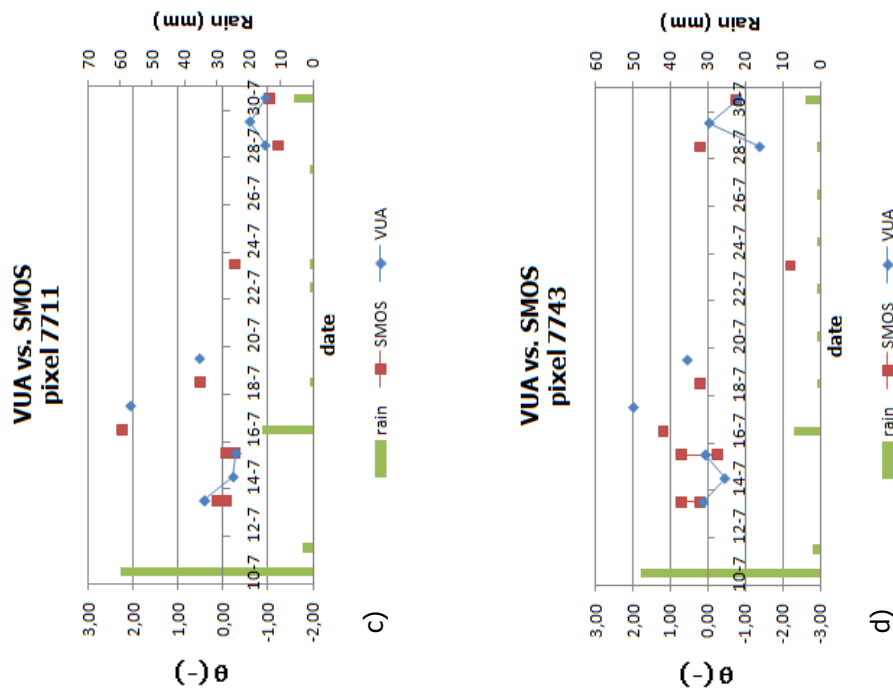
Figure 6.6a-d: Graphs of VUA and SMOS satellite data, including accumulated precipitation over the day in the specific area. Statistics can be found in paragraph 6.3.5

The little data that is available for VUA and SMOS show a relatively good similarity, especially when one takes into account that VUA might sometimes reacts slower than SMOS (as it does compared to EPFL). Looking at SMOS data from July 16th to 18th, the patterns seems to be similar to that of VUA from July 17th to 19th. Just as with the other comparisons, it is not very clear whether the morning or evening overpass shows more similarity with the VUA data.

6.3.4 SCATTER PLOTS OF SATELLITE DATA

Besides comparing data in time, it is also possible to compare values at a certain moment and place with each other in a scatter plot. On the x-axis, the value of the first dataset is placed. On the y-axis, the value of the second. If two datasets show a correlation, it is possible to draw a line in a scattering plot. In appendix E, the locations of the pixels with the highest count can be found.

Figure 6.7 shows the scattering plots of SMOS data compared to another satellite dataset over the pixels with the highest count of the research area. In Appendix E, scattering plots can be found over the entire research area. The area is divided into 9 area, North West, North, North East, West, Middle, East, South West, South and South East.
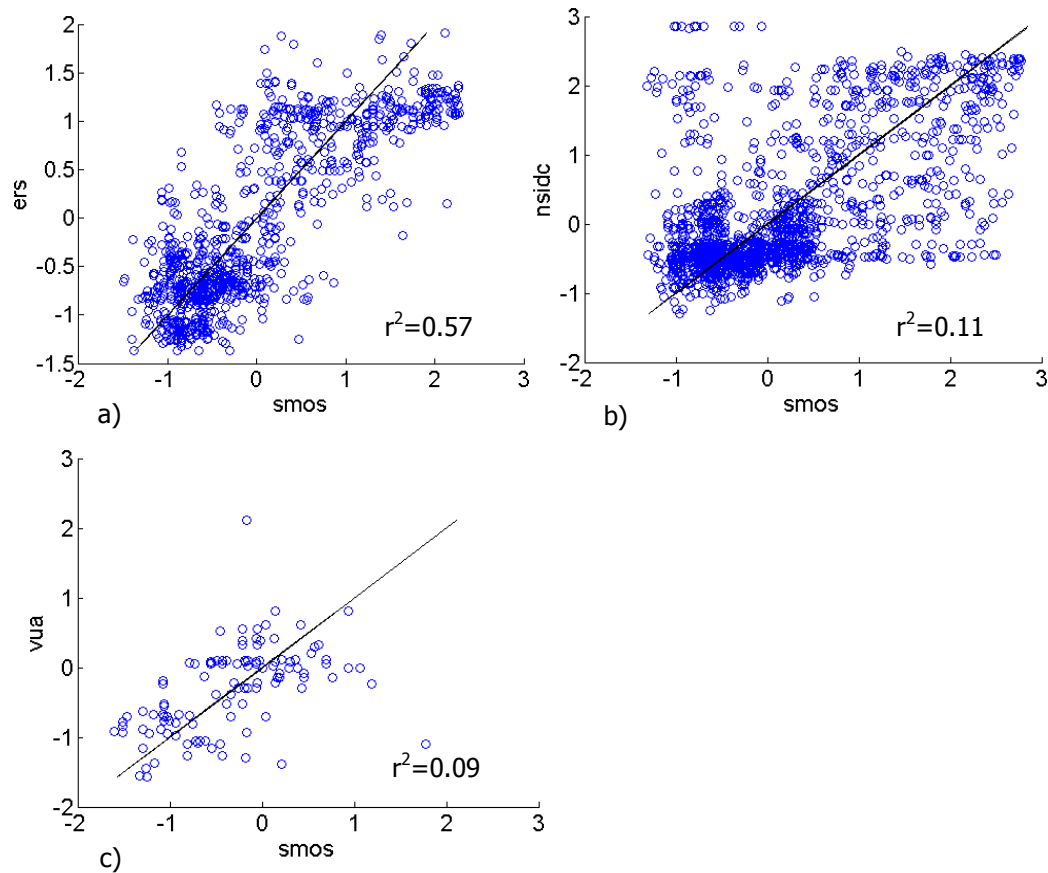
Figure 6.7a-c: Scattering plots of pixels with the highest count when comparing SMOS to respectively ERS, NSIDC and VUA data

At first sight, the scattering plots of ERS-SMOS and VUA-SMOS seem to show the highest correlation. The r^2 value tells a different story however. Looking at those values, only ERS has a reasonable correlation with ERS, whereas VUA has a lower value than NSIDC, both very low, possibly because of the effect the amount of data has on the mean. Both ERS as NSIDC show a thick cloud of data points around the lower regions of the normalized values and a hint of a cloud at the higher regions. The cloud in the lower regions of the ERS-SMOS graph seems to show a better correlation than that of the NSIDC-SMOS graph. The clouds in the higher regions show less correlation.

Other scatter plots can be found in appendix E. Overall, they show a better correlation for northern areas. Again, r^2 is higher for NSIDC than for VUA, while the scatter plots seem to show a better fit (at least for some areas) for VUA.

6.3.5 VALUES FOR COEFFICIENTS

In Table 6.1, the values for the different coefficients calculated for the pixels that are also shown in the above graphs in Figure, 6.3 to Figure 6.6 are given. In Table 6.2, the values averaged over all pixels with the highest count are given.

The ERS and VUA data is derived from morning overpasses. Because the graphs from paragraph 6.3.3 show such large difference between morning and evening overpasses, MD, RMSE and Pearson and Nash coefficient are also determined for a comparison with only morning SMOS data. The outcome can be found in Table 6.3.

6.3.6 COMPARING SATELLITE DATA WITH TRMM DATA

Table 6.4 finally gives the values for the correlation of the different *SM* datasets with TRMM data. It also gives the chance of such a correlation happening for completely uncorrelated data.

Data is always calculated with normalized data, also for the comparison of the different satellite datasets with EPFL. With the bias known from the visual inspection, it is interesting to look at the difference in pattern from the normalized value.

Table 6.1: Coefficients for different graphs as found in Figure 6.3- Figure 6.6

Figure	Comparison (Pixel no.)	Count	Mean Difference	RMSD	Pearson sample coefficient	Nash coefficient
6.3a	EPFL-ERS	5	0.358	1.17	0.15	-0.41
6.3b	EPFL-NSIDC	9	0.15	0.17	0.17	0.08
6.3c	EPFL-VUA	7	-0.23	1.19	0.28	-6.14
6.3d	EPFL-SMOS	8	2.55	2.63	-0.36	-5.11
6.4a	SMOS-ERS (5190)	8	-0.153	0.12	0.93	0.89
6.4b	SMOS-ERS- (6433)	8	-0.05	0.48	0.69	0.28
6.4c	SMOS-ERS (6621)	8	-0.19	0.55	0.68	0.23
6.4c	SMOS-ERS (6680)	8	0.01	0.67	0.50	-1.42
6.5a	SMOS- NSIDC (4266)	10	-0.18	0.42	0.76	0.48
6.5b	SMOS-NSIDC (4382)	10	0.03	0.64	0.69	0.43
6.5c	SMOS-NSIDC (4516)	10	-0.15	1.12	0.01	-8.81
6.5d	SMOS-NSIDC (7919)	10	0.12	3.30	-0.27	-1.01
6.6c	SMOS-VUA (6770)	6	0.05	0.156	0.91	0.73
6.6b	SMOS-VUA (7041)	6	-0.12	1.594	-0.03	-0.69
6.6c	SMOS-VUA (7711)	6	0.01	0.08	0.92	0.81
6.6d	SMOS-VUA (7743)	6	0.01	0.57	0.20	-0.28

The statistical coefficients of the comparison of EPFL with the different datasets confirms the conclusions from the visual inspection; the correlation is not very high. SMOS has a very high MD

and RMSD and negative Pearson and Nash values. VUA has a relatively low MD, although it RMSD is high. This is caused by its large standard deviation and its shift in time. This is also reflected in its low Nash value. ERS shows a relatively high RMSD, possibly because of the extreme value on July 19th. NSIDC, with its small range, shows the best correlation with EPFL, due to the fact that data is normalized. Its Nash and Pearson correlation coefficients are however still low.

Looking at the statistical properties of the different datasets for the four pixels who's graph is given, it makes a big difference which pixel is chosen. ERS pixel 6621 has the (absolute) largest value for MD. The lowest value can however be found with ERS pixel 6680 (and VUA pixels 7711 and 7743). The lowest RMSE is found for VUA pixel 7711, but the highest value is also found for a VUA pixel (7041). The highest Pearson coefficient can be found at ERS and VUA, the lowest at NSIDC. The same goes for the Nash coefficient. The best comparisons overall seems to be found in the north of Ghana, at the western border, and to the north of Burkina Faso (with the exception of pixel 7919). Thus, the VUA pixels most to the north and the ERS pixels show the best correlations. For NSIDC, data, this distinction is not as clear.

Table 6.2: Statistics properties for pixels with the highest count

SMOS compared to	Number of pixels	Count	MD minimum	RMSE minimum	Pearson minimum	Nash minimum
ERS	115	8	-0.26	0.07	0.33	-1.42
NSIDC	204	10	-0.22	0.08	-0.41	-18.57
VUA	20	6	-0.12	0.08	-0.20	-2.37
SMOS compared to	Number of pixels	Count	MD maximum	RMSE maximum	Pearson maximum	Nash maximum
ERS	115	8	0.04	1.07	0.96	0.92
NSIDC	204	10	0.13	3.70	0.96	0.89
VUA	20	6	0.09	1.85	0.92	0.81
SMOS compared to	Number of pixels	Count	MD average	RMSE average	Pearson average	Nash average
ERS	115	8	-0.16	0.32	0.81	0.49
NSIDC	204	10	-0.06	0.73	0.59	-0.72
VUA	20	6	-0.00	0.37	0.65	0.22
SMOS compared to	Number of pixels	Count	MD median	RMSE median	Pearson median	Nash median
ERS	115	8	-0.16	0.25	0.85	0.68
NSIDC	204	10	-0.08	0.59	0.67	0.34
VUA	20	6	-0.01	0.21	0.78	0.46

Table 6.2 shows that ERS has the smallest range for Pearson. All its values are positive. Its median value and its average value are thus also the highest. ERS median and average Nash coefficient are also the highest, and its range for RMSE is the smallest. The smallest median value for RMSE however can be found for VUA. The same goes for MD. Median and average value of Pearson and Nash are lowest for NSIDC, although it does not have the highest maximum value for these coefficients.

In order to calculate the chance of the statistic parameters found above to be coincidental, the same calculations were done with random numbers instead of the ERS, NSIDC and VUA data.

The change for a median value or higher cq. lower of MD and RMSE as found in Table 6.2 are relatively large, 20 to 45%, depending on the amount of pixels with the highest count. The chance for a median value or higher for Pearson and Nash was 1% or less. The low chance for Pearson and Nash is in accordance what was found in automated Matlab calculations. The big difference in chance for MD and RMSD on one side and Pearson and Nash at the other side is possibly caused by high mean values of the datasets. This has no direct effect on MD and RMSD, but can cause Pearson and Nash to become high. The low values for Pearson and Nash seem to suggest that the median values are caused by actual correlation between the datasets rather than chance.

Table 6.3: Statistics properties for pixels with the highest count when comparing ERS and VUA data to only SMOS morning overpasses

SMOS compared to	Number of pixels	Count	MD minimum	RMSE minimum	Pearson minimum	Nash minimum
ERS	115	2-4	0.00	0.00	-0.74	-17.52
VUA	20	7-9	0.00	0.01	0.33	-0.72
SMOS compared to	Number of pixels	Count	MD maximum	RMSE maximum	Pearson maximum	Nash maximum
ERS	115	2-4	0.00	2.53	1.00	1.00
VUA	20	7-9	0.00	0.84	1.00	0.98
SMOS compared to	Number of pixels	Count	MD average	RMSE average	Pearson average	Nash average
ERS	115	2-4	0.00	0.47	0.70	0.17
VUA	20	7-9	0.00	0.34	0.77	0.33
SMOS compared to	Number of pixels	Count	MD median	RMSE median	Pearson median	Nash median
ERS	115	2-4	0.00	0.22	0.89	0.79
VUA	20	7-9	0.00	0.29	0.83	0.38

Table 6.3 shows the values for the different statistic coefficients, but in this case for comparisons of only morning overpasses. Taking out evening overpasses changes MD to zero, because the mean values of the different datasets becomes equal. The minimum value for RMSE goes down, but the maximum goes up, meaning the average and median does not consequently go up or down. The median value for VUA goes up with 0.08, but goes done with 0.03 for ERS. The median value for Pearson goes up for both datasets, but Nash goes only up for ERS, not for VUA. The maximum change for the median of Pearson is 0.05, for Nash this is 0.11 (but its drop for VUA is 0.08).

6.3.7 COMPARING SATELLITE DATA WITH TRMM DATA

Table 6.4: Correlation between different datasets and chance of this correlation happening to unrelated datasets. The first number under the random correlation is the average value, the second the standard deviation

Dataset compared to TRMM	Correlation (average no.)				Random correlation (average no./ standard deviation)			
	TP	TN	FP	FN	TP	TN	FP	FN
EPFL	0.3636	0.3636	0.1818	0.0909	-	-	-	-
SMOS	0.1637	0.4787	0.1383	0.2193	0.1979/ 0.1392	0.3002/ 0.1465	0.3034/ 0.1473	0.1985/ 0.1378
NSIDC	0.2640	0.3743	0.1760	0.1857	0.1892/ 0.1121	0.3124/ 0.1218	0.3113/ 0.1203	0.1871/ 0.1099
VUA	0.3067	0.3989	0.1783	0.1182	0.1868/ 0.1071	0.3167/ 0.1179	0.3155/ 0.1183	0.1810/ 0.1070
ERS	0.2625	0.2970	0.3098	0.1307	0.1686/ 0.1074	0.3341/ 0.1216	0.3302/ 0.1194	0.1671/ 0.1074

Table 6.4 shows the outcome of the comparisons of the different datasets with TRMM data. Determining true/false positive/negative gives an idea of the overall reaction to rainfall of the different datasets. The numbers are also compared to numbers determined from a random sample of the same size, compared to the same rainfall. EPFL measurements are not compared to a random sample. The size of that sample would become 1*11, which is too low to get an representative average. EPFL values are compared to values of the other random samples.

From a random sample, approximately half of the derivatives can be called TP or TN. All datasets show a higher value for the sum of TP and TN than 0.5, ranging from 0.5539 for ERS to 0.728 for EPFL. The sum of TP and TN for SMOS lies between the values of ERS and VUA, the two datasets that, according to the other statistics, show most resemblance to SMOS. The NSIDC value however lies far closer to SMOS. TN for NSIDC is actually lower than for a random sample, as is the TN of ERS. On the other hand, the TP for SMOS is lower that the value of the random sample. All satellite datasets show a high value for FN, compared to the random sample. The chance that the values are caused by chance is relatively high. Values for TP and TN are never larger than the mean + 2 * standard deviation as calculated with random numbers. The chance of the valued obtained by chance is thus always larger than 5%.

EPFL data shows the highest sum of TP and TN, despite the shift in time seen in the graphs. Assuming this is the highest value possible with the available rain data (and the gap of rain and satellite measurement), any value that falls within 10% of this may still be considered well correlated to rain. Only VUA falls within 10% of the EPFL value. SMOS falls within 12% of it. The other values show bigger differences with EPFL.

Besides a calculation, it is also possible to check visually whether *SM* has an impact on *SM*. Because of the changing orbits and time between overpasses, it is not always possible to check whether rainfall results in a rise in *SM*. It is possible to see whether *SM* is higher on places where rain has fallen than in the surroundings. This is often the case, but not easy to spot. Often,

rainfall of the previous day has to be taken into account as well, to understand how *SM* is divided. Besides that, the orbit of the satellite can prohibit a visual inspection, if it does not pass the area where rain has fallen. An example of a comparison of rainfall and *SM* is given in Figure 6.8. This comparison also shows the high *SM* values in the middle of Burkina Faso for July 23rd.
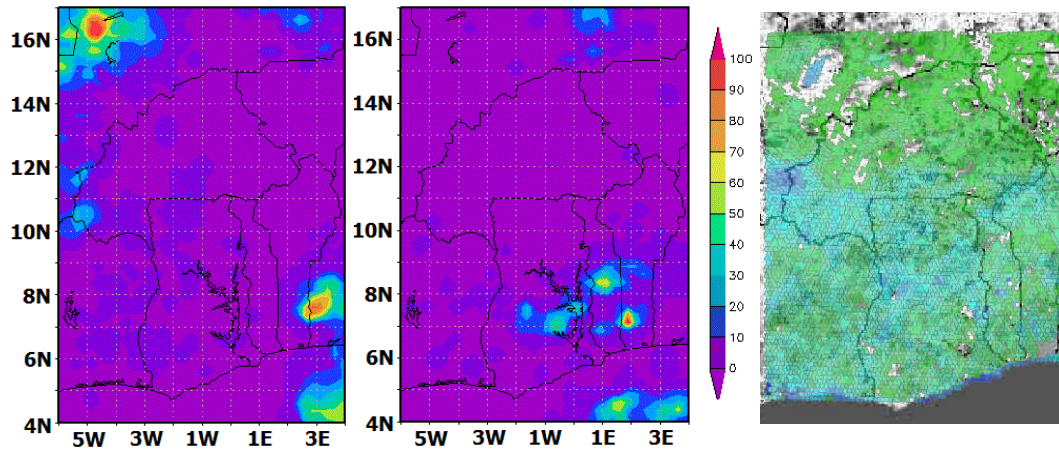
Figure 6.8: Comparing *SM* according to SMOS (right) on July 13th with accumulated rainfall in mm on July 12th (left) and July 13th (middle). In the left figure, green stands for low *SM* content, blue for high
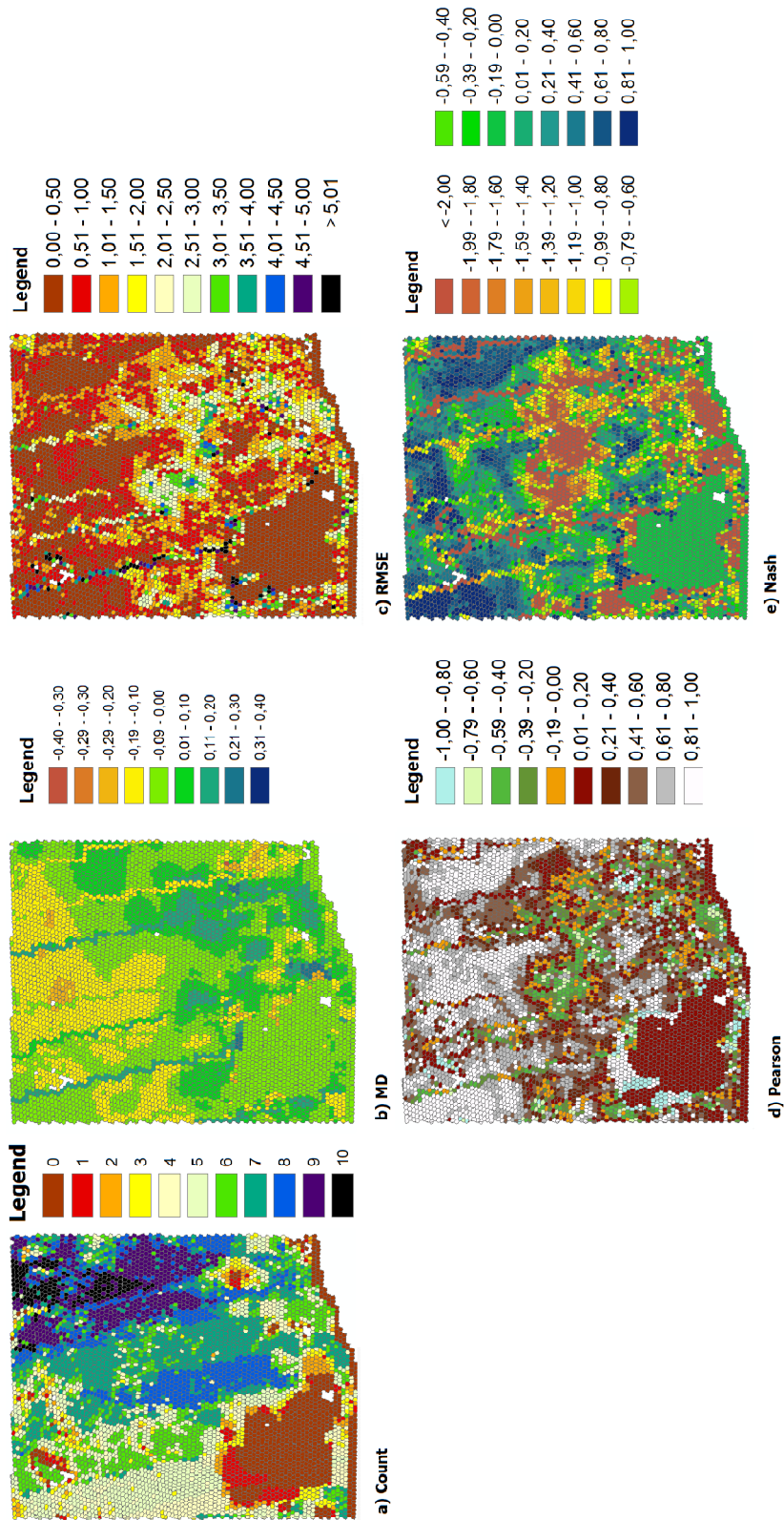
Figure 6.9: Count and statistical properties NSIDC-SMOS comparison

The graphs of the different statistics show that the best correlation can be found in the north. In the south, it depends very much on which comparison you are looking at which area has the best correlation. The strange lines from north to south (slightly going to the east) in the NSIDC figures are created by the fact that the overlap of the satellite's overpasses is not perfect. Please note the large area in the south west with a count of 0. None of the satellite datasets show a clear (relative) better correlation with SMOS.

6.4 DISCUSSION

6.4.1 COMPARING EPFL MEASUREMENTS WITH OTHER DATASETS

Visual inspection of the satellite data compared to the EPFL data shows a large difference between absolute values. The EPFL measurements show as expected higher *SM* values. This is probably caused by the fact that the EPFL meters are located near a stream and over only part of the pixel studied, while the satellite measurements give an average *SM* value over the complete pixel. The high EPFL values causes all satellite datasets to have a dry bias. This in contract with Rudiger *et al.*, who show a wet bias for VUA. What is in agreement with this paper is the fact that NSIDC shows a lack in dynamic.

Both the visual inspection as the statistical coefficients show that the correlation between EPFL measurements and satellite data is very limited. This was not expected. Coefficients reveals NSIDC to be the dataset that correlate best to EPFL, despite its limited dynamics. The low values for VUA can be explained by an apparent delay, that of SMOS is hard to determine visually.

A possible explanation for the low correlation of normalized values (thus not considering the dry bias) is the depth satellites measure. Only the SMOS satellite measures up to a depth of five centimeter, while EPFL is measured at five centimeter. Since the top soils gets wet faster and dries out faster, this causes differences in measured *SM*. The delayed reaction of VUA might be caused by the fact that the *SM* value for EPFL is determined by taking the average over the entire day, while VUA measurements are done in the morning. If the precipitation falls after the overpass, the satellite cannot respond till the next day. Finally, it could very well be that the *SM* situation at the EPFL measurement site shows a different pattern than the complete pixel, due to the influence of the stream.

6.4.2 COMPARING SATELLITE DATASETS WITH SMOS

Both VUA and ERS show a relative good correlation with SMOS when comparing them with each other, as was expected. However, none of the graphs gives a clear indication why some data points are closer to each other than others, nor do the statistical coefficients. Pixels with the best correlation lay to the north and on the western border of Ghana. These pixels lie in an area the total amount of rainfall over the entire period is actually relatively low. It thus seems that not a single precipitation event, but rather the total amount of rainfall in an area is important for a

good correlation. The ERS data does not show a better correlation in the south with SMOS than NSIDC or VUA.

Overall, the ERS data shows the best Pearson and Nash values when compared to SMOS, while VUA shows the best RMSE and MD. The best Pearson and Nash value of ERS can be explained by the delay of VUA. The best RMSE and MD values can be explained by the fact that VUA and SMOS have values that lie closer to each other and show a larger range than ERS does.

Taking out evening overpasses from SMOS does not necessarily improve the correlation between SMOS and other datasets. This can be explained by the fact that it very much depends on the date and location, whether the other datasets are more similar to the morning or the evening overpass.

It should be noted that the count for all comparisons is relatively low. This means that their values could be based on coincidence. The data that can be compared might be erroneous. Because there are only a few dates comparable, the impact of this erroneous data on statistic properties can be big. Taking this into account, the values of the comparisons with VUA and ERS do not differ much, while the comparison with the NSIDC dataset, with a higher count, shows statistical values that represent both a different pattern and different values. This is in according with findings from Rudiger *et al.*

The graphs of the different statistics show that the best correlation can be found in the north. In the south, best correlations can be found in areas with little precipitation and vegetation. In the comparisons with EPFL, both ERS and VUA showed a larger range than SMOS. Although for both dataset this was caused by only one data point, it is expected that this is the reason for the low correlation of SMOS in wet areas. SMOS apparently responds with a smaller rise or fall in *SM* due to precipitation/drying out. Again, it should be noted that this conclusion is based on little data.

6.4.3 SCATTER PLOTS

Although the r^2 values beg to differ, visual inspection of the scatter plots shows again the best correlation between SMOS and VUA and SMOS and ERS. ERS has the highest r^2 value, but VUA has far less data to compare with. The plots over the entire area confirm that correlations get better to the north. Above however, it was already concluded that some of the pixels with the highest correlation coefficients are located in places more to the south with little precipitation.

6.4.4 COMPARING SATELLITE DATA WITH TRMM DATA

A visual inspection of TRMM data compared to rainfall data does indeed show that rainfall causes higher *SM* values. However, sometimes there is a shift in time of the reaction. A good impression of the reaction of *SM* to precipitation can be acquired by looking at rainfall of more than one day.

In the true/false positive/negative comparison, VUA shows the highest TP+TN of all satellite datasets, despite its apparent delay in comparison with SMOS and EPFL. Apparently, over the complete research area, it has found a balance between the EPFL and SMOS dataset that gives such a high value. SMOS has the second highest value. It is however not within 10% of the

TP/TN values of EPFL, thus not as good as expected. NSIDC, despite its lack in *SM* change as displayed in the comparison with EPFL, does move enough to have a higher TP+TN than ERS. Since SMOS has a very good correlation with both ERS and VUA, this might seem strange at first. However, the comparison with TRMM data is completely separate from the comparisons of satellite data with each other. It can therefore well be that the data on which rainfall is checked with satellite data are different data than satellite data are compared to each other.

The high value for FN for all satellite data can be caused by the fact that precipitation amounts are often small. Rain data was chosen that falls at the most nine hours before the satellite overpass. This water should have already disappeared for the *SM* content to go done. This is very well possible, as with small amounts of rainfall a large part is already evaporated before it can become *SM*.

It should be noted that the chance that values for TP etc. is obtained by chance is relatively high. This despite the effort of adapting the rainfall data to the time of the satellite overpass.

6.5 CONCLUSIONS

In paragraph 6.5.1, first of all the hypotheses from paragraph 6.2.4 are tested with the information of the previous paragraphs. Paragraph 6.5.2 then continues with an overall conclusion on the reliability of SMOS data.

6.5.1 HYPOTHESES TESTED

1. All datasets will react to precipitation by a raise in *SM* content. This raise will be relatively equal.

With a visual inspection, most of the time, *SM* data does indeed go up with precipitation and down in dry periods. TP, TN, FP, FN statistics do however not always agree with that conclusion. Reason for this is that in the visual inspections, it is possible to look at the data over time, and see delayed reactions to precipitation. This is not possible with the statistics as used in this research. Whenever there is a raise due to precipitation, it is not equal for the different datasets. This is clearly shown in Figure 6.3.

2. The middle of Burkina Faso receives a lot of rain from July 20th to July 22th. Because the rest of the area does not receive much rain in the same period, the *SM* content in the middle of Burkina Faso on 23rd of July will be significantly higher than the rest of the area.

SM content in the middle of Burkina Faso is significantly higher than the preceding days on July 23th, due to the rain that fell in the period before. It does not necessarily show the highest value, because of differences in rainfall over time in the area.

3. Temperatures in an evening overpass are higher than for overpasses in the morning. For the τ-ω model, this means that with all other parameters kept the same, this leads to a higher calculated temperature brightness. This in turn means a lower SM content.

As long as it does not rain in between two overpasses, SMOS SM values of a morning overpass are indeed higher than that of a evening overpass.

4. AMSR-E VUA and SMOS use a same kind of model, the τ-ω model, to determine SM content. The values of these datasets will therefore be closest to each other and show the highest coefficients.

The best fit according to statistical coefficients between two normalized satellite datasets is between ERS and SMOS data, not between VUA and SMOS data. This is intriguing, because ERS data gives relative SM data, as opposed to the absolute values of SMOS and VUA. It should be noted however that the count of both the ERS-SMOS and the VUA-SMOS and statistical values are relatively close to each other, compared to the NSIDC-SMOS comparison.

5. The different datasets, although they show different absolute values, should react in the same manner to precipitation. Therefore, the Pearson coefficient of the comparisons of SMOS data with other datasets will be positive and closer to 1 than to 0.

The Pearson coefficient is often, but not always positive. Its median value however is always positive. Negative values are probably caused by a delay in reaction of a dataset compared to the reaction of SMOS

6. The Nash coefficient will be closer to 1 than to 0 for the comparison of SMOS and VUA data.

The median Nash coefficient of the VUA-SMOS comparison is 0.46, so it is not closer to 1 than to 0. Most probably, because of the delay of VUA compared to SMOS, data shows data that is too different to get a higher value.

7. For the other comparisons, except for those with EPFL data, the Nash coefficient will still be higher than 0, because they are all estimates over larger areas.

All median Nash values are indeed positive, except for the comparison with EPFL.

8. Because EPFL data has high values compared to the other datasets, the Nash coefficient of the comparison with this dataset will be lower than the other coefficients.

The Nash coefficient of the comparisons with EPFL data are all lower than the median values of the other comparisons.

9. The Mean Difference of the different datasets will be comparable to each other, because the reaction to precipitation is the same.

Looking at the different pixels and the median value, it is clear that MD is not the same for all pixels, mostly because the range of the different datasets differs too much.

10. The RMSE will be largest for the comparison with EPFL data and smallest for the comparison of SMOS and VUA data.

The RMSE of the comparisons VUA-EPFL and SMOS-EPFL are indeed large, but those of the comparisons ERS-EPFL and NSIDC-EPFL are not that large at all. They are actually in the same range as the medians of the comparisons ERS-SMOS and VUA-SMOS. Apparently, although the pattern of ERS and NSIDC does not seem to be comparable to EPFL at a visual inspection, the values that can be compared are close enough to the EPFL data to obtain a low RMSE.

6.5.2 ANWERING THE RESEARCH QUESTIONS

Comparing SMOS and other satellite data with EPFL measurements shows a big difference in absolute values. This is partly caused by the fact that the EPFL meters are located near a stream, on a stretch of only a couple of kilometers, instead of a complete pixel. Besides a difference in absolute values, there is also a difference in the reaction of the different datasets to precipitation, both time wise and looking at the amount of change. Statistical properties suggest that ERS and NSIDC show the best correlation with EPFL, although it is clear from visual inspection that the reaction of NSIDC to precipitation is small, which is not likely to depict the real situation well.

Comparing SMOS to the three other satellite datasets shows the best correlation with VUA and ERS. This fits with the findings of Rudiger *et al.* (2007). This is positive, because these datasets were considered reliable according to Rudiger *et al.* However, it should be taken into account that the error of both datasets is larger than 4%. In order to be sure that SMOS does indeed give the right values, more comparison with field data is still necessary. The good correlation of SMOS TRMM seems to suggest that SMOS does react to precipitation. However, looking at the complete area, it is found that the correlation of SMOS with other datasets is best for the northern part of the research area. This area has less rainfall and less vegetation. Some of the pixels with the best correlation can be found more to the south, but in areas with a small total amount of precipitation. The scattering plots of the pixels with the highest counts also suggest that the very low values are best correlated. This concurs with the observation that the changes due to precipitation are at times smaller for SMOS than for ERS or VUA, causing low correlation in areas with larger amounts of rainfall. The lesser correlation in more vegetated areas can be caused by the fact that the vegetation is too dense for TB to penetrate, but can also simply be caused by the fact that the default parameters used are even less suitable for more these areas than for areas with less vegetation.

It should be noted that the difference between SMOS *SM* data in the morning and evening can become larger than 4%. This does not seem to resemble reality. Besides that, SMOS seems to react quickly to precipitation and drying out, but more data is needed to draw a definite conclusion about this.

Validation of SMOS satellite data over Ghana and Burkina Faso

7 CONCLUSIONS AND RECOMMENDATIONS

7.1 INTRODUCTION

This chapter starts with a summary of the most important conclusions of the three previous chapters. Together, these conclusions give an answer to the research question formulated in paragraph 1.3:

How well does the algorithm as presented in the SMOS *level 2 Processor for SM Algorithm Theoretical Based Document*[17] *perform for northern Ghana and Burkina Faso, West-Africa?*

Besides an answer to the main question of this research, the chapter gives recommendations on how to proceed with further research on validation and calibration of the algorithm for West-Africa.

7.2 SUMMARY PREVIOUS CHAPTERS

7.2.1 CHAPTER 4: SENSITIVITY ANALYSIS

In chapter 4, a sensitivity analysis is executed on the SMOS algorithm. The analysis is done with the so called L2PP processor and with a Matlab code. Because not all auxiliary data was available in a format readable in Matlab, this code uses a less complicated algorithm than the processor.

All parameters have a default value. These default values are used to calculate *SM*. The default values are then, one by one, varied in order to see the effect on the *SM* value. The parameters values are changed in extreme values and location values. The first are the maximum and minimum value the parameter can become globally. The second pair of values are the minimum and maximum value the parameter can obtain in the research area (Ghana/Burkina Faso).

The parameters that affect the outcome of the algorithm most are the scattering albedo, the roughness of the area, temperature and litter properties. The main problem with three of these parameters is that determination of values for an inhomogeneous area is very hard. For temperature, it is important to get a separate estimate for soil and canopy temperature. The

---

[17] ATBD, see paragraph 1.3

difference measured in the field can have a big impact on the outcome of the algorithm. Litter properties finally may well change significantly over the year in West-Africa.

For these parameters, graphs were created that showed *SM* as a function of the parameter. The change of *SM* is often not linear and depends on the original value, the amount of change and the measured *TB*.

7.2.2 CHAPTER 5: COMPARING SMOS TO FIELD DATA

In chapter 5, SMOS data is compared to field data. In April and May 2010, field work was executed in northern Ghana. Its goal was to obtain a well defined average value over an area of 15 by 15 km (a so called SMOS pixel) in order to compare this with SMOS data from the same period.

In order to obtain a well defined average *SM* value for the SMOS pixel, the area was divided into so called hydrotopes. A hydrotopes shows a temporal pattern that is distinguishable from other hydrotopes in the area. Comparing two hydrotopes should always show a difference in their mean larger than their variability. By using hydrotopes, less measurements have to be done in order to get a well defined average. It turned out that for the top five centimeters of soil however, vegetation should also be taken into account in order to get a well defined average.

The final outcome showed that the SMOS value is relatively low compared to the field measurements. The roughness value was adapted in order to get a *SM* value higher than 0, namely 0.2%. With a field average of 2.83%, this does fall within the 4% error bounds set by ESA. It is therefore concluded that, in the dry season, SMOS does fall in these error bounds.

7.2.3 CHAPTER 6: COMPARING SMOS TO TIMESERIES

In chapter 6, SMOS data was compared to other datasets over time. Datasets used were EPFL field measurements, AMSR-E VUA, AMSR-E NSIDC and ERS-Ascat satellite data and TRMM satellite precipitation data.

Because of a lack of data, not only were statistical properties determined, but a visual inspection of the data was also executed. The latter shows that there is some resemblance between the different *SM* datasets and precipitation, whereas the statistical properties do not always confirm this, due to e.g. delayed reactions. It also show a small range for SMOS, compared to ERS and VUA.

SMOS data seems to respond better to precipitation than other datasets, in the sense that its correlation with precipitation is the second best of all satellite datasets. Correlation with other datasets in wet areas is however poor. This is due to the fact that, although SMOS reacts to precipitation, it does so with smaller changes in *SM* than ERS and VUA. Correlation with other *SM* datasets in terms of statistical coefficients is best for VUA and ERS. ERS shows a relative good visual correlation with EPFL due to its relative high values (compared to the other satellite datasets). Statistical properties however show a poor correlation for all datasets with EPFL. The best correlation is found for NSIDC, which on visual inspection shows a lack or dynamics. Looking

at the complete research area, it seems that the best correlations can be found in the north of the research area or in general in areas with little precipitation. The best median values for SMOS are found with ERS, which is surprising, as ERS is an active satellite and uses a different algorithm than SMOS. Values for the SMOS-VUA comparison, however, do not stay far behind. Interesting are the differences between morning and evening SMOS *SM* values. Without rain, *SM* values in the evening are lower than in the morning, while measured *TB* are not necessarily different.

Concluding, it seems that SMOS is able to show a pattern that resembles reality relatively well for dry areas. However, in wet and more densely vegetated areas, the correlation is lower.

7.3 DISCUSSION

In chapter four, the relative importance of the parameters in the SMOS algorithm is determined for specific ranges. Assuming the used ranges are correct, the most influential parameters for West-Africa are the scattering albedo ω, the roughness *HR*, temperature *T* (from both ground and canopy) and litter parameters, influencing the optical depth τ_c. The exact influence of an erroneous value for these parameters depends however on the measured *TB*, the original value for the parameter and the change the parameter undergoes. Using the default parameters from the L2PP processor, SMOS shows low *SM* values as compared to field measurements in Ghana. Only by changing the *HR* value, a *SM* value higher than 0 was computed that could be compared to the field average. In further research into this area, it is thus very important to get values that are representative for West-Africa. Finding the right value for scattering albedo, roughness and litter properties will require knowledge on the area's characteristics and their effect on the parameters.

The actual value for SMOS computer for comparison with the field average did, after changing *HR*, fall within the 4% error bound as set by ESA. The field work was executed at the end of the dry season. Comparisons in chapter six however show the low correlation of SMOS with other datasets for wet areas, despite a good correlation of SMOS with precipitation data. Visual inspection seem to suggest that the range of SMOS is smaller than that of VUA and ERS. Together, this seems to lead to a *SM* dataset that rises when it is supposed to rise and drops when it is supposed to drop, but has a problem with absolute values. Also, it seems that the right value is easier to achieve in very dry areas.

7.4 ANSWERING THE RESEARCH QUESTION

With the available data, it has to be concluded that, at the end of the dry period, SMOS measurements fall within the 4% range of the real value, provided that the roughness parameter

is given a higher value than its default. Comparing SMOS to EPFL field data does not give any statistical correlation, although visual inspection shows there is some resemblance. SMOS data however has absolute values that are lower than that of EPFL. Good correlation with other satellite imagery suggest that the low correlation with EPFL is partly caused by the fact that the meters do not give a well defined SM average for the complete pixel. Correlation with other satellite datasets is especially good in dry areas however.

It thus seems that the SMOS algorithm does not stay within the 4% bound for West Africa, because its values are too low compared to the real value, especially in wet areas. With the correct values for parameters like surface roughness, litter and scattering albedo determined however, it is expected that the SM value at the very least can fall within the 4% error bounds in the dry periods. Further research will have to show whether these changes in parameter values will also realize a better correlation with other datasets in wet areas.

RECOMMENDATIONS

Following this research, several recommendations can be given for further research:

1. The sensitivity analysis has shown that the values for HR, ω, τ_c and T_c and T_g have a large impact on the SM values calculated. Further research should be done into whether the values for the location analysis as used in chapter four are indeed representative for West-Africa. It is very likely that some values change over the course of a year. More insight in this would enable a better determination of SM. Possibly, ω and τ_c can be determined by satellite imagery. ECMWF has data for T_c and T_g, that could possibly be used. For HR, an empirical formula based on field capacity is suggested, but not yet validated, by ESA.

2. Part of the sensitivity analysis in this report is done with a Matlab code. This code is a simplification of the minimization process executed in the L2PP processor. It is recommended to obtain auxiliary data files with different values for the litter and temperature data. This way, the sensitivity of these parameters can be tested in the processor, so that a better impression of the algorithm for these parameters can be obtained.

3. The sensitivity analysis was done for one specific area. In order to obtain a better understanding of the effect of the different parameters, sensitivity analyses of other areas should be executed. It might for example be interesting to have a look at more densely vegetated areas. For the location analysis, this means the values used might be different than in this research.

4. With default values for parameters that represent the area well, it would be interesting to see whether the correlation for more densely vegetated areas becomes better than in the comparisons done in this report.

5. It would be interesting to do a new field work in the same area as the one in the report during the wet season. That would give insight into how SMOS reacts in the wet season as compared to the dry season.

6. Another subject for future field work could be to determine whether the *HR* value of the area can indeed be determined with the formula as suggested in the ATBD. *HR* in this formula is dependent on the field capacity and the sand and clay fraction of the soil.

7. Finally, field work could give insight into how the EPFL measurements relate to a well defined average for the pixel they are located in. This knowledge would enable a better comparison of SMOS data with EPFL measurements.

8. Validation in chapter six was done with a limited amount of data. It is recommended to do the validation with more data, in order to e.g. determine the correlation between the different datasets over the different seasons.

9. In earlier research, a soil moisture model was created for West-Africa [7.1]. This model is not yet validated. However, comparing SMOS and other data to the output of this model can help obtain a better understanding of their error. With more knowledge on the error of the different datasets, it is also possible to work the other way around and assimilate the *SM* satellite or field data into the model. This is a way of improving the model.

Validation of SMOS satellite data over Ghana and Burkina Faso

REFERENCES

[1.1] Barré, H.M.J.P., Duesmann, B., and Kerr, Y.H., *SMOS: The Mission and the System*, IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no. 3, March 2008, p. 587-593

[1.2] http://www.uwsp.edu/geo/faculty/ritter/geog101/textbook/energy/energy_balance.html, viewed on 04-10-2010

[1.3] McMullan, K.D., Brown, M.A., Martin-Neira, M., Rits, W., Ekholm, S., Marti, J., Lemanczyk, J., *SMOS: The Payload*, IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no. 3, March 2008, p. 594-605

[1.4] CESBIO, IPSL-Service d'Aéronomie, INRA-EPHYSE, Reading University, Tor Vergata University, SMO*S level 2 Processor for Soil Moisture Algorithm Theoretical Based Document (ATBD) Issue 3*, December 2007

[1.5] Saleh, K., Wigneron, J.-P., Rosnay, de, P., Calvet, J.-C., Escorihuela, M. J., Kerr, Y., Waldteufel, P., *Impact of rain interception by vegetation and litter on the L-band emission of natural grasslands (*SMO*SREX Experiment)*, J. Geophys. Res., submitted 2005

[1.6] Holmes, T. R. H., Rosnay, de, P, Jeu, de R., Wigneron, J.-P., Kerr, Y. H., Calvet, J.-C., Escorihuela, M. J., Saleh, K., Lemaitre, F., *A new parameterization of the Effective Temperature for L-band Radiometry*, Geophys. Res. Letters, vol. 33, 2006

[1.7] Wigneron, J.-P., Chanzy, A., Laguerre, L., Kerr, Y., *A simple parameterization of the L-band microwave emission from rough agricultural soils*, IEEE Trans. Geosci. Rem. Sens., vol. 39, pp. 1697-1707, 2001

[1.8] Wigneron, J.-P, Pardé, M., Waldteufel, P., Chanzy, A., Kerr, Y., Schmidl, S., Skou, N., *Characterizing the dependence of vegetation model parameters on crop structure, view angle and polarization at L-Band*, IEEE Trans. Geosci and Remote Sensing, vol. 42, pp. 416-425, 2004

[1.9] Wigneron, J.-P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M.-J., Richaume, P., Ferrazzoli, P., Rosnay, de, P., Gurney, R., Calvet, J.-C., Grant, J. P., Guglielmetti, M., Hornbuckle, B., Mätzler, C., Pellarin, T., Schwank, M., *L-band Microwave Emission of the Biosphere (L-MEB) Model: description and calibration against experimental data sets over crop fields*, Rem. Sens. Environ., vol. 107 pp. 639–655, 2007

[1.10] Friesen, J., Rodgers, C., Oguntunde, P.G., Hendrickx, M.H., Giesen, van de, N., *Hydrotope-Based Protocol to Determine Average Soil Moisture Over Large Areas for Satellite Calibration and Validation With Results From an Observation Campaign in the Volta Basin, West Africa*, IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no.7, p.1995-2004, July 2008

[1.11] Friesen, J., *Spatio-temporal Rainfall Patterns in Northern Ghana*, Geographische Institute der Rheinischen Friedrich-Wilhelms-Universität Bonn, Bonn, June 2002

[2.1] Barry, B., Obuobie, E., Andreini, M., Andah, W., Pluquet, de, M., *The Volta River Basin, Comprehensive Assessment of Water Management in Agriculture, Comparative study of river basin development and management*, draft version, January 2005

[2.2] "FAOCLIM 2: World-Wide Agroclimatic Database," *Environment and Natural Resources Working Paper Number 5*, Food Agriculture Org. United Nations (FAO), Rome, Italy, 2001, Version 2.01.

[2.3] http://www.fao.org/ag/AGP/agpc/doc/Counprof/Ghana/Ghana.htm#2. SOILS AND TOPOGRAPHY (viewed on 15-04-2010)

[2.4] http://www.worldwildlife.org/wildworld/profiles/terrestrial/at/at0722_full.html, viewed on 01-09-2010

[2.5] http://www.fao.org/ag/AGP/AGPC/doc/Counprof/BurkinaFaso/burkinaFeng.htm, viewed on 01-09-2010

[2.6] http://www.uni.edu/gai/Nigeria/Background/Standard4.html, viewed on 04-10-2010

[3.1] McMullan, K.D., Brown, M.A., Martin-Neira, M., Rits, W., Ekholm, S., Marti, J., Lemanczyk, J., *SMOS: The Payload*, IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no. 3, March 2008, p. 594-605

[3.2] ESL & Shengli Dai, Mahmoodi, A., Graham, A., *Table Generation Requirement Document (TGRD) for the SMOS Level 2 Soil Moisture Prototype Processor Development (SMPPD), issue 0.3*, CBSA, UoR, TV, INRA, and ARRAY, March 2006

[3.3] Marquardt, D.W., *An algorithm for least-squares estimation of non linear parameters*, J. Soc. Ind. Appl. Math, vol. 11, p. 431-441, 1963

[3.4] Holmes, T. R. H., Rosnay, de, P, Jeu, de R., Wigneron, J.-P., Kerr, Y. H., Calvet, J.-C., Escorihuela, M. J., Saleh, K., Lemaitre, F., *A new parameterization of the Effective Temperature for L-band Radiometry*, Geophys. Res. Letters, vol. 33, 2006

[3.5] Wigneron, J.-P., Laguerre, L., Kerr, Y.H., *A Simple Parameterization of the L-Band Microwave Emission from Rough Agricultural Soils*, IEEE Transactions on Geoscience and Remote Sensing, vol. 39, no.8, p.1697-1707, August 2001

[3.6] Wigneron, J.-P., Pardé, M., Waldteufel, P., Chanzy, A., Kerr, Y., Schmidl, S., Skou, N., *Characterizing the dependence of vegetation model parameters on crop structure, view angle and polarization at L-Band*, IEEE Trans. Geoscience and Remote Sensing, vol. 42, pp. 416-425, 2004

[3.7] Wigneron, J.-P., Pellarin, T., Calvet, J.-C., Rosnay, de, P., Saleh, K., Kerr, Y., *L-MEB: A simple model at Lband for the continental areas - Application to the simulation of a half-degree resolution and global scale data set*, Radiative Transfer Models for Microwave Radiometry, C. Mätzler, Ed. Stevenage, UK: Institution of Electrical Engineers, 2005

[3.8] Wigneron, J.-P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M.-J., Richaume, P., Ferrazzoli, P., Rosnay, de, P., Gurney, R., Calvet, J.-C., Grant, J. P., Guglielmetti, M., Hornbuckle,

B., Mätzler, C., Pellarin, T., Schwank, M., *L-band Microwave Emission of the Biosphere (L-MEB) Model: description and calibration against experimental data sets over crop fields*, Rem. Sens. Environ., vol. 107 p. 639–655, 2007

[4.1] http://www.brockmann-consult.de/beam-wiki/display/SBOX/SMOS+Toolbox+for+BEAM;jsessionid=267EC27448B3CA6724923863868C76C7, viewed on 30-6-2010

[4.2] CESBIO, IPSL-Service d'Aéronomie, INRA-EPHYSE, Reading University, Tor Vergata University, SMO*S level 2 Processor for Soil Moisture Algorithm Theoretical Based Document (ATBD) Issue 3*, December 2007

[4.3] Wigneron, J.-P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M.-J., Richaume, P., Ferrazzoli, P., Rosnay, de, P., Gurney, R., Calvet, J.-C., Grant, J. P., Guglielmetti, M., Hornbuckle, B., Mätzler, C., Pellarin, T., Schwank, M., *L-band Microwave Emission of the Biosphere (L-MEB) Model: description and calibration against experimental data sets over crop fields*, Rem. Sens. Environ., vol. 107 p. 639–655, 2007

[4.4] Wigneron, J.-P., Pardé, M., Waldteufel, P., Chanzy, A., Kerr, Y., Schmidl, S., Skou, N., *Characterizing the dependence of vegetation model parameters on crop structure, view angle and polarization at L-Band*, IEEE Trans. Geosci and Remote Sensing, vol. 42, p. 416-425, 2004

[4.5] Wigneron, J.-P., Pellarin, T., Calvet, J.-C., Rosnay, de, P., Saleh, K., Kerr, Y., *L-MEB: A simple model at Lband for the continental areas - Application to the simulation of a half-degree resolution and global scale data set*, Radiative Transfer Models for Microwave Radiometry, Mätzler, C., Stevenage, Ed., UK: Institution of Electrical Engineers, 2005

[4.6] Della Vecchia, A., Ferrazzoli, P., *A large scale approach to estimate L band emission from forest covered surfaces*, Tor Vergata University, ESA Technical note SO-TN-TV-GS-0001-01.a, 21/02/2006 2006

[4.7] http://www.britannica.com/EBchecked/topic/357237/magnetic-permeability, viewed on 31-08-2010

[4.8] Hendrickx, J.M.H., Bruce, J., Harrison, J., Dam, van, R.L., Borchers, B., Norman, D.I., Dedzoe, C.D., Antwi, B.O., Asiamah, R.D., Rodgers, C., Vlek, P., Friesen, J., *Magnetic soil properties in Ghana*

[4.9] Wigneron, J.-P., Chanzy, A., Laguerre, L., Kerr, Y., *A simple parameterization of the L-band microwave emission from rough agricultural soils*, IEEE Trans. Geosci. Rem. Sens., vol. 39, p. 1697-1707, 2001

[4.10] Saleh, K., Wigneron, J.-P., Rosnay, de, P., Calvet, J.-C., Escorihuela, M. J., Kerr, Y., Waldteufel, P., *Impact of rain interception by vegetation and litter on the L-band emission of natural grasslands (SMOSREX Experiment)*, J. Geophys. Res., submitted 2005

[4.11] Wigneron, J.-P., Chanzy, A., Calvet, J.-C., Bruguier, N., *A simple algorithm to retrieve soil moisture and vegetation biomass using passive microwave measurements over crop fields*, Remote sens. environ., vol. 51, p. 331-341, 1995

[5.1]Friesen, J., Rodgers, C., Oguntunde, P.G., Hendrickx, M.H., Giesen, van de, N., *Hydrotope-Based Protocol to Determine Average Soil Moisture Over Large Areas for Satellite Calibration and Validation With Results From an Observation Campaign in the Volta Basin, West Africa*, IEEE Transactions on Geoscience and Remote Sensing, vol. 46, no.7, p.1995-2004, July 2008

[5.2] Park, S.J., Van de Giesen, N., *Soil-landscape delineation to define spatial sampling domains for hillslope hydrology*, Journal of Hydrology 295, pages 28-46, 2003

[5.3] Western, A.W., Blöschl, G., *On the spatial scaling of soil moisture*, Journal of Hydrology 217, pages 203–224, 1999

[5.4] Chanzy, A., Bruckler, L, Perrier, A., *Soil evaporation monitoring: a possible synergism of microwave and infrared remote sensing*, Journal of Hydrology 165, p.235-259, 1995

[5.5] Schmugge, T.J., Jackson, T.J., McKim, H.L.,, *Survey of Methods for Soil Moisture Determination*, Water Resource Research, vol. 16, no. 6, pages 961-979, December 1980

[5.6] Reynolds, S.G., *The gravimetric method of soil moisture determination part I, a study of equipment, and methodological problems*, Journal of Hydrology 11, pages 258-273, 1970

[5.7] Agyare, W.A., *Ecology and Development Series No. 17, 2004; Soil characterization and modeling of spatial distribution of saturated hydraulic conductivity at two sites in the Volta Basin of Ghana*, Cuvillier Verlag Göttingen, 2004

[5.8] CESBIO, IPSL-Service d'Aéronomie, INRA-EPHYSE, Reading University, Tor Vergata University, SMO*S level 2 Processor for Soil Moisture Algorithm Theoretical Based Document (ATBD) Issue 3*, December 2007

[6.1] Rudiger, C., Calvet, J.-C., Gruhier, C., Holmes, T., Jeu, de, R., Wagner, W., *An Intercomparison of ERS-Scat and AMSRE Soil Moisture Observations with Model Simulations over France*, Journal of Hydrometeorology

[6.2] http://nsidc.org/data/ae_land3.html, viewed on 30-7-2010

[6.3] http://geoservices.falw.vu.nl/amsr_soil_moisture_description.html, , viewed on 30-7-2010

[6.4] http://oiswww.eumetsat.org/WEBOPS/eps-pg/ASCAT/ASCAT-PG-4ProdOverview.htm, viewed on 30-7-2010

[6.5] http://trmm.gsfc.nasa.gov/3b42.html, viewed on 15-08-2010

[6.6] http://www.ipf.tuwien.ac.at/radar/index.php?go=ascat, viewed on 15-08-2010

[7.1] Smits, J., Assimilation of remote sensed soil moisture data into a hydrological model using the Ensemble Kalman Filter, A West Africa case study, TU Delft, May 2008

[A.1] Dobson, M. C., Ulaby, F. T., Hallikainen, M. T., Elrayes, M. A., *Microwave Dielectric Behavior of Wet Soil .2. Dielectric Mixing Models*, Ieee Transactions on Geoscience and Remote Sensing, vol. 23, pp. 35-46, 1985.

[A.2] Lane, J.A., Saxton, J.A., *Dielectric dispersion in pure polar liquids at very high radio frequencies, measurements on water, methyl and ethyl alcohols*, Proc. Royal Soc., vol. A213, pp. 400–408, 1952

[A.3] Wigneron, J.-P., Chanzy, A., Laguerre, L., Kerr, Y., *A simple parameterization of the L-band microwave emission from rough agricultural soils*, IEEE Trans. Geosci. Rem. Sens., vol. 39, pp 1697-1707, 2001

[A.4] http://www.terragis.bees.unsw.edu.au/terraGIS_soil/sp_water-soil_moisture_classification.html, viewed on 28-03-2010

[A.5] Jackson, T.J., Schmugge, T.J., *Vegetation Effects on the Microwave emission of soils*, Remote Sensing of Environment, vol. 36, pp. 203-212, 1991

[A.6] Wigneron, J.-P., Chanzy, A., Calvet J.-C., Bruguier, N., *A simple algorithm to retrieve soil moisture and vegetation biomass using passive microwave measurements over crop fields*, Remote sens. environ., vol. 51, pp. 331-341, 1995

[A.7] Wigneron, J.-P., Pellarin, T., Calvet, J.-C., Rosnay, de, P., Saleh, K., Kerr, Y., *L-MEB: A simple model at L-band for the continental areas - Application to the simulation of a half-degree resolution and global scale data set*, Radiative Transfer Models for Microwave Radiometry, C. Mätzler, Ed. Stevenage, UK: Institution of Electrical Engineers, 2005

[A.8] Wigneron J.-P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M.-J., Richaume, P., Ferrazzoli, P., Rosnay, de, P., Gurney, R., Calvet, J.-C., Grant, J.P., Guglielmetti, M., Hornbuckle, B., Mätzler, C., Pellarin, T., and Schwank, M., *L-band Microwave Emission of the Biosphere (L-MEB) Model: description and calibration against experimental data sets over crop fields*, Rem. Sens. Environ., vol. 107 pp. 639–655, 2007

[A.9] Wigneron J.-P., Calvet, J.-C., Rosnay, de, P, Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M.J., Kruszewski, A., *Soil moisture retrievals from biangular L-band passive microwave observations*, Geoscience and Remote Sensing Letters, IEEE, vol. 1, pp. 277-281, 2004

[A.10] Wigneron, J.-P., Chanzy, A., Calvet, J.-C., Olioso, A., and Kerr, Y., *Modeling approaches to assimilating L-Band passive microwave observations over land surfaces*, J. Geophys. Res., vol. 107, pp. DOI10.1029/2001JD000958, 2002

[A.11] Saleh, K., Wigneron, J.-P., Rosnay, de, P., Calvet, J.-C., Escorihuela, M.J., Kerr, Y., Waldteufel, P., *Impact of rain interception by vegetation and litter on the L-band emission of natural grasslands (SMOSREX Experiment)*, J. Geophys. Res., submitted 2005

[A.12] Saleh, K., Wigneron, J.-P., Waldteufel, P., Rosnay, de, P., Schwank, M., Calvet, J.-C., Kerr, Y., *Estimates of surface soil moisture under grass covers using L-band radiometry*, Rem. Sens. Environ., vol. 109, pp. 42-53, 2007

[A.13] Della Vecchia, A., Ferrazzoli, P., *A large scale approach to estimate L band emission from forest covered surfaces*, Tor Vergata University, ESA Technical note SO-TN-TV-GS-0001-01.a, 21/02/2006 2006

[A.14] Waldteufel, P., Vergely, J.-L., and Cot, C., *A cardioid model for multi-angular radiometric observations*, IEEE Trans Geosci.Remote Sens., vol. 42, pp. 1059-1063, 2004

[A.15] Klein, L.A., Swift, C.T., *An improved model for the dielectric constant of sea water at microwave frequencies*, IEEE Trans Antennas and Propagation, vol. AP-25, pp. 104-111, 1977

[C.1] Mahmoodi, A., Wong, C., *L2 SM Open Prototype Software User Manual (SUM) for the SMOS Level 2 Soil Moisture Processor Development Continuation Project*, Array, 2008

[C.2] Wong, C., *L2 SM Open Prototype Software Release Document (SRD) for the SMOS Level 2 Soil Moisture Processor Development Continuation Project, Software Release 03.07*, Array, 2010

[D.1] Agyare, W.A., *Soil characterization and modeling of spatial distribution of saturated hydraulic conductivity at two sites in the Volta Basin of Ghana*, Ecology and Development Series No. 17, Cuvillier Verlag Göttingen, 2004

APPENDICES

A FORMULAS FOR REFLECTIVITY, DIELECTRIC CONSTANT AND VEGETATION PARAMETERS

INTRODUCTION

This appendix gives the most important formulas used in the SMOS algorithm to obtain the reflectivity of several surface types with an explanation of the parameters. Whenever a parameter is computed through a different formula, this formula is given below the explained parameters. Most information comes directly from the ATBD.

There are three completely defined models at this point, the Nominal Model (MN), the water model (MW) and the cardioid model (MD). Figure A.1 shows how the different formulas work schematically and the parameters for which the formulas are given in this appendix. For all models, the dielectric constant ε is determined. The way this parameter is determined depends on the model. Besides that, there are some cases, such as a rocky soil or ice, for which ε is determined with very simple formulas. These are also given in this appendix. For the nominal model, it is also shown how ε is connected to soil emissivity (through reflectivity) and thus to *TB*. Furthermore, the formulas for single scattering albedo and the optical depth, used to model vegetation, are described.
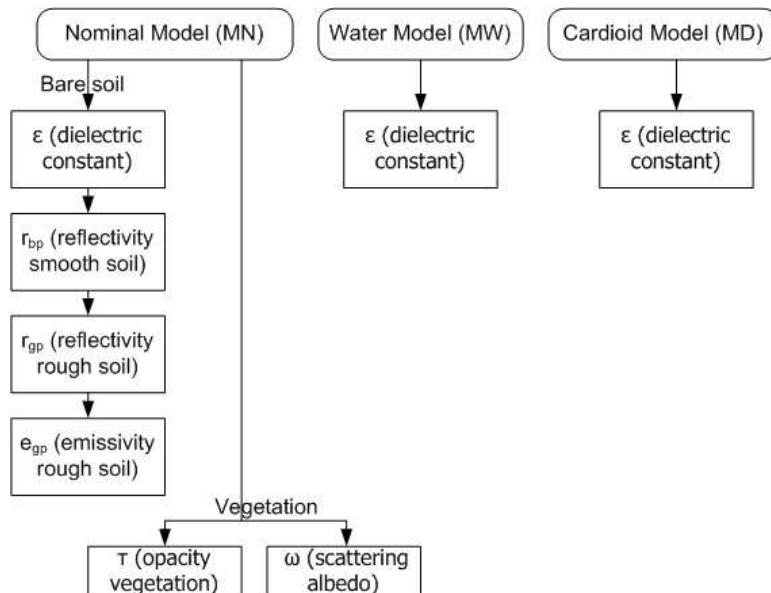

Figure A.1: The three defined models of the algorithm with the parameters described in this appendix

In the formula for brightness temperature, the terms that depend on the fractions are called R1 and R2. The parameters in these terms are the reflectivity r, the emissivity e and the optical depth of the canopy τ_c.

For τ_c, a general formula is $\tau_c = \int_0^{LHsec(\theta)} \kappa_{ext} dx$, with LH the layer height of the canopy, θ is the incidence angle of the view and κ_{ext} is a scattering coefficient, a measure for light extinction. For vegetation characterised by cylindrical features, scattering is non-isotropic. This means the optical depth can be replaced by the effective optical depth, $\tau_c^* = (1-a_{sct}\omega)\tau_c$. In this case, a_{sct} is the canopy structure coefficient that characterizes the proportion of radiation scattered in forward direction and ω the isotropic single scattering albedo. These general formulas have been adapted for forest and low vegetation.

For nominal cases (bare soil, forest and low vegetation) it is possible to calculate the reflectivity *r* and through that the emissivity *e* (because *e=1-r*) with the nominal model. The other models are first and foremost used to calculate the dielectric constant of an area.

NOMINAL MODEL (MN)

In this model, the reflectivity (for a smooth surface) is written as a function of the complex dielectric constant $\varepsilon = \varepsilon' - j\varepsilon''$ and *SM*. There are three nominal cases that can be calculated with the nominal model: bare soil, low vegetation and forest. There are also formulas for the dielectric constant for non-nominal cases.

BARE SOIL

For a SMOS pixel composed of several surface types, the equation for the specular reflectivity H and V for a smooth air-soil r_{bp} is:

$$r_{bH}(\theta) = \left| \frac{\mu_s cos(\theta) - \sqrt{\mu_s \varepsilon_b - sin^2(\theta)}}{\mu_s cos(\theta) + \sqrt{\mu_s \varepsilon_b - sin^2(\theta)}} \right|^2 \qquad r_{bV}(\theta) = \left| \frac{\varepsilon_b cos(\theta) - \sqrt{\mu_s \varepsilon_b - sin^2(\theta)}}{\varepsilon_b cos(\theta) + \sqrt{\mu_s \varepsilon_b - sin^2(\theta)}} \right|^2 \qquad \text{(eq. A1a and b)}$$

Parameter	Name	Range	Units	Default	Source	Remark
ε_{bS}	Bare soil dielectric constant		F/m			To be computed
θ	Incidence angle	0-1.25	rad		SMOS	Depends on view
μ_s	Soil magnetic permeability			1	Field	Expected to be unity

$$\varepsilon_b = \left(1 + \frac{\rho_b}{\rho_s}\left(\varepsilon_{pa}^a - 1\right) + SM^{\beta'}\varepsilon_{fw}'^a - SM \right)^{\frac{1}{a}} - j\left(SM^{\beta''}\varepsilon_{fw}''^a\right)^{\frac{1}{a}} \qquad \text{(eq. A2)}$$

Parameter	Name	Range	Units	Default	Source	Remark
ρ_b	Dry bulk soil density	0.5-2.5	g/cm3	1.3	Field	
ρ_s	Soil particle density	2 - 3	g/cm3	2.664	Field	
ε_{pa}	Dielectric constant of solid particles			4.7		To be computed
a	Shape factor, depending on amount of free and bound water			0.65	[A.1]	Determined with measured dielectric constants over all used frequencies and soils
$\beta = \beta' - j\beta''$	(Empirically-derived complex function of) soil texture parameter					To be computed, gives the influence of the soil texture on the dielectric constant.
SM	Soil moisture	0-0.5	m3/m3		Field	
Sal	Soil salinity	0-123	ppt		Field	Not used currently
F	Frequency	1.4	Ghz			
$\varepsilon_{fw} = \varepsilon'_{fw} - j\varepsilon''_{fw}$	Dielectric constant of free water					To be computed, for this case adapted to the specific case of soil by approximating the effective conductivity of water using the Stern-Gouy double layer theory

$$\varepsilon_{pa} = (CPA_1 + CPA_2 *\rho_s) + CPA_3 \qquad \text{(for 'normal' soils)} \qquad \text{(eq. A3)}$$

Parameter	Name	Range	Units	Default	Source	Remark
CPA_1	Coefficients for computation of dielectric constant of solid particles		$[F/m]^{1/2}$	1.01	UPF [A.1]	These parameters represent the empirical fitting of measured data for soils with extremely low moisture content
CPA_2			$[F\ m^2/g]^{1/2}$	0.44		
CPA_n			$[F/m]$	-0.062		

$$\beta' = BERE_1 + BERE_2 S + BERE_3 C \qquad \text{(eq. A4a and b)}$$
$$\beta'' = BEIM_1 + BEIM_2 S + BEIM_3 C$$

Parameter	Name	Range	Units	Default	Source	Remark
S	Sand fraction	0-100	%		ECOCLIMAP	
C	Clay fraction	0-100	%			
$BERE_1$	Parameters to calculate soil texture parameter β	1.2748	-		TGRD [A.1]	Values determined through fitting with measurements
$BERE_2$		-0.519	-			
$BERE_3$		-0.152	-			
$BEIM_1$		1.338	-		TGRD [A.1]	
$BEIM_2$		-0.603	-			
$BEIM_3$		-0.166	-			

$$\varepsilon'_{fw} = \varepsilon_{w\infty} + \frac{\varepsilon_{w0} - \varepsilon_{w\infty}}{1 + (2\pi f \tau_w)^2}$$

$$\varepsilon''_{fw} = \frac{2\pi f \tau_w (\varepsilon_{w0} - \varepsilon_{w\infty})}{1 + (2\pi f \tau_w)^2} + \frac{\sigma_{eff}}{2\pi f \varepsilon_0} \frac{\rho_s - \rho_b}{\rho_s SM}$$

(eq. A5a and b)

Parameter	Name	Range	Units	Default	Source	Remark
ε_{w0}	Static dielectric constant of water ε_{w0}					See water model
$\varepsilon_{w\infty}$	High frequency limit of the dielectric constant of water ε_{fw}			4.9	[A.2]	
F	Mean SMOS frequency	Hz				
τ_w	Relaxation time of water					See water model
ε_0	Permittivity of free space		Fm^{-1}	8.854×10^{-12}		Constant (vacuum permittivity)
σ_{eff}	Effective conductivity					To be computed

$$\sigma_{eff} = SGEF_1 + SGEF_2 \rho_b + SGEF_3 S + SGEF_4 C$$

(eq. A6)

Parameter	Name	Range	Units	Default	Source	Remark
$SGEF_1$	Parameters for soil texture				TGRD [A.1]	Values determined through measurements
$SGEF_2$						
$SGEF_3$						
$SGEF_4$						

Most surfaces are not completely smooth. In that case, we cannot use the bare soil reflectivity, but need to determine the ground reflectivity r_{gp}:

$$r_{gp}(\theta) = \left((1-QR)r_{bp} + QR r_{bp} \right) exp\left(-HR(SM) cos^{NR_P}(\theta) \right)$$

(eq. A7)

Parameter	Name	Range	Units	Default	Source	Remark
QR	(Semi-empirical) H/V polarization coupling factor	0.0-0.5		0	[A.3]	At L-band, models polarization-mixing effect
HR	(Semi-empirical) surface roughness parameter	2 - 3	-			To be computed, models intensity of roughness effects
NR$_P$	Power law of cos(θ)	0-0.5		0	[A.3]	

HR depends on *SM*. It has a minimum value of 0 and a maximum value of 3. The minimum value belongs to the amount of *SM* that belongs to field capacity *FC*. The maximum value belongs to a transition moisture point *XMVT*. This point is calculated through the wilting point. Both the wilting point and the field capacity depend on the sand and clay fraction of a soil. *XMVT* is determined as follows:

WP(C,S)=CWP1+CWP2*S+CPW3*C   (eq. A8a and b)
XMVT(C,S)=CXMVT1*WP(C,S)+CXMVT2

Parameter	Name	Range	Units	Default	Source	Remark
CWP1,2,3				0.06774, -0.00064, 0.00478	TGRD UPF	
CXMVT1,2				0.49, 0.165	TGRD UPF	

The determination of the field capacity is done with Figure A.2:
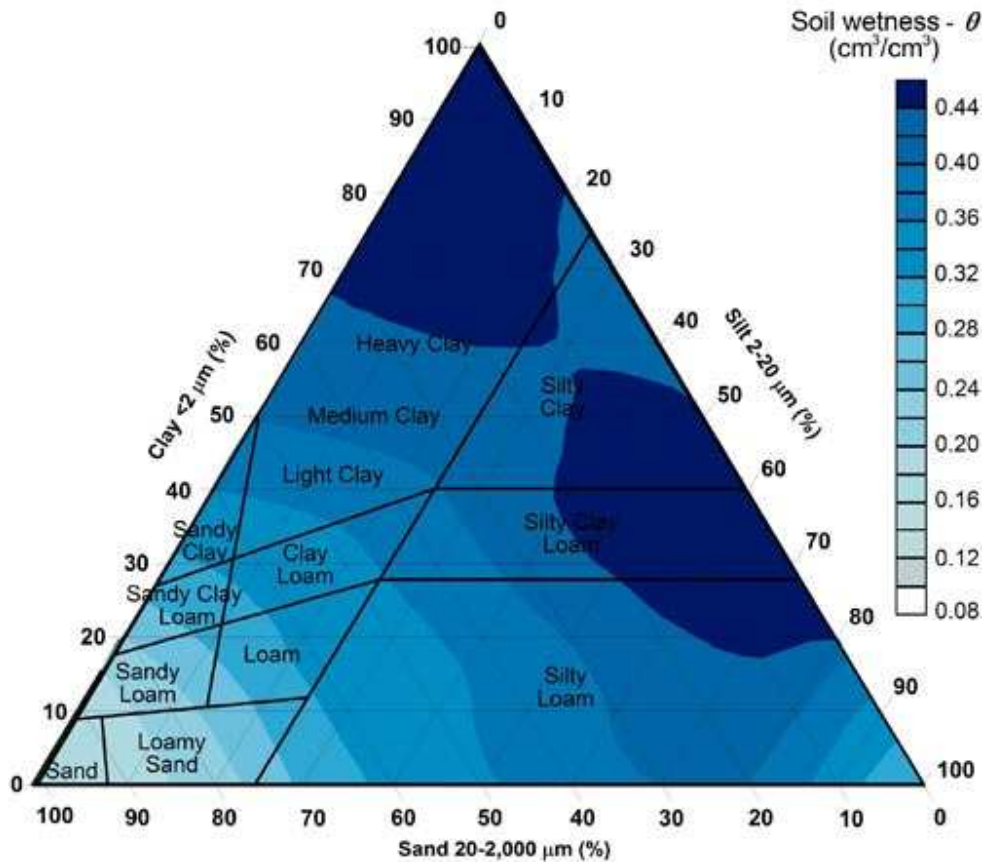
Figure A.2: Triangle for determination field capacity [A.4]

VEGETATION

For vegetation, not only the reflectivity, but also the opacity τ and the scattering albedo ω are necessary to calculate R1 and R2. The formulas are however different for different sorts of vegetation.

There are several formulas for τ. More general formulas are:

$\tau_p = b_p \, VWC$ and $\tau_p / LAI = 0.06 - 0.08$ (eq. A9a and b)

Parameter	Name	Range	Units	Default	Source	Remark
b_p	Vegetation parameter			0.12 +/- 0.03 (most agricultural crops)	[A.5]	0.4 obtained over prairies, probably because of litter
VWC	total vegetation water content				Field	VWC is for the algorithm replaced by LAI, because LAI is easier to determine

Besides that, $\tau_p = \tau_{NAD}$ x function(θ, p) is always true. τ_p depends on the incidence angle and one can calculate:

$\tau_H (\theta) = \tau_{NAD}$ and $\tau_V (\theta) = \tau_{NAD} [\cos^2 \theta + C_{pol} \sin^2 \theta]$ (eq. A10a and b)

Parameter	Name	Range	Units	Default	Source	Remark
C_{pol}	polarization correction factor				[A.6]	Cpol > 1 for a vertical structure

LOW VEGETATION (GRASSLAND, CROP)

Within a large-scale SMOS scene, it is likely that the effects due to the vegetation structure for a variety of vegetation types are averaged, so that the dependence of τ_p (and ω_p) on polarization and incidence angle can be neglected over most pixels. In that case, the following generalization can be used (for crops with a vertical structure):

$\tau_P = \tau_{SP} + \tau_L + \tau_{IP}$ (eq. A11)

Parameter	Name	Range	Units	Default	Source	Remark
τ_S, τ_{S_NAD}	optical depths of the standing vegetation canopy (all the canopy but excluding the litter and the intercepted water)	0 - 3				To be computed. Correlated to VWC and LAI
τ_L	optical depth of the litter layer	0 - 3				To be computed
τ_I	increase in optical depth due to intercepted water in the standing vegetation canopy	0 - 3				Flagged, not taken into account

$\tau_{SH}(\theta=0) = \tau_{SV}(\theta=0) = \tau_{S\_NAD} = b'_S . LAI + b''_S$
$\tau_{SH}(\theta) = \tau_{S\_NAD} (\sin^2(\theta).tt_H + \cos^2(\theta))$ (eq. A12a, b and c)
$\tau_{SV}(\theta) = \tau_{S\_NAD} (\sin^2(\theta).tt_V + \cos^2(\theta))$

Parameter	Name	Range	Units	Default	Source	Remark
b'_S	parameter of the relation τ_S / LAI (for $\theta=0$)	0.01-1	m2/m2		[A.7], [A.8]	Function of canopy type
b''_S		0. - 3	m2/m2			
tt_H	angular correction parameter at H polarization (accounting for the dependence of τ_{SP} on incidence angle)	0.1 – 15		1	[A.9]	Function of canopy type
R_{tt} $(=tt_V/tt_H)$	ratio of angular correction parameters	0.05-20		1		Function of canopy type
LAI	Leaf Area Index	0-10	m2/m2	See [A7]	Satellite images or SVAT modeling (see [A.10])	

$$\tau_L = c_L \, LWC \tag{eq. A13}$$

Parameter	Name	Range	Units	Default	Source	Remark
LWC	Water content of the litter	0-50	kg/m2			To be computed
c_L	attenuation coefficient of litter($c = \tau_L$ / LWC)	0.01-1	m2/kg	0.24	[A.11]	Function of canopy type

LWC and thus τ_L are dependent on SM and canopy type

$$LWC = [Mg\_L / (1-Mg\_L)] \, B_S\_L \tag{eq. A14}$$

Parameter	Name	Range	Units	Default	Source	Remark
B_S_L	dry biomass of litter	0-50	kg/m2	0.3	SMOSREX	
Mg_L	moisture content of litter	0-0.9	kg/kg			To be computed

$$Mg\_L = a\_L \,.\, SM + b\_L \tag{eq. A15}$$

Parameter	Name	Range	Units	Default	Source	Remark
a_L	parameters used to compute the litter moisture content (Mg_L) from soil moisture SM	0 – 10		2.33	SMOSREX	
b_L		0, 1		0		Default corresponding to litter moisture content of 70% for a soil moisture SM = 0.3 m3/m3

The single scattering albedo is dependent on its polarization. The polarization dependent albedos are used in the term *R2*.

Parameter	Name	Range	Units	Default	Source	Remark
ω_H	scattering albedo at H polarization	0.- 0.2		0	[A.5]	Can be non-zero, especially for grasses [A.12]
$\omega_V - \omega_H$	difference of albedo at H and V polarization	-0.2-0.2		0		

FOREST

For forest, the opacity can be calculated with the following formula:

$$\tau_{F\_NAD} = b'_F \cdot LAI\_max + b''_F \qquad \text{(eq. A16)}$$

Contributions to τ_{NAD} are contribution of crown (τ_1), contribution of litter (τ_2) and contribution of understory (τ_3). This leads to:

$$\tau_{FNAD} = \tau_1 + \tau_2 + \tau_3 = 2 * (b_{F1}' * 0.6) LAI\_max + 0.06 * 0.4 LAI\_max \qquad \text{(eq. A17)}$$

Parameter	Name	Range	Units	Default	Source	Remark
b'_F	parameter of the relation τ_{FNAD} / LAI_max	0.01-1	m2/m2	deciduous broadleaf, evergreen broadleaf, woodland 0.295; needle leaf 0.337; mixed forest 0.31	[A.13]	
b''_F		-1 - 1	m2/m2	0		
LAI_max	Maximum yearly value of Leaf Area Index	0-10	m2/m2			Assumed constant during SMOS lifetime

NON-NOMINAL CASES

For several non-nominal cases, that is cases of which the brightness temperature cannot be computed so easily, suggested formulas for the dielectric constant are given.

Dry sand:
$$\varepsilon_{dry-sand} \approx 2.53 - 0.05i \qquad \text{(eq. A18)}$$

Rocks and rocky out crops:
$$\varepsilon_{rock} = 5.7 - 0.074i \qquad \text{(eq. A19)}$$

Frozen soils:
$$\varepsilon_{frz} = 5 - 0.5i \qquad \text{(eq. A20)}$$

Ice:
$$\varepsilon_{ice} \approx 3.17 - \varepsilon_{ice}''i \qquad \text{(eq. A21)}$$

CARDIOID MODEL (MD)

This model is used whenever surface reflectivity cannot be written as a function of the *SM*. This is the case for example for iced surfaces. The assumption is however that the surface still behaves as a dielectric reflecting medium. The dielectric constant can however not be determined directly from the measured *TB*, the errors become too great. However, the root mean square error of the difference between measured and calculated brightness temperature divided by squared assumed uncertainty forms a cardioid. This can be modelled with a model that depends on three parameters [A.14]. This gives the formulas for the real and imaginary part of the dielectric constant:

ε'=A_card(1+cos(U_card))cos(U_card)+B_card
ε''=A_card(1+cos(U_card))cos(U_card)

(eq. A22a and b)

Parameter	Name	Range	Units	Default	Source	Remark
A_card	Dielectric constant index				[A.14]	Can be retrieved from angular dependent radiometric data
U_card	Polar angle					Emissivity is almost independent of this parameter
B_card	Corrective term					This is a constant in order to fit the model with reality, the optimal value is 0.8

This can be rewritten as:

A_card=m_card^2+(m_card+ε'-B_card)
U_card=tan^{-1}(ε''/(ε'-B_card))
with m_card=((ε'-B_card)^2+ε''^2)$^{1/2}$

(eq. A23a, b and c)

With the SMOS data and auxiliary information (with which the dielectric constant can be determined), it is possible to estimate *A_card* and *U_card*. With additional information, these can then be used for the determination of the dielectric constant of the area.

RETRIEVAL VERSUS DIRECT SIMULATION

The model works slightly different for retrieval and direct simulation. To calculate fixed contributions, the obtained values for ε' and ε'' (for the different fractions) are used directly. However, for non-nominal fractions (e.g. rocks, dry soils etc.) or group of fractions over which retrieval is carried out, the following steps are taken:

- Build fraction weighted averages of ε' and ε''
- Obtain the a priori value for *A_card* and the reference value for *U_card* with eq. A27
- In the forward model for iterative retrieval, *A_card* is floating; ε' and ε'' are computed from *A_card*, *U_card* and the constant *B_card* using eq. A26. Following steps (surface

roughness, vegetation layer) are carried out in the usual way. Depending on the decision tree, the vegetation optical depth is either a fixed or a floating parameter

- Using the retrieved *A_card* and the constant *U_card* and *B_card*, ε' and ε'' can be computed and should appear in the output product

WATER MODEL

For open water, the difference is in the determination of the dielectric constant.

The complex dielectric constant for free water at a given radiometer frequency f is $\varepsilon_w = \varepsilon'_w - j\,\varepsilon''_w$, with:

$$\varepsilon'_{fw} = \varepsilon_{w\yen} + \frac{\varepsilon_{w0} - \varepsilon_{w\yen}}{1 + (2\pi rT_w f)^2}$$

$$\varepsilon''_{fw} = \frac{2\pi rT_w f(\varepsilon_{w0} - \varepsilon_{w\yen})}{1 + (2\pi rT_w f)^2} + \frac{\sigma_i}{2\pi\varepsilon_0 f}$$

(eq. A24a and b)

Parameter	Name	Range	Units	Default	Source	Remark
ε_{w0}	static dielectric constant of water ε_{w0}					To be computed
$\varepsilon_{w\infty}$	high frequency limit of the dielectric constant of water ε_{fw}			4.9	[A.2]	
F	mean SMOS frequency	Hz				
rT_w	relaxation time of water					To be computed
ε_0	permittivity of free space		Fm^{-1}	8.854 10^{-12}		
σ_i	Ionic conductivity saline water		(S/m)			To be computed

The static dielectric constant, ε_{w0}, depends on whether the water is fresh or saline.

For fresh water, the equation is:
$$\varepsilon_{w0}(T) = ow_1 + ow_2 T + ow_3 T^2 + ow_4 T^3$$
(eq. A25)

For saline water:
$$\varepsilon_{sw0}(S,T) = \varepsilon_{sw0}(0,T) * a_{ST}(S,T)$$
$$\varepsilon_{sw0}(T,0) = ow_5 + ow_6 T + ow_7 T^2 + ow_8 T^3$$
$$a_{ST}(S,T) = ow_9 + ow_{10} * S * T + ow_{11} S^2 + ow_{12} S^3$$
(eq. A26a, b and c)

The relaxation time, $2\pi rT_w$, also depends on the salinity of the water.

For pure water:
$$2\pi rT_w(T) = ow_{14} + ow_{15} T + ow_{16} T^2 + ow_{17} T^3$$
(eq. A27)

For saline water:

$2\pi r\tau_{sw}(T) = 2\pi r\tau_w(T)*b_{ST}(S,T)$ (eq. A28a and b)

$b_{ST}(S,T) = ow_{18} + ow_{19}*S*T + ow_{20}S^2 + ow_{22}S^3$

σ_i is the ionic conductivity for saline water (in S/m) function of temperature and salinity:

$\sigma_i(S,T) = \sigma_i(25,S) \times e^{-\varphi(S,\Delta)}$ (eq. A29)

Parameter	Name	Range	Units	Default	Source	Remark
$\sigma_i(25,S)$	the ionic conductivity of sea water at 25°C					
T	Temperature water				Satellite images	
S	Salinity water				ECOCLIMAP	0 for pure water
φ	Function depending on S and $\Delta = 25 - T$					To be computed

For pure water, $\sigma_i(0,T) = 0$

The ionic conductivity of sea water at 25°C then becomes:

$\sigma_i(25,S) = S*(ow_{23} + ow_{24}S + ow_{25}S^2 + ow_{26}S^3)$ (eq. A30)

Φ is computed as follows:

$\varphi(\Delta,S) = \Delta*(ow_{27} + ow_{28}\Delta + ow_{29}\Delta^2 - S(ow_{30} + ow_{31}\Delta + ow_{32}\Delta^2))$ (eq. A31)

All *OW* Coefficients are based on regression. The values are supplied in TGRD UPF. They are obtained from Klein and Swift [A.15].

B VIEWS, WEIGHING FUNCTION, DECISION TREE AND COST-FUNCTION

INTRODUCTION

This paragraph gives more detailed information on some aspects of the decision tree are mentioned in chapter 3, but not elaborated on. The appendix starts with the explanation on L1c views and the elimination of these views for further use. Elimination is done for single views (for dual polarization, either H or V polarized) or, in the case of Radio Frequency Interference (RFI) for paired views (H and V polarization views that follow each other). Only with enough views, *SM* for an L1c pixel is determined. For this, a working area around each L1c pixel, the areas of the fractions in this working area and thus the weighing function of the different fractions have to be determined. The weighing function can change due to change in environment, e.g. snow. With the weighing function known, it is possible to determine the retrieval case and with that the model (nominal, water, cardioid) that is used to determine the contribution of a certain area. The appendix ends with an explanation on the minimization (in this appendix called retrieval) works and *SM* is determined.

ELIMINATION OF L1C VIEWS

L1c views give among other information on the brightness temperature as calculated from the measured radiance. For a given Discrete Global Grid (DGG, the SMOS grid) pixel, there are a number of available views, M_AVA0. Not all these views can be used for the final retrieval. First of all, the views have to meet a spatial resolution criterium. L1c views are eliminated for two reasons. The first is when the area of the half maximum contour of the WEF (weighing function, see below) is bigger than 55x55 km^2. This constrain is called TH_SIZE. TH stands for threshold. The second reason is that the elongation (major axis to small axis ratio) of this contour is bigger than 1.5 (TH_ELON):

$$\sqrt{4 \times axis_1 \times axis_2} > 55 \times 55 km^2 \quad \text{or} \quad \frac{axis_1}{axis_2} > 1.5 \qquad \text{(eq. B1a and b)}$$

These criteria lead to the reduced number of available views. M_AVA is further decreased by eliminating views that:

- Have *TB* values and/or amplitudes of paired views ($\sqrt{TB_X^2 + TB_Y^2}$) outside the expected ranges. This is seen as Radio Frequency Interferences (RFI) contamination. This filters however only the strong RFI.

- Have a SUN_POINT flag

A filter for softer RFI is created by looking at paired views (that is (TB_X, TB_Y) or (TB_Y, TB_X). When *M_AVA* is above a certain threshold, pairs are eliminated when their halved Stokes parameter

$(0.5*(TB_X+TB_Y))$ minus the mean value of the halved Stokes parameter of all views is bigger than a threshold. This threshold depends on the radiometric uncertainty on TB_X.

ELIMINATING L1C PIXELS FOR DETERMINATION OF *SM*

The remaining views of a pixel have a certain information content. This is roughly expressed by the initial validation index *MVAL0*. It is a weighted sum of the number of available measurements:

$$MVAL0=C_{val}DTB\_F*sum(1/DTBa) \hspace{4cm} \text{(eq. B2)}$$

With:
DTBa = radiometric uncertainties over each *TB* at antenna level
DTB_F = scaling factor (can be found in the TGRD)
C_{val} = coefficient depending on polarization mode (can be found in the TGRD)

The sum is carried out over every view and polarizations TB_X and TB_Y. When *MVAL0* is smaller than a certain threshold (see the TGRD), the complete pixel is eliminated.

THE WORKING AREA WA$_{DFFG}$ AROUND A DGG NODE

To determine the surface types of a DGG pixel, a working area of 123 x 123 km^2 is used around every DGG node. The grid used for this determination is called the Discrete Fine Flexible Grid (DFFG) and has (as used in the L2PP processor) a resolution of ± 4 km, which gives ± 900 cells. This DFFG working area (WA$_{DFFG}$) is used whenever the default fraction computation cannot be used (that is, whenever the fraction is not static, e.g. because of snowfall). Because of the angle of the satellite with the earth, contribution to the *TB* measurements can come from a larger area than the SMOS pixel. This is the reason why the WA$_{DFFG}$ is larger than the average size of the SMOS pixels

WEIGHING FUNCTIONS

Weighing functions are used for three reasons:

1. To compute the fractions of the various land types (FM0). These are used to drive the decision tree
2. To compute reference values for each mean fraction of group of fractions (FM)
3. To determine the weight of each incidence angle dependent fraction of a L1c view in order to build the aggregated forward model that is used for the retrieval.

For point 1 and 2, the mean weighing function (*MEAN_WEF*) is used. For point 3, the incidence angle dependent weighing function (*WEF*) is needed.

WEF

The incidence angle dependent weighing function (WEF) is defined in the ATBD as: "Function derived from the apodization function, to be applied to every elementary area inside the SMOS pixel in order to give the proper weight to the corresponding contribution to up-welling radiation."

An incidence angle dependent fraction does not only depend on the surface rations. The measured TB is the sum of integrated radiances through the syntetic antenna pattern. In other words, because the observation has its own geometry, weights differ for each incidence angle. The biggest contribution to the signal a satellite receives comes from the earth directly below. In order to take this into account, a synthetic antenna pattern is calculated. For this, an apodization function is used. Truncating this function to the main lobe (centre) of the antenna, does not generate major errors, which makes it possible to simplify the WEF, calculated by intersecting the synthetic antenna patterns with the earth surface, to $WEFA$.

$$WEFA(\rho_{DC}) \approx \frac{sinc(C_{WEF1} \times \rho_{DC})^{C_{WEF2}}}{1 + C_{WEF3} \times \rho_{DC}^{C_{WEF4}}} \ if \ C_{WEF1} \times \rho_{DC} £\pi$$

$$WEFA(\rho_{DC}) = 0 \ otherwise$$

(eq. B3a and b)

With:

$$\rho_{DC} = \sqrt{(\xi'-\xi)^2 + (\eta'-\eta)^2}$$

$$sinc(x) = \frac{sin(x)}{x} \ (x<>0); \ sinc(0)=1$$

(eq. B3c and d)

ρ_{DC} = distance in the DC coordinates
ξ, η = central director cosines (DC) coordinates
ξ', η' = running DC coordinates
$$\xi = \sin\theta_a \cos\phi_a ; \ \eta = \sin\theta_a \sin\phi_a$$

(eq. B4e and f)

With:
θ_a and φ_a = polar angles in the antenna reference frame, computed from θ_g, φ_g and the Attitude and Orbit Control System (AOCS)
θ_g, φ_g = geographical polar angles computed from spacecraft position and DGG node coordinates

MEAN_WEF

The $MEAN\_WEF$ is calculated in the same way as the WEF. However, to make sure that no fraction is ignored, a background component is added. The diameter of this component is WEF_SIZE. It must then be normalized to unity integral in order to compute mean fractions.

$$MEAN\_WEF(\rho_{earth})=C_{MWEF2}+WEF_A\left(\frac{\rho_{earth}}{C_{MWEF1}}\times\frac{\pi}{C_{WEF1}}\right)for\ \rho_{earth}\ \hat{I}[0,C_{MWEF1}]$$

$$MEAN\_WEF(\rho_{earth})=C_{MWEF2}\ for\ \rho_{earth}\ \hat{I}[C_{MWEF1},\frac{WEF\_SIZE}{2}]$$ **(**eq. B4a, b and c)

$$MEAN\_WEF(\rho_{earth})=0\ otherwise$$

With:
WEF_SIZE = 123 km (size of the WA$_{DFFG}$)

Values for C_{MWEF1}, C_{MWEF2} and C_{WEF1} can be found in the TGRD.

With the weighing functions known, the fractions can be determined.

CHANGING MAGNITUDE

Any fraction of a pixel can be appointed a FM0 and FM class through a map that is formed by aggregation of the ECOCLIMAP (a data set that gives the 215 different ecosystems over the world). The aggregated fractions that are used can be found in Table B.1: Aggregated fractions FM0 and FMTable B.1. Classes are called complementary when together they cover the whole surface of a pixel. The classes topography and snow are supplementary. In order to introduce them, the previous complementary classes have to be overwritten so that the resulting set again becomes complementary. The order of overwriting is: high topography, frost, snow.

Table B.1: Aggregated fractions FM0 and FM

FM0 class A	Aggregated land cover B	FM class C	Complementarity D	E
FNO	Vegetated soil + sand	FNO	Complementary	Sum of Complementary fractions equals unity
FFO	Forest	FFO		
FWL	Wetlands	FWL		
	Open fresh water	FWP		
	Open saline water	FWS		
FWO	Open water			
FEB	Barren	FEB		
FTI	Total Ice Fraction	FTI		
	Ice & permanent snow			
	Sea Ice			
FEU	Urban	FEU		
FTS	Strong topography		Supplementary	Supplementary fractions are superimposed
FTM	Moderate topography			
FRZ	Frost	FRZ		
FSW	Non permanent dry snow	FSN		
FSM	Non permanent wet snow			
	Non permanent mixed snow			

The aggregated ECOCLIMAP can well be used to appoint classes to fractions. However, non-permanent (NPE) surface conditions (soil that gets frozen, flooding) are not taken into account. Before further steps in the decision tree can be taken, the pixels first need to be updated by checking for these conditions. Both the reference values and the fraction values need to be adjusted. This should be done in the order that is mentioned below.

1. Check for frozen soil:
   - Trigger: *STL1 < TH_SOIL_FRZ* in which *STL1* is the surface soil temperature.
   - Actions on reference values:
   - *TAU_FRZ = (FNO*TAU_FNO+FFO*TAU_FFO+FEU*TAU_FEU+FEB*TAU_FEB) / (FNO + FFO + FEU + FEB)*
   - Actions on fractions values: *FRZ = FNO + FFO + FEU + FEB; FNO=FFO=FEU=FEB=0*
2. Check for sea ice
   - Trigger: *FSI ≠ 0*
   - Actions on reference values: *TAU_FSI = (FWP*TAU_FWP + FWS*TAU_FWS)/ (FWP + FWS)*
   - Actions on fractions values: *FWS = (1-FSI)*FWS, FWP=(1-FSI)*FWP*
3. Check for pure water that freezes over
   - Trigger: (*LSM>0.5* and *2T< TH_PWATER_FRZ*) or (*LSM<0.5* and *SSTK<TH_PWATER_FRZ)* in which the size of *LSM* depends on whether land or sea is the dominant in the cell and SSTK stands for sea surface temperature.
   - Actions on reference values: *TAU_FEI = (FEI*TAU_FEI + FWL*TAU_FWL + FWP*TAU_FWP) / (FEI + FWL + FWP)*
   - Actions on fractions values*: FEI = FEI + FWL + FWP; FWL = FWP = 0*
4. Check for saline water that freezes over
   - Trigger: *(LSM > 0.5 and 2T < TH_SWATER_FRZ) or (LSM < 0.5 and SSTK < TH_SWATER_FRZ)*
   - Actions on reference values: *TAU_FSI = (FSI*TAU_FSI + FWS*TAU_FWS) / (FSI + FWS)* Actions on fractions values: *FSI= FSI + FWS; FWS = 0*
5. Check for a mixed snow cover
   - Trigger: *FSM ≠ 0*
   - Actions on reference values: TAU_FSM = (FNO*TAU_FNO + FFO*TAU_FFO + FEI*TAU_FEI + FSI*TAU_FSI + FRZ*TAU_FRZ)/ (FNO + FFO + FEI + FSI + FRZ)
   - Actions on fractions values: FNO = FNO*(1-FSM), FFO = FFO*(1-FSM), FEI = FEI*(1-FSM), FRZ = FRZ *(1-FSM)
6. Check for a wet snow cover
   - Trigger: FSW ≠ 0
   - Actions on reference values: *TAU_FSW=(FNO*TAU_FNO + FFO*TAU_FFO + FEI*TAU_FEI + FSI*TAU_FSI + FRZ*TAU_FRZ)/ (FNO + FFO + FEI + FSI + FRZ)*
   Actions on fractions values: *FNO = FNO*(1-FSW), FFO = FFO*(1-FSW), FEI= FEI*(1-FSW), FRZ = FRZ *(1-FSW)*

For both snow covers goes that the given snow cover parameter is a percentage. If this percentage is smaller than 100%, it is to be applied to each pre-existing aggregated land fraction, selecting for each of them the zone with the highest absolute latitude.

7. Check for freezing water content standing vegetation
   Trigger: $SKT < TH\_VEG\_FRZ$ in which SKT is the temperature of the vegetation
   Actions on reference values: $TAU\_FNO = 0$
   Actions on fractions values: -
8. Check for flooding by precipitation
   Trigger: $LSCP+CP>TH\_FLOOD$ in which $LSCP$ stands for large scale precipitation and CP for convective precipitation.
   Actions on reference values: $TAU\_WL=(FNO*TAU\_FNO+FFO*TAU\_FFO+FWL*TAU\_FWL) / (FNO + FFO + FWL)$
   Actions on fractions values: $FWL = FWL + FNO + FFO, FNO = FFO = 0$

The final integrated fractions FM and FM0 for the working area WA_{DFFG} are simply computed using the weighted mean of the local DFFG pixel fractions over all the DFFG pixels that belong to WA_{DFFG}, calculated with the $MEAN\_WEF$ or WEF.

THE FOREST WINTER EXCEPTION CASE

For an area with almost no vegetation except for a forest fraction, it is expected that using the forest fraction (being the second main fraction) as for retrieval enables a better estimate of the forest vegetation opacity, even though this goes at the expense of the retrieved SM quality. However, this retrieval should only be done once, in order to change a missing optical depth value.

- Trigger: ($TAU\_NOFFO < TH\_TAU\_F1$) and ($TH\_F1 < FFO < TH\_F2$) and ($FNO > TH\_NO$) and ($DGG\_CURRENT\_TAU\_NAD\_FO == -999$)
With TAU_NOFFO being the mean optical depth of all the 10 fractions but forest.
$TAU\_NOFFO = (FNO*TAU\_FNO + FWL*TAU\_FWL + FWP*TAU\_FWP + FWS*TAU\_FWS + FEB*TAU\_FEB + FTI*TAU\_FTI + FEU*TAU\_FEU + FRZ*TAU\_FRZ + FSN*TAU\_FSN) / (FNO + FWL + FWP + FWS + FEB + FTI + FRZ + FSN)$
- Actions on reference values: -
- Actions on fractions values (follow the order): $FNO = FNO - (TH\_F1 - FFO);$
$$FFO = FFO + (TH\_F1 - FFO) = TH\_F1$$

DETERMINING RETRIEVAL CASE

In order to find out which retrieval case is to be used for a pixel, the fractions are compared to a specific value that depends on a threshold and whether this threshold is meant for the whole fraction, or for the land fraction. In other words: F** > THVAL * FREF

With:
$THVAL$ = threshold TH_*** (%)
$FREF$ = fraction of reference that depends on a key TH_***_D, either:

- equal to 1, when the threshold value is meant for the whole fraction,
- equal to the land total fraction $FLA = 1 - FWO$, when the threshold value is meant for the land fraction.

Values for these parameters can be found in the TGRD. The order in which fractions have to be considered can be found there as well. There are 19 retrieval cases.

DETERMINING MODEL

With the retrieval case known, one can determine which model should be used for every fraction, in order to calculate their contribution. There are four different models: the nominal model, the water model, the cardioid model and the snow model (which for now is replaced by the cardioid model, as long as there is no actual snow model). A table that shows the different options can be found in the TGRD. Depending on the retrieval case, some contributions are default. This is indicated by a 0.

RETRIEVAL

In order to retrieve the SM from the measured TB value at the satellite, the modelled TB values for all polarized views have to be subtracted from the measured values and the difference has to minimized. This is done through iteration. However, before this can happen, the modelled TB has to be transformed in order to correspondent to the measured TB. The modelled TB is namely the Top Of Atmosphere (TOA) TB for horizontal and vertical TB components, TB_H and TB_V, while the satellite measures TB_X and TB_Y. This means the modelled TB's need to be adjusted for a rotation of the electric fields, due to geometrical considerations and the Faraday rotation induced by the ionosphere.

In order to do this, a rotation angle $\boldsymbol{a}$ is defined:

$\boldsymbol{a} = -a_r + \omega_{Fa}$

With:
a_r = rotation that transfers the TOA signals to the antenna reference frame
ω_{Fa} = the angle that corrects the Faraday rotation

Both are directly supplied by the L1c data for SM.

For dual polarization mode, the TB can then be calculated with the following matrix:

$$\begin{bmatrix} A1 \\ \hline A2 \end{bmatrix} = \begin{bmatrix} cos^2(a) & sin^2(a) \\ sin^2(a) & cos^2(a) \end{bmatrix} [MR2] \begin{bmatrix} TB_H \\ TB_V \end{bmatrix}$$

(eq. B5)

With:
$A1 = BT_{XX}$
$A2 = BT_{YY}$

This way, the antenna TB's are modelled twice more than the observed ones. Therefore, only the relevant modelled antenna polarizations (i.e., those that match the observed ones) should be selected.

The next step is to define a cost function, in which the modelled *TB* is subtracted from the measured *TB*. By minimizing this function, it is possible to retrieve parameters such as *SM*. The cost function is:

$$COST=(TBM_m-TBF(\theta_m,p_i))^t [COV_T]^{-1} (TBM_m-TBF(\theta_m,p_i))+\sum_i \frac{[p_i-p_{i0}]^2}{\sigma_{i0}^2}$$ **(eq. B6)**

With:

TBM_M = valid measured values of brightness temperature
TBF(θ, pi...) = direct (forward) model for brightness temperature at the antenna level
[COV_T] = is the variance matrix for the observed *TBM*, given in the L1c data
θ = incidence angle
p_i = physical parameters
p_{i0} = prior estimates of the free physical parameters
σ_{i0}^2 = prior variances of the free physical parameters

This can be written as:

$$COST=(TBM_m-TBF(\theta_m,p_i))^t [COV_T]^{-1} (TBM_m-TBF(\theta_m,p_i))+(p_i-p_{i0})^t [COV_P]^{-1}(p_i-p_{i0})$$ **(eq. B7)**

With:

$$COV_P=\begin{bmatrix} \sigma_{10}^2 & 0 & ... & 0 \\ 0 & \sigma_{20}^2 & 0 & 0 \\ M & & O & 0 \\ 0 & ... & 0 & \sigma_{NP0}^2 \end{bmatrix}$$ **(eq. B8)**

Or even:

$$COST=(DIFF)^t [COV_Z]^{-1} (DIFF)$$
(eq. B7)

With:

$$[COV_Z]=\begin{bmatrix} [COV_a] & 0 \\ 0 & [COV_p] \end{bmatrix}$$: square (N*T+NP*) ranked matrix built by aligning [*COV_T*] and

[*COV_P*] along the main diagonal
DIFF: Vector with length NT+NP, consisting of NT terms *(TBM_m - TBF(θ_m, p_i))*, then P terms *(p_i – p_{i0})*
NT = number of validated L1c observed views (= *M_AVA*)

C L2PP PROCESSOR

INTRODUCTION

The L2PP processor is a device created by the Canadian company Array Systems Computing Inc.. It is used to determine *SM* content from so called L1c data. L1c data files are provided by ESA. They give among other information on *TB*. The L2PP processor uses the algorithm as described in chapter 3 to determine *SM*.

For this research, version 3.07 of the L2PP processor was used. This appendix gives information on the installation and use of this version of the processor. Part of the information comes straight from the processor's Software User Manual (SUM) [C.1] or the Software Release Document (SRD) [C.2], other from experience. It should be noted that at the time of printing, a newer version of the processor was already available. Some of the information in this appendix might be obsolete or incorrect when using newer version. It is therefore advised to use this appendix next to the official documents provided by Array. This appendix does not contain all information on the L2PP processor, but is merely a collection thought most useful by the author.

At this point, the processor only runs on the operating system Linux. The appendix starts therefore with a short paragraph on Linux. It continues with the installation of the processor and finally explains its most basic operations. More detailed information can be found in the SUM.

LINUX

For the use of the L2PP processor, the Linux distribution Red Hat version 4.0 is needed. Other distributions (e.g. Ubuntu) might not work. There is a version of Red Hat that can be used free of charge (without the helpdesk from Red Hat), called CentOS. It is unclear whether newer versions of Centos than 4.0 will also work.

The (Dutch) site http://beeman.nl/howto/oracle-xe-met-apache-en-php-installeren-op-centos/ (last viewed on 04-10-2010) gives an explanation on how to install CentOS and where to find different versions. Under paragraph "1.2 CentOS installeren", one of the final Lines read: "Vink alle opties uit en selecteer alleen de optie 'minimal'... Hiermee heb je de installatie teruggebracht van 2475MB naar 679MB !!" When you have no experience with Linux, it is advised not to mark the option 'minimal'. Installing the complete package ensures that a desktop similar to Windows is installed, which is not the case with the option 'minimal'.

If there is already an operating system installed, it is possible to use a virtual machine to run Linux on. Make sure this virtual machine has enough memory. The SRD suggest a RAM of 16 GB. With a smaller memory, the processor might not be able to execute any calculations. Please see the SRD for other requirements.

INSTALL THE PROCESSOR

For the installation of the processor, two compressed files can be found on ftp://www.array.ca/L2SM_Processor/Older_Versions/20100326_V3_7/smos-l2sm-03.07-20100325/ (last viewed on 04-10-2010); smos-l2sm-03.07-20100325-prototype-bin.tar.gz and smos-l2sm-03.07-20100325-data.tar.gz. As the names suggest, the first comprises the actual processor, while the latter holds the auxiliary data files. In order to enter the site, you will need a username and password.

The auxiliary data provided with the processor is test data. It is data from a specific date and location that fits the L1c data that is provided with the processor as well and used to test whether the processor works correctly. In order to calculate *SM* over Ghana and Burkina Faso, at the very least another ECMWF data file is needed, because the ECMWF test file does not cover this area. A file that does cover the area can be found on http://www.brockmann-consult.de/beam-wiki/display/SBOX/SMOS+Toolbox+for+BEAM;jsessionid=09C58D08153BF08D7BBA8B94A946FBF1 (last viewed on 04-10-2010). This data is considered trainings data by Brockmann Consult, and is again data of a specific date and location. If it is important to get a representative *SM* value, one might want to consider replacing some of the other auxiliary data files as well. For a sensitivity analysis as executed in chapter 4, this is less important.

In order to get the processor running, at the very least a Java Development Kit needs to be installed. Furthermore, some libraries might have to be installed. This will however depend on what is already installed on the computer that is used.

The installation of the processor is rather simple, even with only some basic Linux knowledge. Whenever in doubt, information on the meaning of most commands can be found on internet. The two compressed files above make it possible to execute a binary installation. Below, the commands needed are given (in italics when directly from SRD), including some extra information for people with no Linux experience.

[0] Open the terminal (used to type commands).

[*1] Before installing the SML2PP, a Java Development Kit must be installed. One can be downloaded*
from http://java.sun.com. The version used in these instructions is JDK 6.0 Update 1, from the file
jdk-6u1-linux-amd64.bin. Other implementations may work.

[2] To install this version:
cp jdk-6u1-linux-amd64.bin /opt
cd /opt
sh jdk-6u1-linux-amd64.bin

[3] The installation directory, such as /opt/jdk1.6.0_01/bin should be added to the PATH; this can be done by adding the following line to the .bashrc file in the home directory:
PATH=/opt/jdk1.6.0_01/bin:$PATH

[4] After doing this, log out and log in again. Run the command "java –version". The output of the
command should match the version installed, in this case 1.6.0_01.

[5] Note that java does not need to be installed in /opt; any convenient location may be used.

Now that Java is installed, it is possible to install the processor.

[1] The binary installation is distributed as two compressed tar files, smos-l2sm-03.07-20100325-prototype-bin.tar.gz and smos-l2sm-03.07-20100325-data.tar.gz. In order to unpack and install the software, the following steps should be taken.

[2] To unpack and install the binary distribution, for the prototype processor:
mkdir <installation>
cp smos-l2sm-03.07-20100325-prototype-bin.tar.gz <installation>
cp smos-l2sm-03.07-20100325-data.tar.gz <installation>
cd <installation>
tar zxvf smos-l2sm-03.07-20100325-prototype-bin.tar.gz
tar zxvf smos-l2sm-03.07-20100325-data.tar.gz
./install

[3] The executables (including L2SM___03_07.exe) are located in <installation>/bin. [C.2]

A short explanation on the commands used during installation:
Mkdir stands for make directory. In other words, you create the directory in which the program is installed.
Cp stands for copy. It is possible to do this with the right mouse button.
Cd stands for change directory. "cd <installation>" Thus brings you to the directory in which you have located the compressed processor files.
Sh stands for: Runs or processes jobs through the Bourne shell. See e.g. http://www.computerhope.com/unix/ush.htm for more information (viewed on 04-10-2010).
Tar zxvf unpacks the compressed files and ./install installs the files.

Sometimes, logging in on Linux in the normal way does not give you all the privileges needed for installing the processor. Instead, you might have to login as a root user. This can simply be done by typing su <username> in the terminal. User name and password of the root user should have been established during installation. An example is that the directory can only be changed into /opt (for the installation of java) when you are logged in as a root user. However, there is a way to work around this. Instead of going to /opt, you can create a directory opt (mkdir opt). This will be placed under the home directory, found on the desktop. In that case, instead of PATH=/opt/jdk1.6.0_01/bin:$PATH, on should type PATH=opt/jdk1.6.0_01/bin:$PATH in order to link java to the processor.

After installation, the graphical user interface (GUI) of the processor should start by typing <installation>/bin/SMl2ppgui. A working interface does unfortunately not necessarily mean that the processor works. In order to check this, typ:
cd <installation>/bin
./self-test
Any problems should reveal themselves in the terminal.

LIBRARIES

During the test phase of the processor for this research, it was revealed that three libraries had to be installed before the processor could do its job. These were compat-libf2c-32-3.2.3-56.fc5.x86_64.rpm, lsb-buildenv-4.0bzr20080722-7.1.x86_64.rpm and xerces-c-2.7.0-6.fc5.x86_64.rpm. These and many other libraries can be downloaded from http://rpm.pbone.net/ (last viewed on 04-10-2010). However, be careful when choosing a library. They all have their own requirements, which might mean again different libraries have to be installed on your computer. Installing the libraries is done by typing rpm –Uhv <library>.

VISUALIZING DATA

Finally, it is possible to install e.g. arcexplorer to visualize the data. The author instead used Beam-VISAT for visualization. This program used to be downloadable from Brockmann Consult, but seems to have been replaced by the so-called SMOS-Box 2.1 Add-on for BEAM 4.8 (30.07.2010) (http://www.brockmann-consult.de/cms/web/beam/software, last viewed on 04-10-2010).

USE OF THE PROCESSOR

This paragraph gives a short overview of the most important functions of the L2PP processor. Only functions actually used during this research are denoted. For a more detailed description, please see the L2PP processor's SUM.

JOB MANAGEMENT

Opening the processor's GUI will give the following screen:
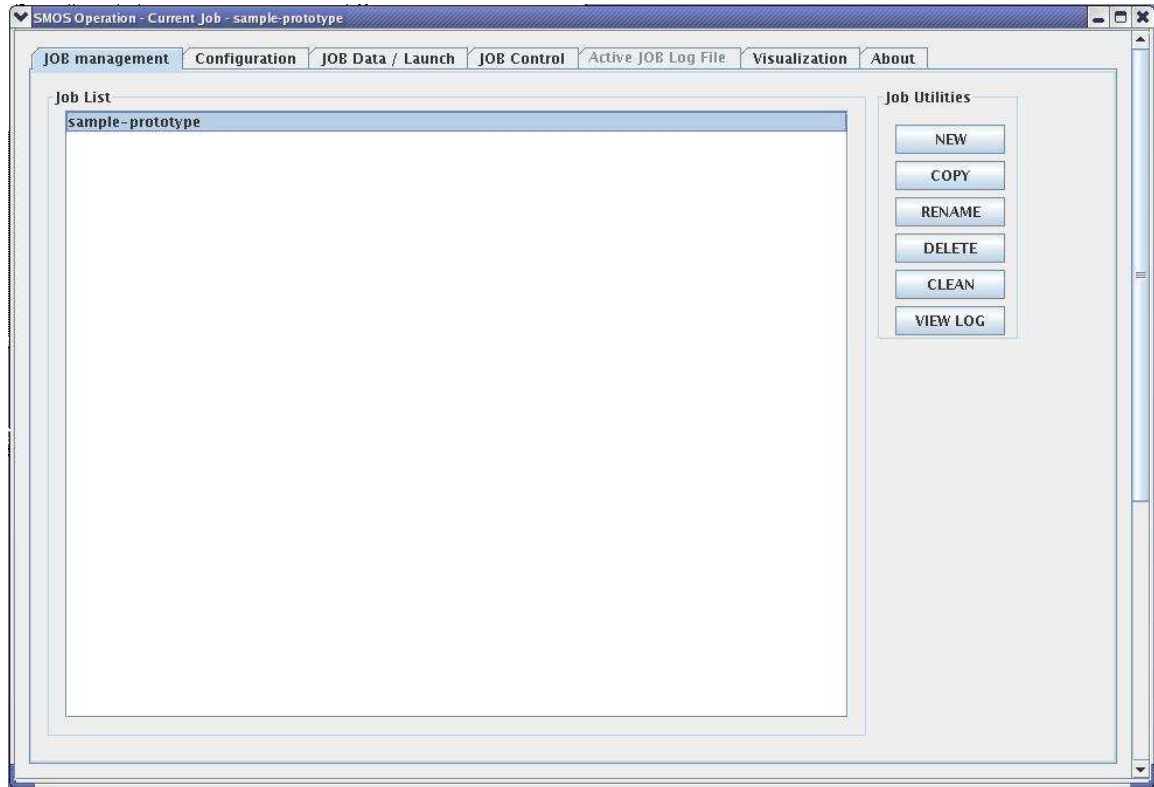
The screen shows all the jobs that are created. A job is a complete package of L1c and auxiliary data that can be run in order to determine an area's *SM* content. New jobs can be created by clicking "new". Jobs can be copied, renamed or deleted by first clicking on the job and then on the respective buttons. In the same way, the log (showing what has happened during the last run) of a job can be shown by clicking the button "view log". The "Clean" button removes intermediate files that were created during the run and are no longer needed. Sample-prototype is the job that is run with the command ./self-test.

When a job is selected, changes can be made to e.g. vegetation data by going to one of the following tabs in the GUI. It should be noted however that these changes are made in the auxiliary files distributed with the processor, not just in the processor. Obviously, when copying a job, all changes made to the first job are copied as well. More importantly however is that any changes made to the second job are also carried through in the first job, unless the two jobs use different auxiliary files. See the following sub paragraphs for more information.

CONFIGURATION
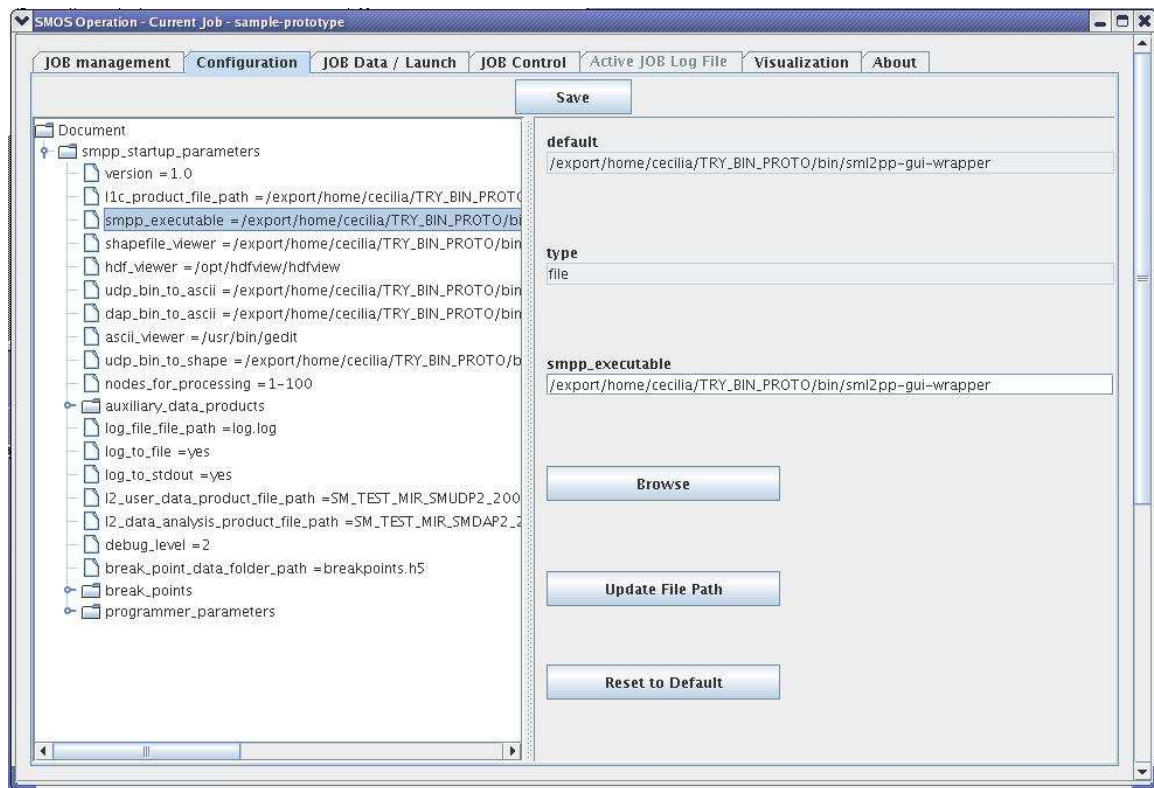
This tab allows you to change the L1c data and auxiliary files used in the calculation. Do not forget to press both the update and the save button when changing a file. Under nodes_for_processing, the nodes that are processed during a run can be adapted. Choosing the right nodes can save a lot of time. If you just copied a job and want to change it parameters, without changing those of the original job, the first thing you will have to do is create new auxiliary files and link the job to those files in this tab. A simple way of creating new files is to copy an existing file.

The figure seems to suggest that in any case, you will be able to return to the default file (or for other tabs, data). This button is not always present. Be careful with changing data (files) if you expect to want to turn back to the default settings.

The ECMWF data file can be adapted under auxiliary_data_products → ecmwf_product_auxiliary_data → ecmwf_product. The data file containing vegetation parameters can be changed under auxiliary_data_products → land_cover_classes_product_auxiliary_data, the data file with more general data under auxiliary_data_products → l2SM_configuration_parameters_product_auxiliary_data.

154

JOB/DATA LAUNCH
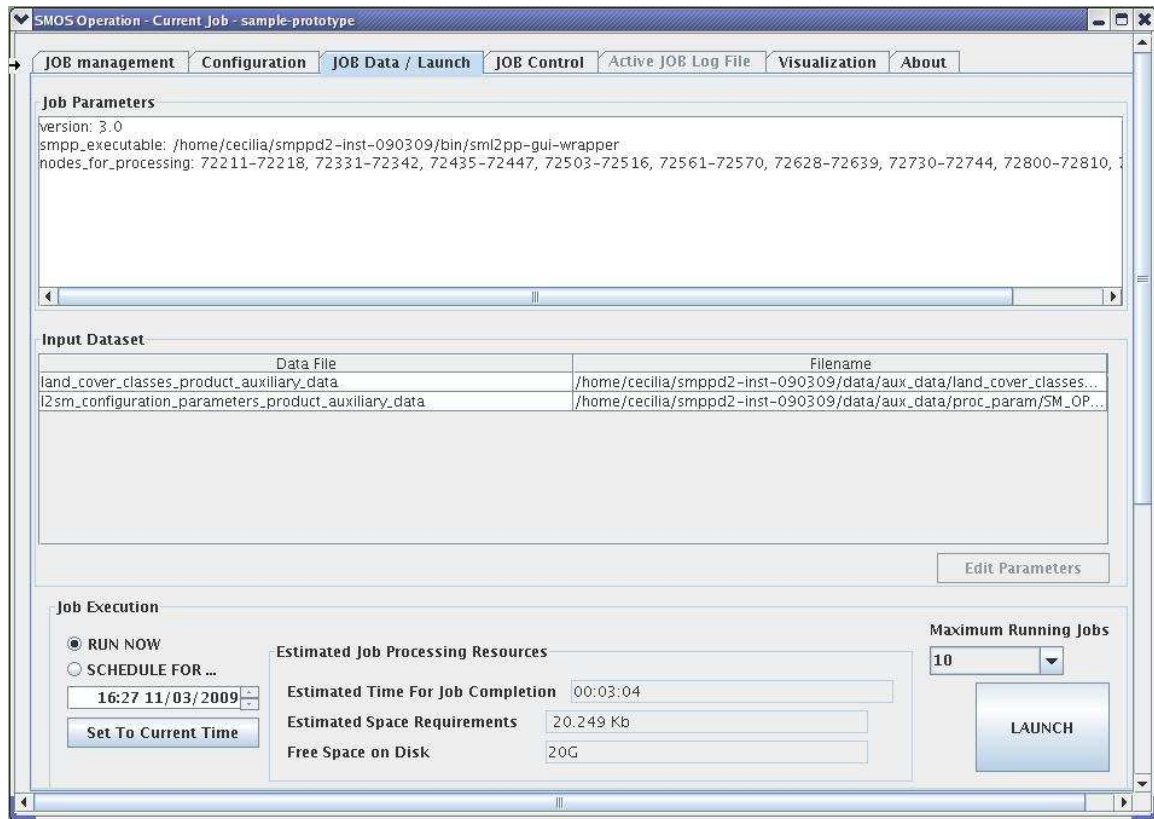
This tab allows you to change data in the landcover or l2sm file, and run jobs by pressing launch. To the left of the launch button, you can define when you want a job to run. For adapting parameters, click on one of the data files and press edit parameters. Then go to the parameter you would like to change. Again, do not forget to press both update and save and pay attention to whether it is possible to return to the default value. Information on the current run is given on the bottom of the tab, information on the job and the nodes being processed on the top.

Validation of SMOS satellite data over Ghana and Burkina Faso

JOB CONTROL

This tab gives information on how long the job has been running and how much longer it will probably take. The process can be paused or stopped by the respective buttons. Only the pause button offers the opportunity of returning to the calculation later. The stop button will make sure the calculation is stopped at a convenient time. The kill process also stops the run, but in this case right away. The remove button allows you to remove a stopped, killed or completed job from the list, so that a new run can be launched.

ACTIVE JOB LOG
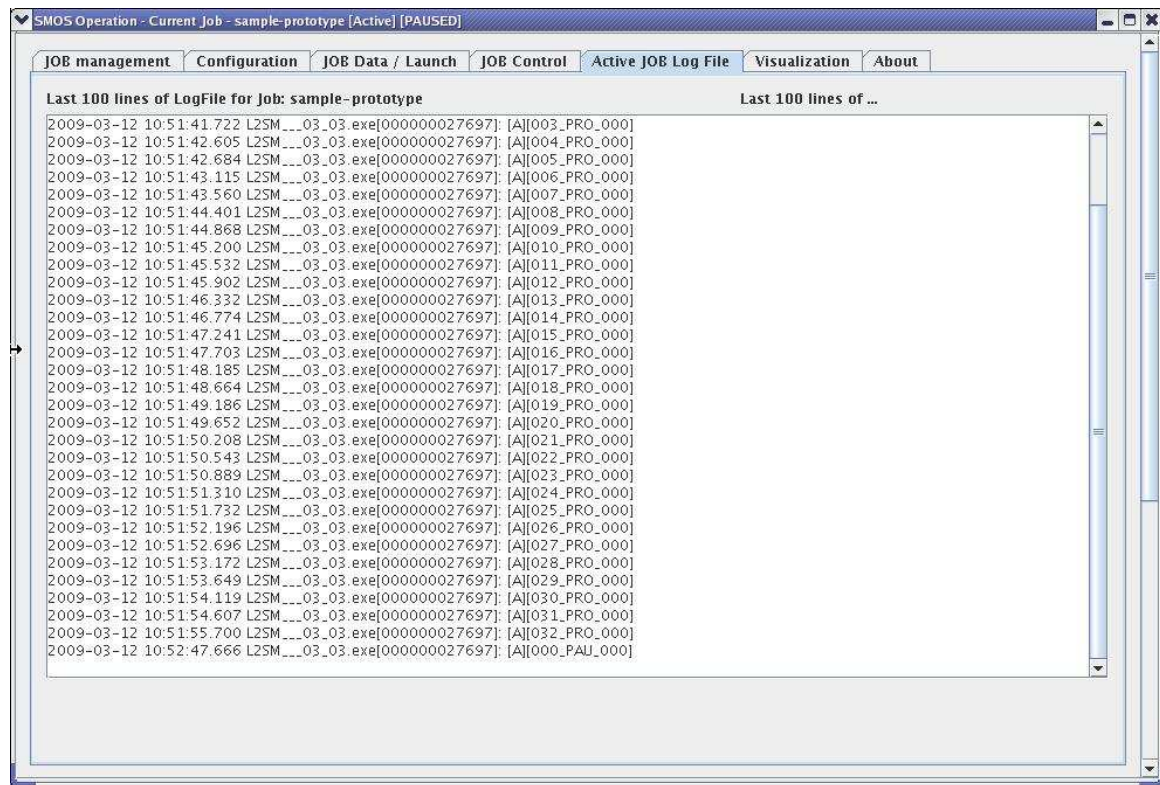
The active job log gives information on the current run. The same information can also be found in the terminal.

Validation of SMOS satellite data over Ghana and Burkina Faso

VISUALIZATION
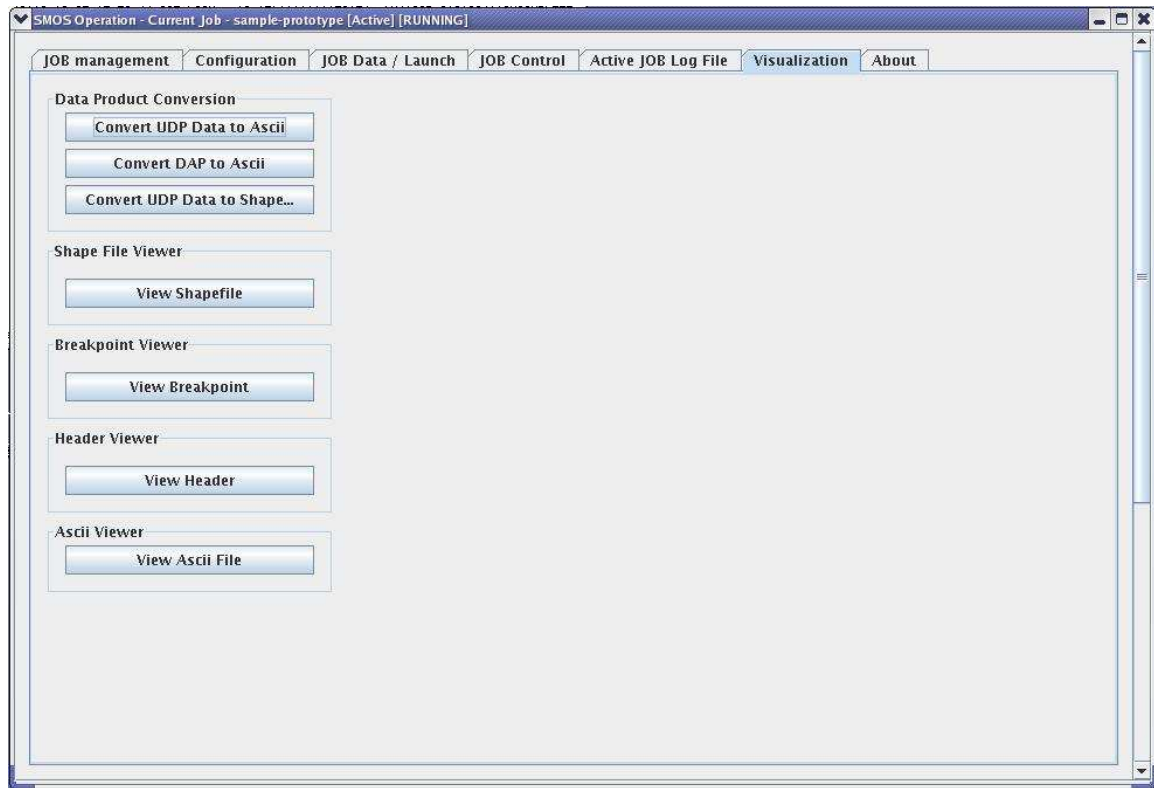
This tab gives the opportunity to visualize the data from a completed run. For this, arcexplorer or a similar program will have to be installed on the computer. Another way is to simple take the data from <installation>-jobs-prototype_jobs-<name_job> and open it in e.g. VISAT.

VALIDATION PERIOD

When using the processor, one should be aware that it is set for the test data. This means the so called validation period, the period over which calculations can done, correspond to the test data's date, but not necessarily with other L1c data. When other data is processed, this validation period might need to be adapted. Make sure the validation period contains the date for which you would like to do your calculations. Any validation periods that do not correspond to the used L1c data can be found in the terminal after launch and (in this case) failure of a run.

D MATLAB CODE

INTRODUCTION

Because some parameters could not be adapted in the L2PP processor, a simple Matlab code was written in order to execute a sensitivity analysis for these parameters. This code calculates *SM* for a single pixel, from a single value for TB_H and TB_V.

The Matlab code is divided into five different m-files. The first two files calculate the R1 and R2 (as defined in chapter 3) for either H or V polarization. These files are called lowvegH and lowvegV. This calculation is done with the formulas of the nominal model. Default values for the parameters are representing low vegetation. The following two files calculate the contributions of sky and atmosphere to *TB* and determine the total theoretical *TB*. These files are called combiH and combiV. In the future, these files can be adapted so that contributions from other models can be added as well. Again, the calculations are done for both H and V polarization. In the final m-file, called iteration, a contribution is added to the theoretical *TB* in order to make the *SM* values comparable to that found in the L2PP processor. Next, *TB* measured by the satellite and the theoretical *TB* are subtracted for the two polarizations. The differences are squared and added. This number is than minimized using the optimtool from Matlab. This tool contains the codes for many minimization algorithm, including Levenberg-Marquardt.

The following paragraph gives the m-files. combiH And combiV are combined into one file, because the differences are minimal. Whenever there is a difference between the two files, this is mentioned. For many parameters, a short explanation is given behind the % sign. More information can be found in chapter 3 and appendix A.

M-FILES

LOWVEGH

```
function R_lv=lowvegH(SM) %lv stands for low vegetation

R_lv=zeros(2,1);

theta=0.74176; %incidence angle from L1c data [rad]

n=1;%1 stands for December 14th, 2 for April 6th, 3 for May 20th 05:15 h
%4 for May 20th, 17:48 h

T(1)=298.876;   %default temperature (both soil and ground)

T(2)=298.87622; %comes from L2PP calculation of default value as found
in

T(3)=298.85815; %table 4.2
```

```
T(4)=298.8835;

tau_s(1)=0.3157793;  %default optical depth

tau_s(2)=0.20133308; %comes from L2PP calculation of default value as
found

tau_s(3)=0.16711998; %in table 4.2

tau_s(4)=0.25356796;

%parameters that depend on location

C=0.06;       %clay fraction [-], average over data from previous field
work [D.1]

S=0.57;       %sand fraction [-], average over data from previous field
work [D.1]

rho_b=1.3;    %soil bulk density, typical value according to ATBD [gcm^-
3]

S_g=0;        %salinity of the soil [ppt]

rho_s=2.664; %soil particle density, typical value according to ATBD
[g/cm]

mu_s=1;       %magnetic permeability [-]

T_g=T(n); %temperature of the ground [K]

T_c=T(n); %temperature of canopy [K]

HR=0.1; % effective surface roughness dimensionless parameter [-]

QR=0;     %polarization coupling factor, related to the fact that
roughness tends to induce polarization mixing, default value according
to ATBD [-]

NR_p=0;   %integer used to parameterize the dependence of the roughness

        %effects on incidence angle, default value according to ATBD [-

%parameters that do not depend on location

SGEF1=-1.645; SGEF2=1.939; SGEF3=-2.256; SGEF4=1.594; % used to
calculate sigma_eff, in turn used to calculate dielectric constant of
free water [-]

f=1.413e9; %frequency of SMOS satellite [Hz]

eps_winf=4.9; %high frequency limit of the static dielectric constant of

%water eps_w0 [F/m]

ow1=88.045; ow2=-0.4147; ow3=6.295e-4; ow4=1.075e-5; ow5=87.134;

ow6=-1.949e-1; ow7=-1.276e-2; ow8=2.491e-4; ow9=1; ow10=1.613e-5;
```

```
ow11=-3.656e-3; ow12=3.210e-5; ow13=-4.232e-7; ow14=1.1109e-10;

ow15=-3.824e-12; ow16=6.938e-14; ow17=-5.096e-16; ow18=1; ow19=2.282e-5;

ow20=-7.638e-4; ow21=-7.760e-6; ow22=1.105e-8; ow23=0.18252;

ow24=-1.4619e-3; ow25=2.093e-5; ow26=-1.282e-7; ow27=2.033e-2;

ow28=1.266e-4; ow29=2.464e-6; ow30=1.849e-5; ow31=-2.551e-7;
ow32=2.551e-8;

%coefficients for static water dielectric constant eps_w0 [1]

BERE1=1.2748; BERE2=-0.519; BERE3=-0.152;

%components for exponent beta (exponent complex function in EPS_s), real

%part [-]

BEIM1=1.338; BEIM2=-0.603; BEIM3=-0.166;

%components for exponent beta (exponent complex function in EPS_s),

%imaginary part [-]

eps_0=8.854187817e-12; %permittivity of free space [F/m]

alpha=0.65; %exponent in eps_s (local use) [-]

CPA1=1.01; CPA2=0.44; CPA3=-0.062; %coefficients for eps_pa,
respecitively [F/m]^1/2, [F m2/g]1/2, [F/m]


%calculation of dielectric constant of area (eps_fw, real and imaginary

%part. For more information, see appendix A

sig_eff=SGEF1+SGEF2*rho_b+SGEF3*S+SGEF4*C;

eps_pa=(CPA1+CPA2*rho_s)^2+CPA3;


eps_s0=ow5+ow6*T_g+ow7*T_g^2+ow8*T_g^3;

a_ST=ow9+ow10*S_g*T_g+ow11*S_g+ow12*S_g^2+ow13*S_g^3;

b_ST=ow18+ow19*S_g*T_g+ow20*S_g+ow21*S_g^2+ow22*S_g^3;


if  S_g<100;  %this boundary might in reality lay somewhere else. This
is not important for this calculation, because in the research area,
salinity is not an issue

    eps_w0=ow1+ow2*T_g+ow3*T_g^2+ow4*T_g^3;
```

```matlab
    t_relax=(ow14+ow15*T_g+ow16*T_g^2+ow17*T_g^3);

    w='fresh';

elseif S_g>=100;

    eps_w0=eps_s0*a_ST;

    t_relax=(((ow14+ow15*T_g+ow16*T_g^2+ow17*T_g^3))*b_ST);

    w='salt';

end


eps_fwr=eps_winf+((eps_w0-eps_winf)/(1+(2*pi*f*t_relax)^2));

eps_fwi=((2*pi*f*t_relax*(eps_w0-eps_winf))/(1+(2*pi*f*t_relax)^2))

+((sig_eff/(2*pi*f*eps_0))*((rho_s-rho_b)/(rho_s*SM)));


beta_r=(BERE1+BERE2*S+BERE3*C);

beta_i=(BEIM1+BEIM2*S+BEIM3*C);


eps_b=(1+(rho_b/rho_s)*(eps_pa^alpha-1)+SM^beta_r*eps_fwr^alpha-SM)

^(1/alpha)-i*((SM^beta_i*eps_fwi^alpha)^(1/alpha));


%calculation reflectivity and emissivity. 1 In r_b1, r_g1, tau_s1 and
%omega1 stands for horizontal polarization

r_b1=abs((eps_b*cos(theta)-sqrt(mu_s*eps_b-sin(theta)^2))

/(eps_b*cos(theta)+sqrt(mu_s*eps_b-sin(theta))));

r_g1=((1-QR)*r_b1+QR*r_b1)*exp(-HR*cos(theta)^NR_p);

e_g1=1-r_g1;


%parameters for calculation R1 and R2

omega1=0; %single scattering albedo, default value low vegetation

        %according to ATBD [-]
```

```matlab
%parameters for calculating contribution litter to optical depth

a_L=2.33;   %parameters used to compute the litter moisture content
(Mg_L)

b_L=0;      %from SM SM, default values according to ATBD [-]

Bs_L=0.3;   %dry biomass of litter, default value according to ATBD
[kg/m^2]

c_L=0.24;   %attenuation coefficient of litter, default value according
to ATBD [m2/kg]


%parameters adaptation optical depth to polarization

tt_H=1; %angular correction parameters of optical depth to incidence
angle for H polarization, default value for low vegetation according to
ATBD [-]


%optical depth without litter contribution

tau_s=tau_s(n); %the actual formula (see appendix A) is not used because

                %value comes directly from L2PP processor


%adaptation optical depth to polarization

tau_s1=tau_s*(sin(theta)^2*tt_H+cos(theta)^2);


%contribution litter to optical depth

Mg_L1=a_L*SM+b_L; %Mg_L=litter moisture content, =<0.8 [kg/kg]

if Mg_L1>0.8

    Mg_L=0.8;

elseif Mg_L1<0

    Mg_L=0

else Mg_L=a_L*SM+b_L;

end

LWC=(Mg_L/(1-Mg_L))*Bs_L; % LWC=litter water content [kg/m^2]

tau_L=c_L*LWC;
```

```
tau_c1=tau_s1+tau_L; %in case of no litter, tau_L is 0


%calculation R1 and R2 for horizontal polarization

R1H=r_g1*exp(-2*tau_c1);

R2H=e_g1*(T_g)*exp(-tau_c1)+(T_c)*(1-omega1)*(1-exp(-tau_c1))

*(1+r_g1*exp(-tau_c1));



R_lv=[R1V;R2V];
```

LOWVEGV

```
function R_lv=lowvegV(SM) %lv stands for low vegetation

R_lv=zeros(2,1);

theta=0.74176; %incidence angle from L1c data [rad]

n=1;%1 stands for December 14th, 2 for April 6th, 3 for May 20th 05:15
h,4 for May 20th, 17:48 h

T(1)=298.876;    %default temperature (both soil and ground)

T(2)=298.87622; %comes from L2PP calculation of default value as found
in

T(3)=298.85815; %table 4.2

T(4)=298.8835;

tau_s(1)=0.3157793;  %default optical depth

tau_s(2)=0.20133308; %comes from L2PP calculation of default value as
found

tau_s(3)=0.16711998; %in table 4.2

tau_s(4)=0.25356796;

%parameters that depend on location

C=0.06;      %clay fraction [-], average over data from previous field
work [D.1]

S=0.57;      %sand fraction [-], average over data from previous field
work  [D.1]
```

```matlab
rho_b=1.3;    %soil bulk density, typical value according to ATBD [gcm^-
3]

S_g=0;        %salinity of the soil [ppt]

rho_s=2.664; %soil particle density, typical value according to ATBD
[g/cm]

mu_s=1;       %magnetic permeability [-]

T_g=T(n); %temperature of the ground [K]

T_c=T(n); %temperature of canopy [K]

HR=0.1;    % effective surface roughness dimensionless parameter [-]

QR=0;      %polarization coupling factor, related to the fact that
roughness tends to induce polarization mixing, default value according
to ATBD [-]

NR_p=0;    %integer used to parameterize the dependence of the roughness

        %effects on incidence angle, default value according to ATBD [-]

%parameters that do not depend on location

SGEF1=-1.645; SGEF2=1.939; SGEF3=-2.256; SGEF4=1.594; % used to
calculate

%sigma_eff, in turn used to calculate dielectric constant of free water
[-]

f=1.413e9; %frequency of SMOS satellite [Hz]

eps_winf=4.9; %high frequency limit of the static dielectric constant of

%water eps_w0 [F/m]

ow1=88.045; ow2=-0.4147; ow3=6.295e-4; ow4=1.075e-5; ow5=87.134;

ow6=-1.949e-1; ow7=-1.276e-2; ow8=2.491e-4; ow9=1; ow10=1.613e-5;

ow11=-3.656e-3; ow12=3.210e-5; ow13=-4.232e-7; ow14=1.1109e-10;

ow15=-3.824e-12; ow16=6.938e-14; ow17=-5.096e-16; ow18=1; ow19=2.282e-5;

ow20=-7.638e-4; ow21=-7.760e-6; ow22=1.105e-8; ow23=0.18252;

ow24=-1.4619e-3; ow25=2.093e-5; ow26=-1.282e-7; ow27=2.033e-2;

ow28=1.266e-4; ow29=2.464e-6; ow30=1.849e-5; ow31=-2.551e-7;
ow32=2.551e-8;

%coefficients for static water dielectric constant eps_w0 [1]

BERE1=1.2748; BERE2=-0.519; BERE3=-0.152;

%components for exponent beta (exponent complex function in EPS_s), real
```

```matlab
%part [-]

BEIM1=1.338; BEIM2=-0.603; BEIM3=-0.166;

%components for exponent beta (exponent complex function in EPS_s),

%imaginary part [-]

eps_0=8.854187817e-12; %permittivity of free space [F/m]

alpha=0.65; %exponent in eps_s (local use) [-]

CPA1=1.01; CPA2=0.44; CPA3=-0.062; %coefficients for eps_pa,
respecitively

%[F/m]^1/2, [F m2/g]1/2, [F/m]


%calculation of dielectric constant of area (eps_fw, real and imaginary

%part. For more information, see appendix A

sig_eff=SGEF1+SGEF2*rho_b+SGEF3*S+SGEF4*C;

eps_pa=(CPA1+CPA2*rho_s)^2+CPA3;


eps_s0=ow5+ow6*T_g+ow7*T_g^2+ow8*T_g^3;

a_ST=ow9+ow10*S_g*T_g+ow11*S_g+ow12*S_g^2+ow13*S_g^3;

b_ST=ow18+ow19*S_g*T_g+ow20*S_g+ow21*S_g^2+ow22*S_g^3;


if  S_g<100;  %this boundary might in reality lay somewhere else. This
is not important for this calculation, because in the research area,
salinity is not an issue

    eps_w0=ow1+ow2*T_g+ow3*T_g^2+ow4*T_g^3;

    t_relax=(ow14+ow15*T_g+ow16*T_g^2+ow17*T_g^3);

    w='fresh';

elseif S_g>=100;

    eps_w0=eps_s0*a_ST;

    t_relax=(((ow14+ow15*T_g+ow16*T_g^2+ow17*T_g^3))*b_ST);

    w='salt';

end
```

```
eps_fwr=eps_winf+((eps_w0-eps_winf)/(1+(2*pi*f*t_relax)^2));

eps_fwi=((2*pi*f*t_relax*(eps_w0-eps_winf))/(1+(2*pi*f*t_relax)^2))

+((sig_eff/(2*pi*f*eps_0))*((rho_s-rho_b)/(rho_s*SM)));

beta_r=(BERE1+BERE2*S+BERE3*C);

beta_i=(BEIM1+BEIM2*S+BEIM3*C);

eps_b=(1+(rho_b/rho_s)*(eps_pa^alpha-1)+SM^beta_r*eps_fwr^alpha-SM)

^(1/alpha)-i*((SM^beta_i*eps_fwi^alpha)^(1/alpha));


%calculation reflectivity and emissivity. 2 In r_b2, r_g2, tau_s2 and
%omega2 stands for vertical polarization

r_b2=abs((eps_b*cos(theta)-sqrt(mu_s*eps_b-sin(theta)^2))

/(eps_b*cos(theta)+sqrt(mu_s*eps_b-sin(theta))));

r_g2=((1-QR)*r_b2+QR*r_b2)*exp(-HR*cos(theta)^NR_p);

e_g2=1-r_g2;


%parameters for calculation R1 and R2

Omega2=0; %single scattering albedo, default value low vegetation
          %according to ATBD [-]


%parameters for calculating contribution litter to optical depth

a_L=2.33;  %parameters used to compute the litter moisture content
(Mg_L)

b_L=0;     %from SM SM, default values according to ATBD [-]

Bs_L=0.3;  %dry biomass of litter, default value according to ATBD
[kg/m^2]

c_L=0.24;  %attenuation coefficient of litter, default value according
to ATBD [m2/kg]


%parameter adaptation optical depth to polarization
```

```matlab
tt_V=1; %angular correction parameters of optical depth to incidence
angle for V polarization, default value for low vegetation according to

     ATBD [-]

%optical depth without litter contribution

tau_s=tau_s(n); %the actual formula (see appendix A) is not used because

               %value comes directly from L2PP processor


%adaptation optical depth to polarization

tau_s2=tau_s*(sin(theta)^2*tt_V+cos(theta)^2);


%contribution litter to optical depth

Mg_L1=a_L*SM+b_L; %Mg_L=litter moisture content, =<0.8 [kg/kg]

if Mg_L1>0.8

    Mg_L=0.8;

elseif Mg_L1<0

    Mg_L=0

else Mg_L=a_L*SM+b_L;

end

LWC=(Mg_L/(1-Mg_L))*Bs_L; % LWC=litter water content [kg/m^2]

tau_L=c_L*LWC;

tau_c2=tau_s2+tau_L; %in case of no litter, tau_L is 0


%calculation R1 and R2 for vertical polarization

R1V=r_g2*exp(-2*tau_c2);

R2V=e_g2*(T_g)*exp(-tau_c2)+(T_c)*(1-omega2)*(1-exp(-tau_c2))

*(1+r_g2*exp(-tau_c2));


R_lv=[R1V;R2V];
```

COMBIH/COMBIV

```matlab
function TBsat_H=combiH(SM) %for vertical polarization, TBsat_H is
replaced by TB_satV and combiH by combiV

TBsat_H=zeros(1,1);

theta=0.74176;

%parameters that depend on location

T0=288;  %value taken from L2PP processor [K]

P0=1013; %value taken from L2PP processor [mbar]

WVC=1;   %total precipiable water vapour content, from ECMWF data
[kg/m^2]


%parameters that do not depend on location

k0_DT_H2O=10.084; k1_DT_H2O=0.00021; k2_DT_H2O=0.025241; %coefficients
for

%H2O layer temperature difference

k0_DT_O2=-3.16387e+0; kT0_DT_O2=1.38628e-1; kP0_DT_O2=3.29731e-3;

kT02_DT_O2=-1.19886e-4; kP02_DT_O2=1.66366e-6; kT0P0_DT_O2=-9.90743e-6;

%coefficients for O2 layer temperature difference

k0_tau_H2O=-107.66; k1_tau_H2O=0.1149; k2_tau_H2O=2.0983;

%coefficients for H2O optical thickness

k0_tau_O2=5.12341e3; kT0_tau_O2=-6.80605e1; kP0_tau_02=2.42216e1;

kT02_tau_O2=1.70616e-1; kP02_tau_O2=6.64682e-3; kT0P0_tau_O2=-7.99404e-
2;

%coefficients for O2 optical thickness


%calculation to determine TB contribution and optical depth atmosphere

DT_O2=k0_DT_O2+kT0_DT_O2*T0+kP0_DT_O2*P0+kT02_DT_O2*T0^2+kP02_DT_O2*P0^2

+kT0P0_DT_O2*T0*P0;

tau_O2=1e-6*(k0_tau_O2+kT0_tau_O2*T0+kP0_tau_02*P0+kT02_tau_O2*T0^2

+kP02_tau_O2*P0^2+kT0P0_tau_O2*T0*P0)/cos(theta);

DT_H2O=k0_DT_H2O+k1_DT_H2O*P0+k2_DT_H2O*WVC;
```

169

```
tau_H2Ox=1e-6*(k0_tau_H2O+k1_tau_H2O*P0+k2_tau_H2O*WVC)/cos(theta);

tau_H2O=max(tau_H2Ox,0);

TB_O2=(T0-DT_O2)*tau_O2;

TB_H2O=(T0-DT_H2O)*tau_H2O;

TB_atm=TB_O2+TB_H2O;

tau_atm=tau_O2+tau_H2O;


%TB contribution sky

TB_sk=5; %value is threshold from TGRD. The ATBD mentiones how to
calculate the exact value, but not enough data is available for this

calculation


%Determine R1 and R2 as the sum of several fractions

R_lv=lowvegH(SM); %in case there area other fractions, these can be

                  %mentioned here and added according to the weighing

                  %function. For vertical polarization, lowvegH is
replaced by lowvegV

R_lv1=R_lv(1,1); %1 and 2 do not stand for horizontal and vertical here

R_lv2=R_lv(2,1);

R1=R_lv1;

R2=R_lv2;


%calculation TB for polarization

TBsat_H=TB_atm+exp(-tau_atm)*(TB_atm+TB_sk*exp(-tau_atm))*R1+exp(-
tau_atm)

*R2; %for vertical polarization, TBsat_H is replaced by TB_satV
```

ITERATION

```
function retr=iteration(SM)

retr=zeros(1,1);
```

170

```matlab
n=1;

a(1)=15; %contribution to horizontal TB to make outcome code comparable
to

a(2)=20; %that of the L2PP processor, as explained in chapter 3

a(3)=0;

a(4)=-5;



b(1)=10; %contribution to horizontal TB to make outcome code comparable
to

b(2)=15; %that of the L2PP processor, as explained in chapter 3

b(3)=5;

b(4)=-5;



TB_sat_H=combiH(SM)+a(n);

TB_sat_V=combiV(SM)+b(n);



TB_mH(1)=261.965;    %data from L1c file, m stands for measured

TB_mH(2)=249.0576;

TB_mH(3)=220.91046;

TB_mH(4)=220.91046;



TB_mV(1)=284.1501;   %data from L1c file, m stands for measured

TB_mV(2)=282.98574;

TB_mV(3)=259.00986;

TB_mV(4)=259.00986;



TB_mH=TB_mH(n);

TB_mV=TB_mV(n);



retrH=TB_sat_H-TB_mH;
```

Validation of SMOS satellite data over Ghana and Burkina Faso

```
retrV=TB_sat_V-TB_mV;
```

```
retr=sqrt(retrH^2+retrV^2);
```

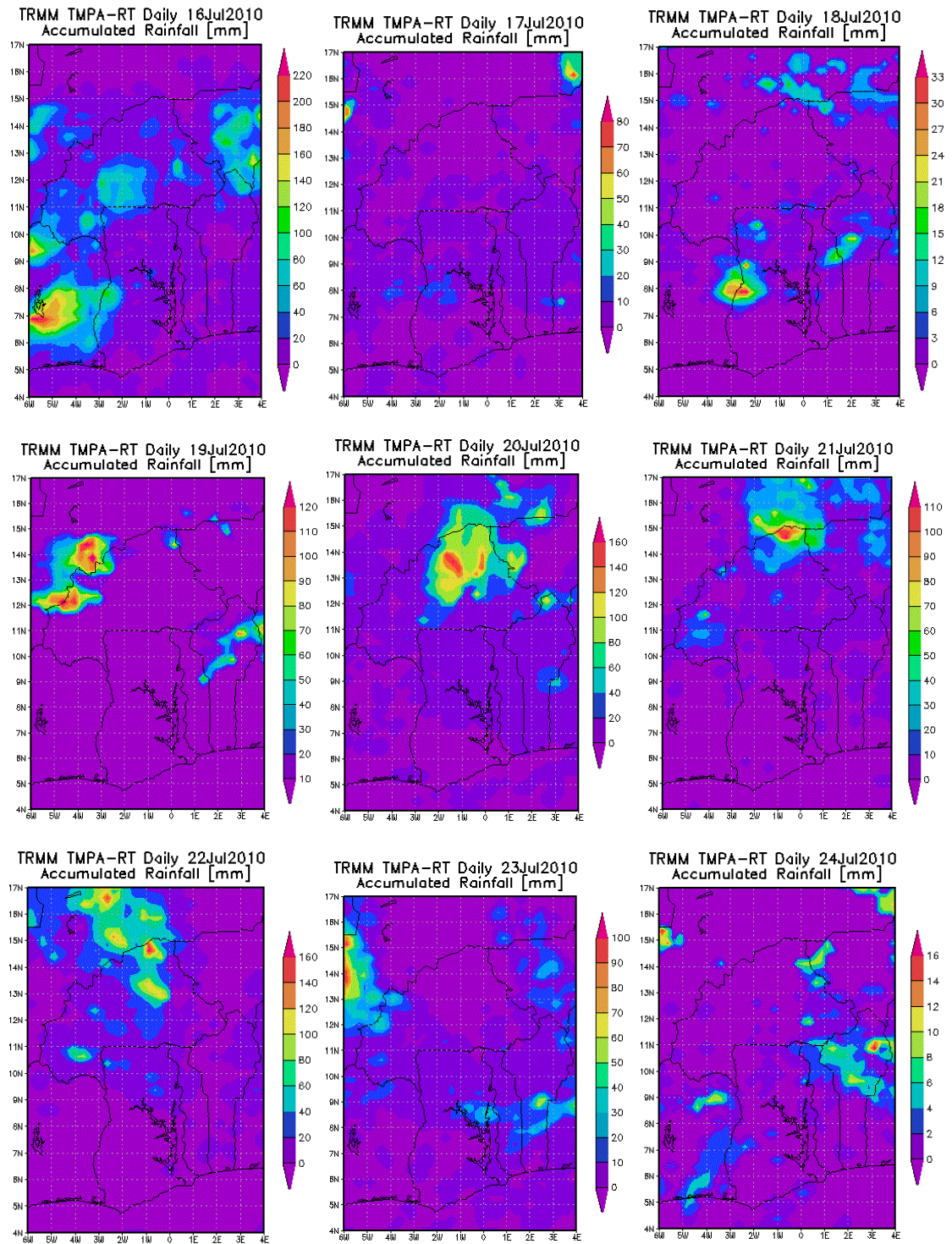# E COMPARING SMOS TO OTHER DATASETS OVER TIME

INTRODUCTION

This appendix gives extra information on validation of chapter 6. It starts with information on the accumulated rainfall in July 2010 over the research area. A map with the locations of the pixels shown in paragraph 6.3.3 is provided. After that, scattering plots of specific parts of the research area for the three satellite datasets compared to SMOS data are given. The appendix is concluded with graphs of the RMSE, MD, Pearson and Nash coefficient and the count for the complete research area are given.

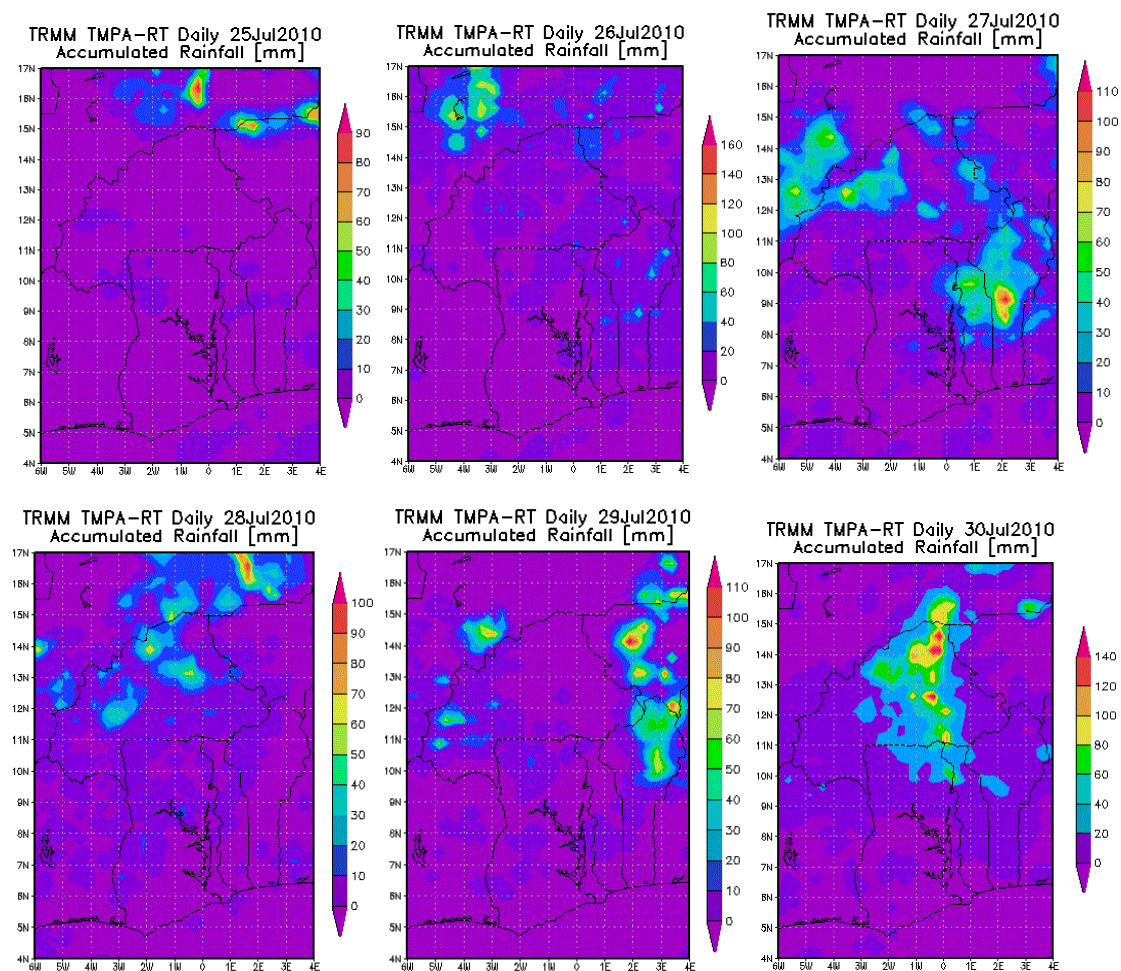## ACCUMULATED RAINFALL IN JULY 2010 OVER GHANA AND BURKINA FASO

Figure E.1a-u: Accumulated rainfall over research area from July 10th to July 30th

Accumulated rainfall in the period from July 11th to August 5th ranges from 0 to 220 mm per day. Most of the time, the largest amount of rain falls in a small area, while the remainder of the area receives 0 to maximum 40 mm. On July 11th, large parts of Burkina Faso and the north east of Ghana get 30 mm rain. Two small parts, in the east and west of Burkina Faso, receive up to 80 mm of rain. On July 13th, rain falls in Ghana, mostly around the Volta Lake. On July 14th, most rain falls in the west of the area. Locally, this can add up to 140 mm. On the mid-west border of Ghana, 220 mm falls. In the east, small amounts of precipitation fall, mostly in the south. On July 15th, small amount of rain fall in the south and sporadically in Burkina Faso. The mid-west border of Ghana is again hit with a lot of rain, up to 180 mm. On July 16th, the whole area is hit with rain up to 40 mm. In Burkina Faso, large areas receive 120 mm. To the west of Ghana, rainfall goes up to 220 mm. On July 17th, Ghana and southern Burkina Faso receives up to 20 mm, with in the mid-west an area with 40 mm of rain. The north of Burkina Faso remains dry. On July 18th, the north of Ghana receives a little rain, with a peak in the west of 33 mm. On July 19th, the north-west of Burkina Faso receives 120 mm. From July 20th to 22nd, the middle of Burkina Faso receives a lot of rain (up to 160 mm per day) On July 23rd, most rain falls to the north-west of Burkina Faso and to the east of Ghana. The rest of the south gets for the most part 20 mm. From

July 24th to July 27th, most rain falls on July 27th, over Burkina Faso and to the east of Ghana up to 110 mm. On July 28th, the middle of Burkina Faso receives up to 80 mm. The rest of the area receives up to 20 mm, with no rain in the south of Ghana. On July 29th, Ghana receives 20 mm in the middle and north. The west of Burkina Faso receives up to 70 mm. To the east of Burkina Faso, up to 110 mm falls. On July 30th, the north of the area receives rain up to 20 mm, with in mid-Burkina Faso up to 140 mm. On July 30th and August 1st, little rain falls, except for the middle of Burkina Faso on August 1st, where up to 120 mm falls. On August 2nd, again the north is hit, with spots with rain up to 100 mm to the north-west and south-east and in the middle Burkina Faso. On August 3rd, a belt of rain spots on the border of Ghana and Burkina Faso is found with rainfall up to 80 mm. On August 4th, on the same belt, rainfall is up to 20 mm, with in the east of Burkina Faso some rainfall up to 90 mm. On August 5th finally, hardly any rain falls in the area.
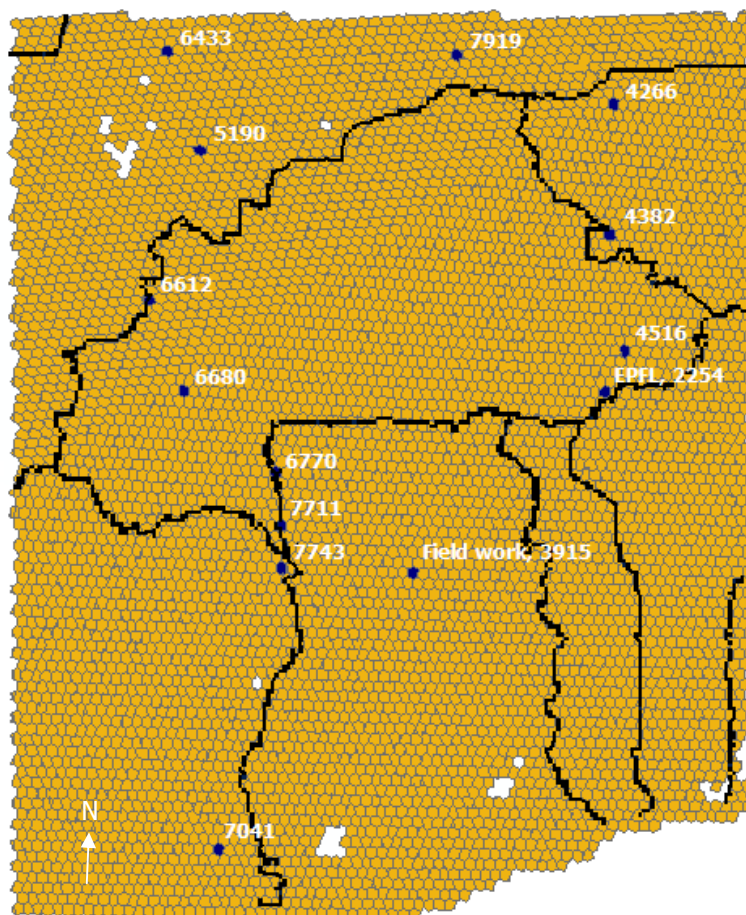
LOCATION OF PIXELS PARAGRAPH 6.3.3

Figure E.2: Locations pixels studies in chapter 6

SCATTERING PLOTS

176

The area is divided into 9 equally sized areas; North West, North, North East, East, Middle, West, South West, South, South East. Scattering plots were created of these areas in order to get a feeling for the scattering due to location. The lowest row of areas has the lushest vegetation and most precipitation. Going north, vegetation changes from forest to savanna and changes the precipitation regime as described in chapter 2. It should be noted that this means that the count of some areas is far lower than other areas.
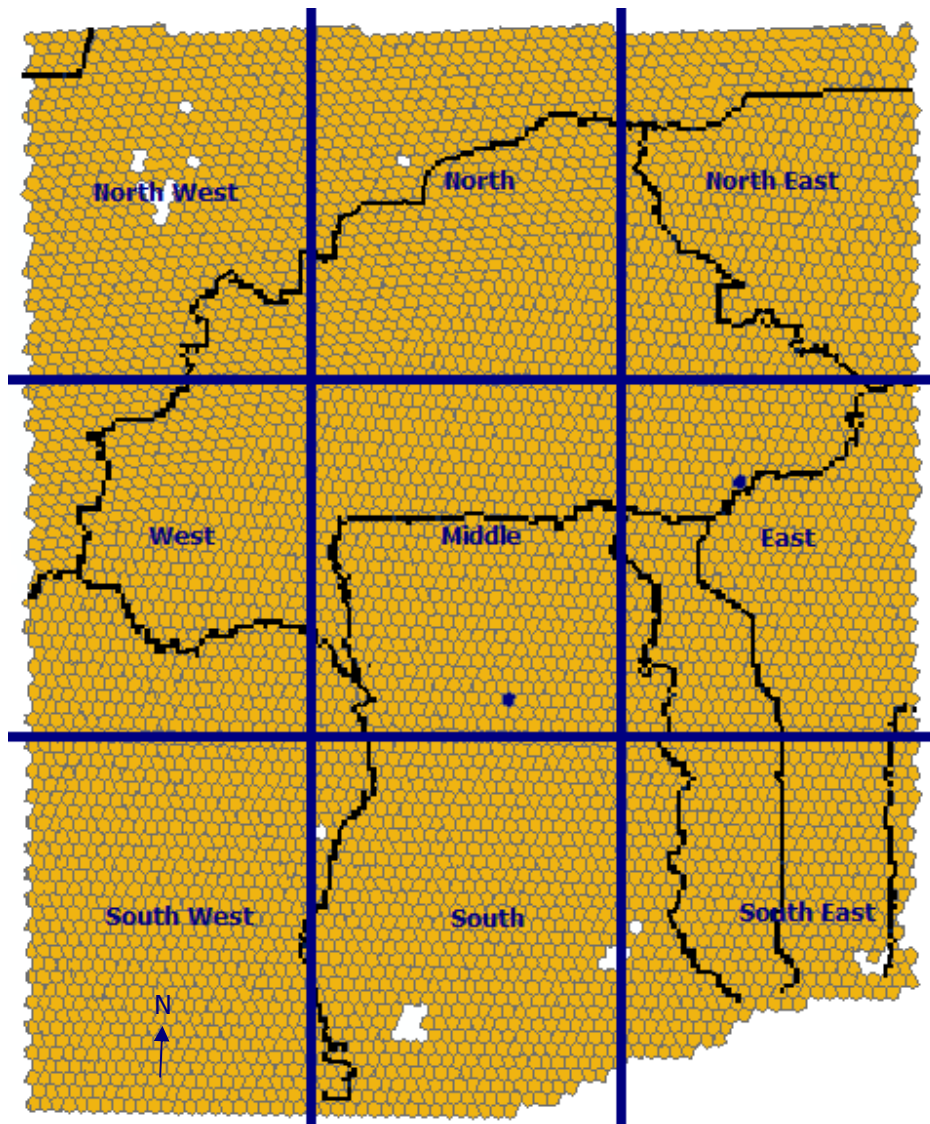

Figure E.3: Division of research area into 9 different parts for scatter plots

177

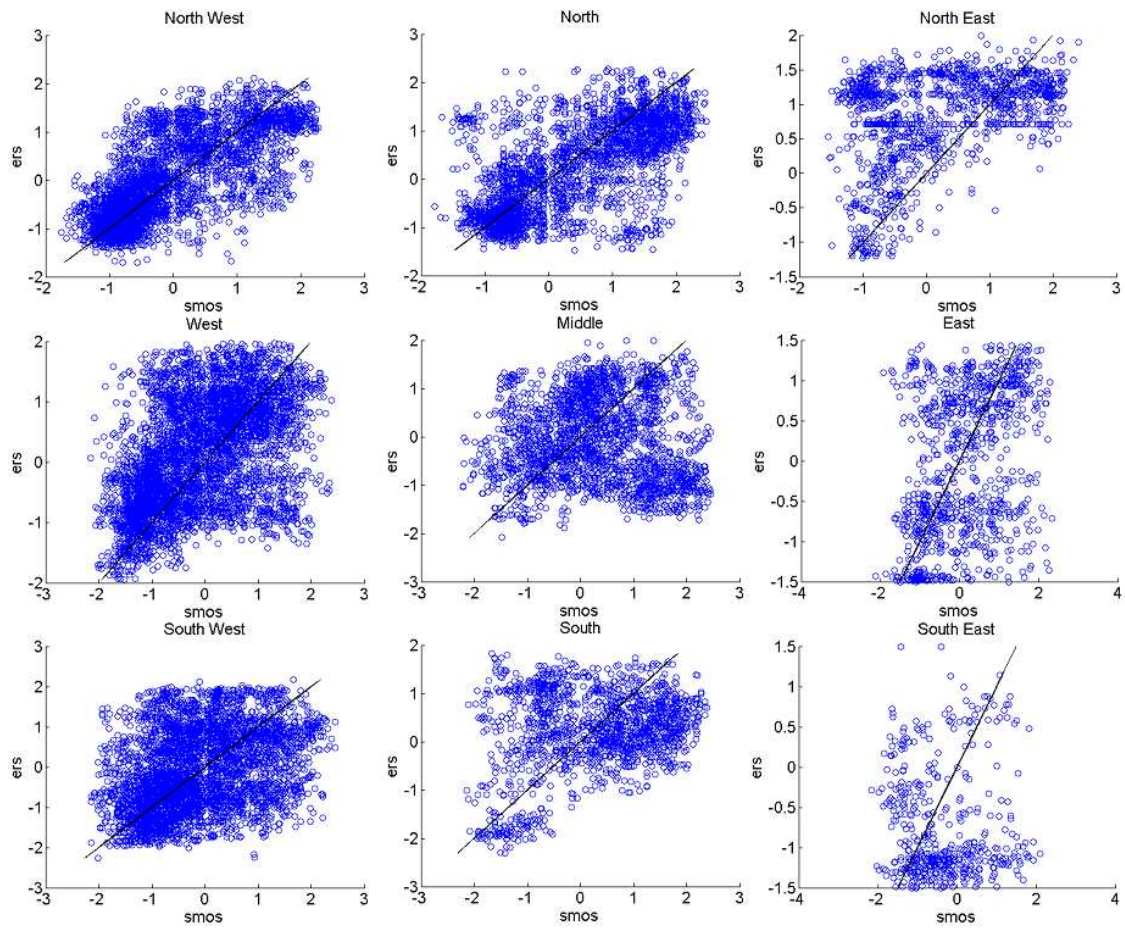Validation of SMOS satellite data over Ghana and Burkina Faso

Figure E.4: Scatter plots ERS vs. SMOS for nine different parts of research area

Table E.1: r^2 for scatter plots ERS vs. SMOS

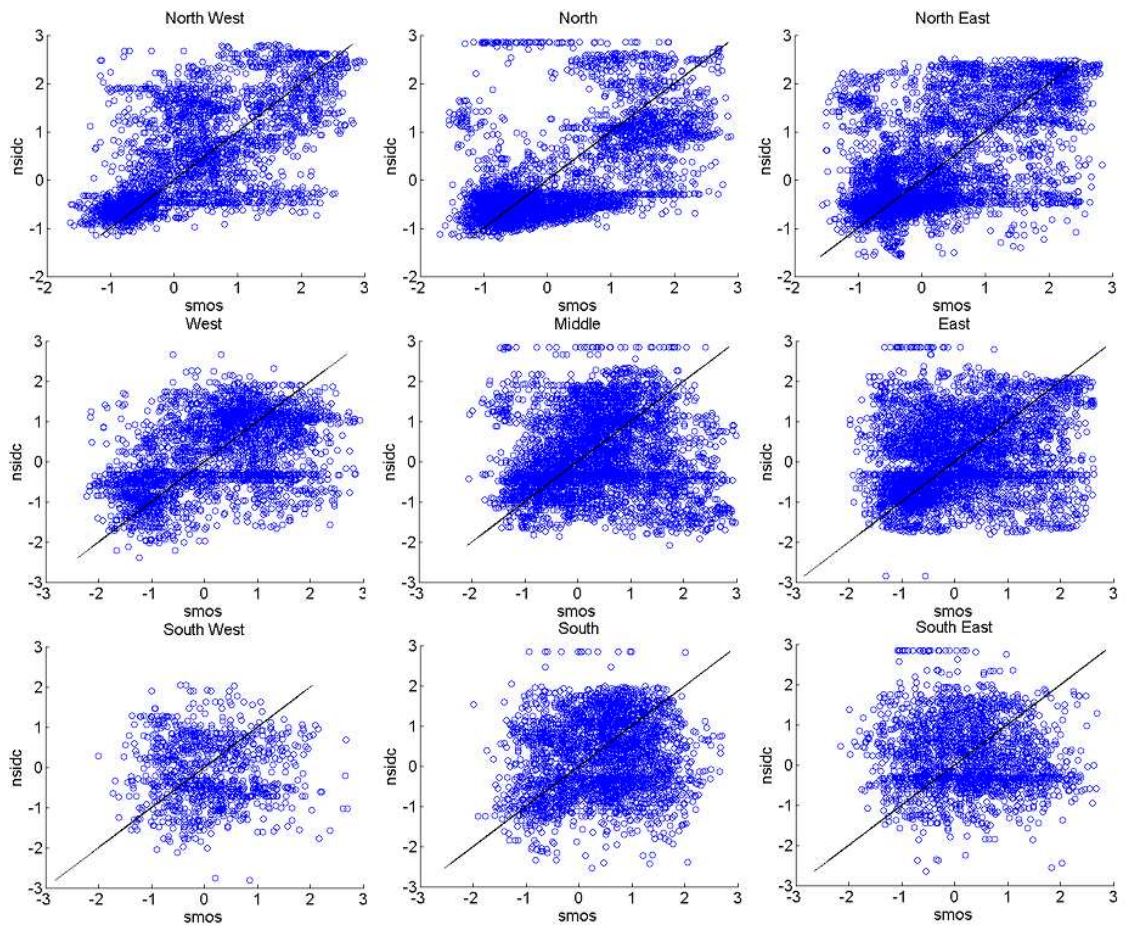| r^2 ERS		
0.27	0.44	NaN
NaN	NaN	NaN
NaN	NaN	NaN

Figure E.5: Scatter plots NSIDC vs. SMOS for nine different parts of research area

Table E.2: r^2 for scatter plots NSIDC vs. SMOS

r^2 NSIDC		
0.3807	0.1278	0.2267
NaN	NaN	NaN
NaN	NaN	NaN

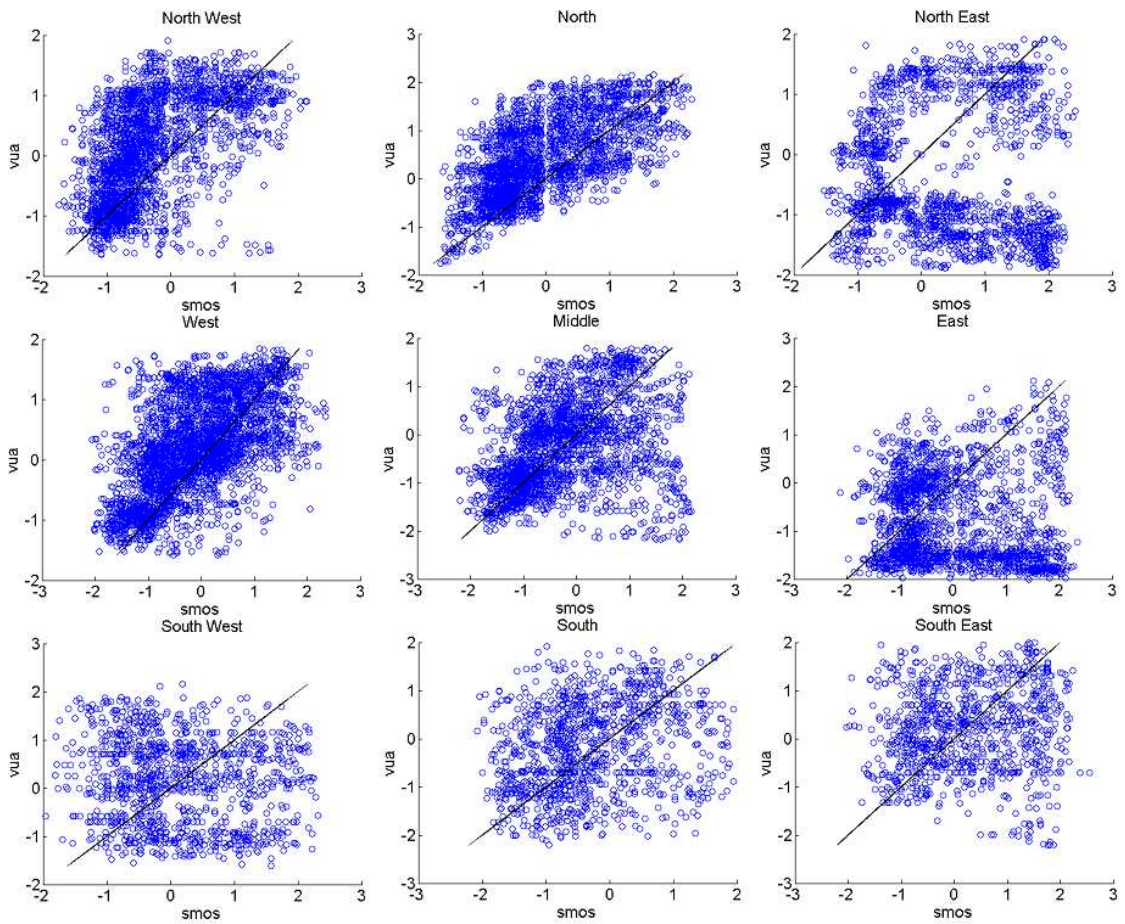Validation of SMOS satellite data over Ghana and Burkina Faso

Figure E.6: Scatter plots VUA vs. SMOS for nine different parts of research area

Table E.3: r^2 for scatter plots VUA vs. SMOS

r^2 VUA		
NaN	0.13	NaN
NaN	NaN	NaN
NaN	NaN	NaN

COUNT, MD, RMSE, PEARSON COEFFICIENT, NASH AND PEARSON
COEFFICIENT

a) ERS

b) NSIDC

c) VUA

Legend
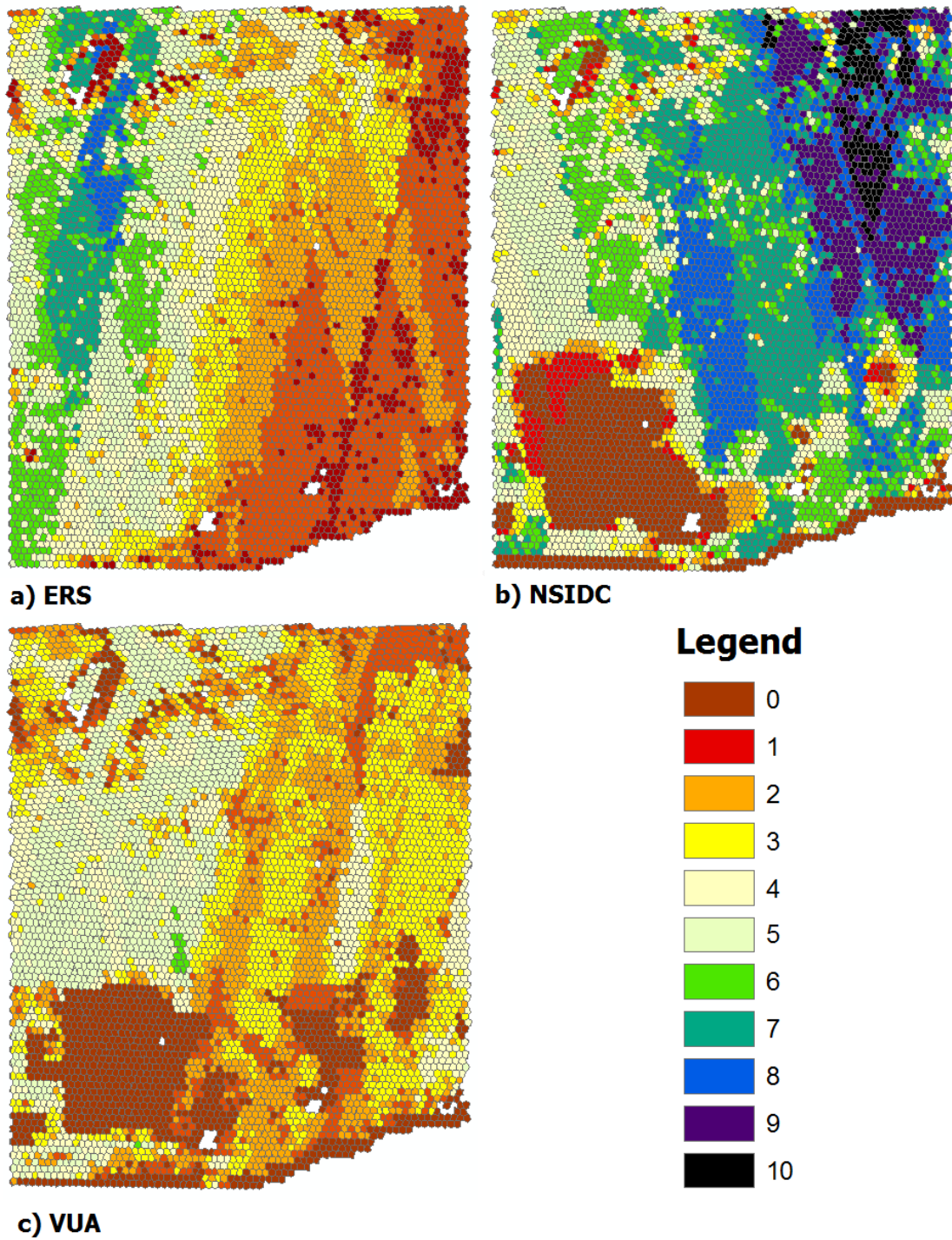
Figure E.7: Count for comparisons a) ERS-SMOS, b) NSIDC-SMOS, c) VUA-SMOS

a) ERS

b) NSIDC

c) VUA

Legend

▮	-0,40 - -0,30
▮	-0,29 - -0,30
▮	-0,29 - -0,20
▮	-0,19 - -0,10
▮	-0,09 - 0,00
▮	0,01 - 0,10
▮	0,11 - 0,20
▮	0,21 - 0,30
▮	0,31 - 0,40
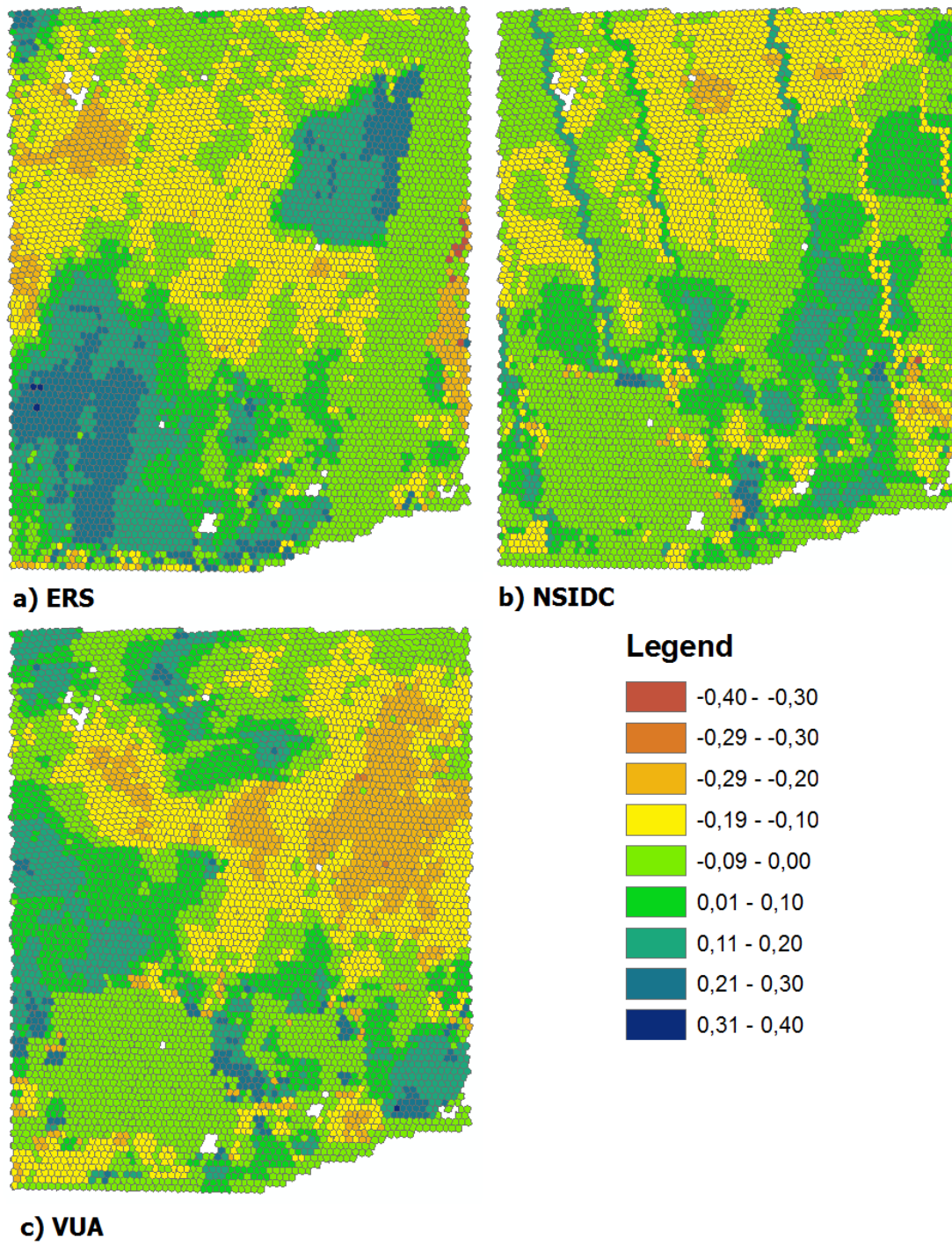
Figure E.8: MD for comparisons a) ERS-SMOS, b) NSIDC-SMOS, c) VUA-SMOS

a) ERS

b) NSIDC

c) VUA

Legend

< -2,00	-0,59 - -0,40
-1,99 - -1,80	-0,39 - -0,20
-1,79 - -1,60	-0,19 - 0,00
-1,59 - -1,40	0,01 - 0,20
-1,39 - -1,20	0,21 - 0,40
-1,19 - -1,00	0,41 - 0,60
-0,99 - -0,80	0,61 - 0,80
-0,79 - -0,60	0,81 - 1,00
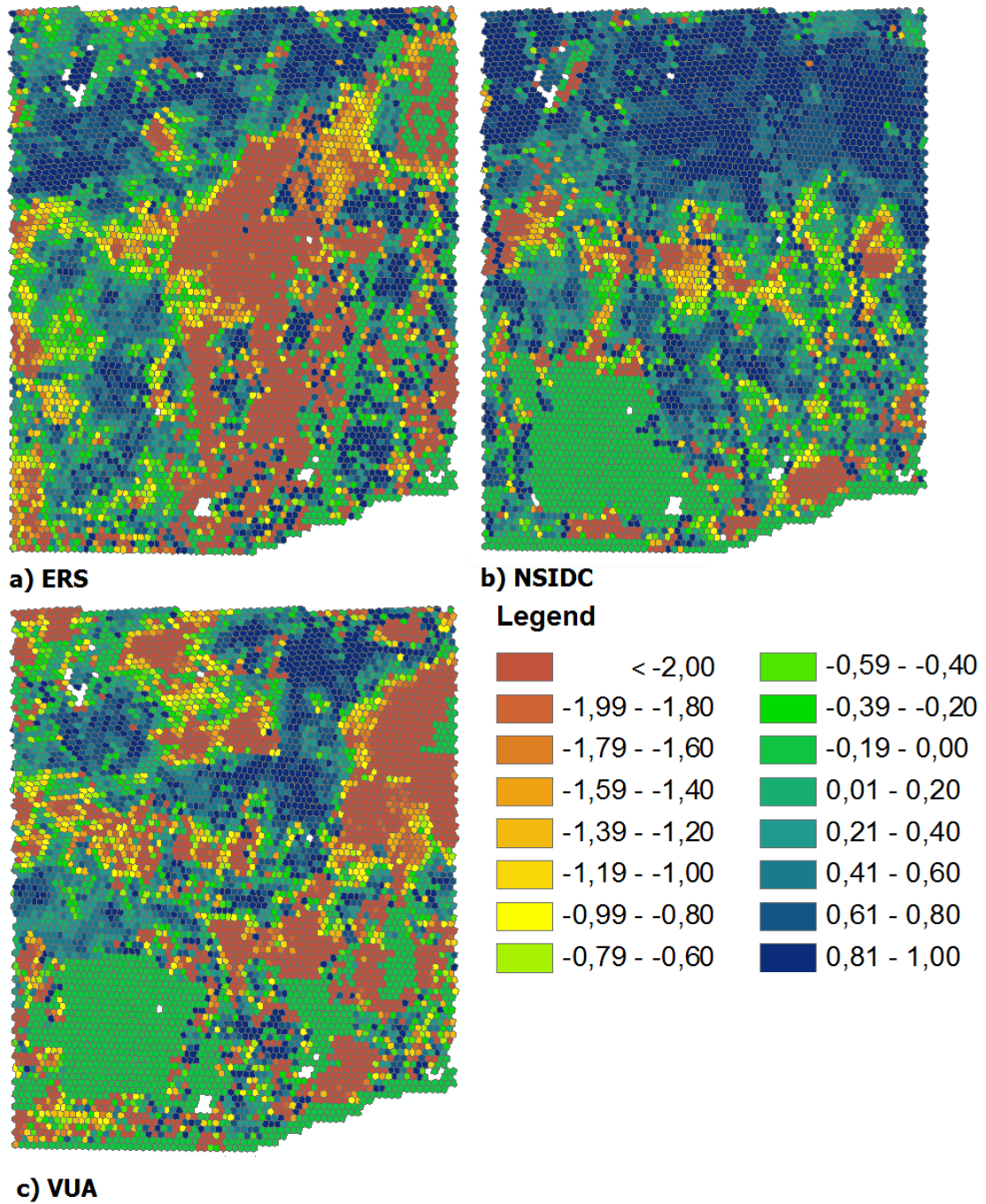
Figure E.9: RMSE for comparisons a) ERS-SMOS, b) NSIDC-SMOS, c) VUA-SMOS

a) ERS

b) NSIDC

c) VUA

Legend

	-1,00 - -0,80
	-0,79 - -0,60
	-0,59 - -0,40
	-0,39 - -0,20
	-0,19 - 0,00
	0,01 - 0,20
	0,21 - 0,40
	0,41 - 0,60
	0,61 - 0,80
	0,81 - 1,00
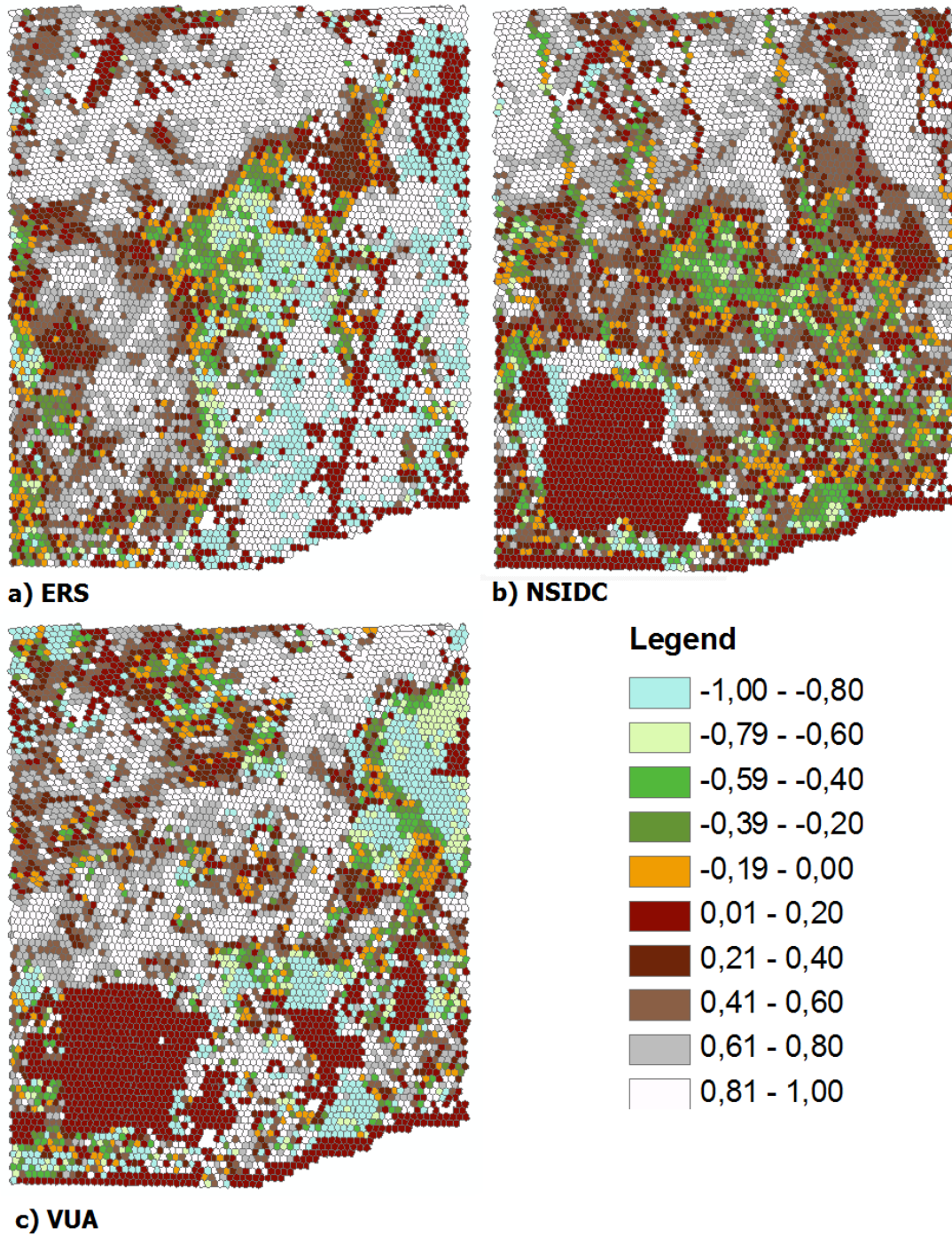
Figure E.10: Pearson for comparisons a) ERS-SMOS, b) NSIDC-SMOS, c) VUA-SMOS

a) ERS

b) NSIDC

c) VUA

Legend

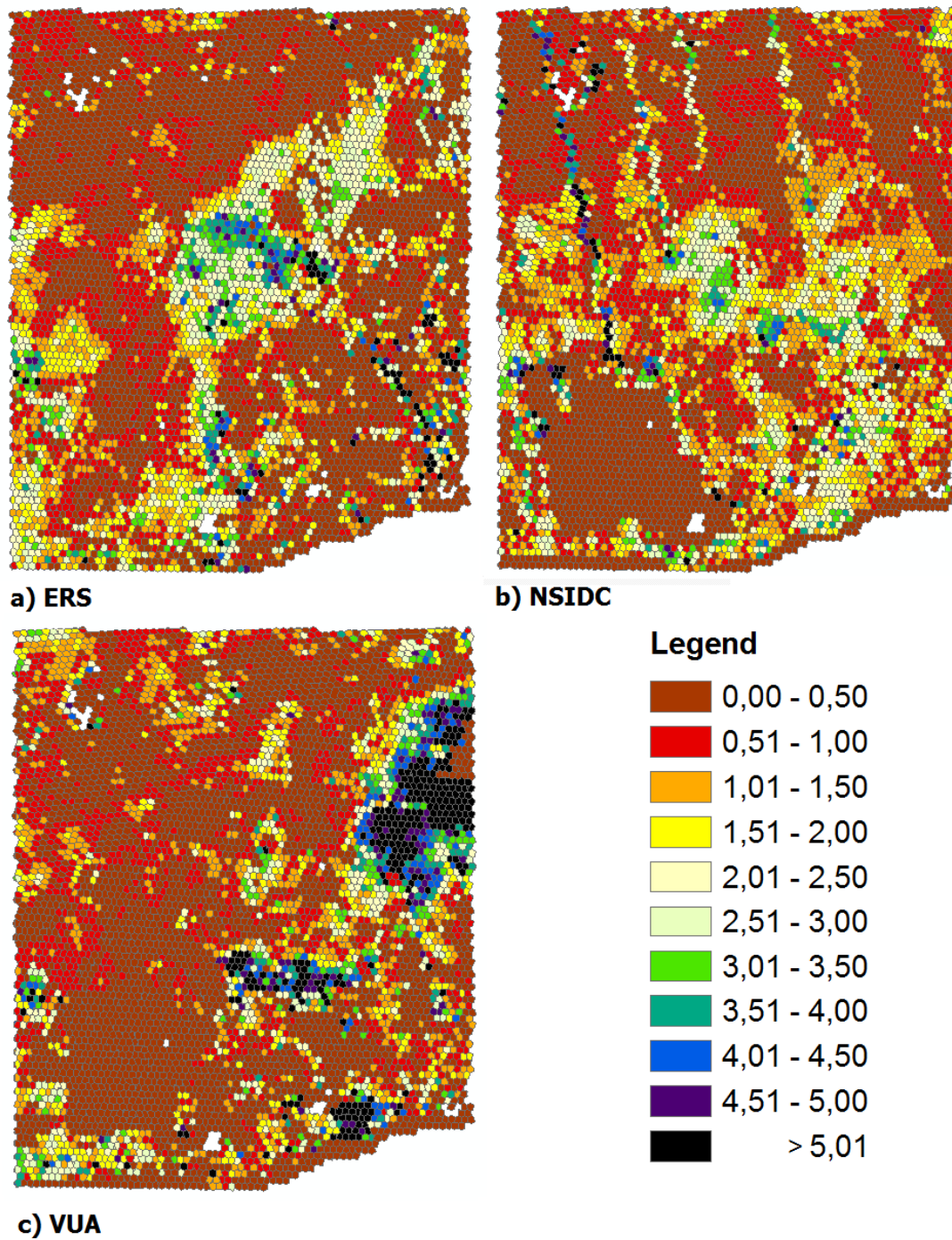|	
	0,00 - 0,50
	0,51 - 1,00
	1,01 - 1,50
	1,51 - 2,00
	2,01 - 2,50
	2,51 - 3,00
	3,01 - 3,50
	3,51 - 4,00
	4,01 - 4,50
	4,51 - 5,00
	> 5,01

Figure E.11: Nash for comparisons a) ERS-SMOS, b) NSIDC-SMOS, c) VUA-SMOS