

Die approbierte Originalversion dieser Dissertation ist an der Hauptbibliothek der Technischen Universität Wien aufgestellt (<http://www.ub.tuwien.ac.at>).

The approved original version of this thesis is available at the main library of the Vienna University of Technology (<http://www.ub.tuwien.ac.at/englweb/>).

DISSERTATION

# Error characterization methods for surface soil moisture products from remote sensing

A thesis submitted in partial fulfilment of the academic degree of “Doktorin der Naturwissenschaften” \*)  
under the supervision of

Univ.Prof. Dipl.-Ing. Dr.techn. Wolfgang Wagner  
E 122  
Institute of Photogrammetry and Remote Sensing

research conducted at the Vienna University of Technology  
Faculty of Mathematics and Geoinformation

by

**Marcela Doubková**

Matriculation number: 0827831  
Obkirchergasse 24/16a  
1190 Wien  
Austria

Wien, 17<sup>th</sup> of October, 2012



\*) comparable to the Doctor in Natural Science

DISSERTATION

# Fehler-Charakterisierungsmethoden für Bodenfeuchtemessungen aus Fernerkundung

Ausgeführt zum Zwecke der Erlangung des akademischen Grades einer Doktorin der  
Naturwissenschaften unter der Leitung von

Univ.Prof. Dipl.-Ing. Dr.techn. Wolfgang Wagner  
E 122  
Institut für Photogrammetrie und Fernerkundung

eingereicht an der Technischen Universität Wien  
Fakultät für Mathematik und Geoinformation

von

**Marcela Doubková**

Matrikelnummer: 0827831  
Obkirchergasse 24/16a  
1190 Wien  
Austria

Wien, am 17 Oktober, 2012

Supervisor: Prof. Dr. Wolfgang Wagner  
Institute of Photogrammetry and Remote Sensing, Vienna University of Technology  
Gusshausstraße 27-29, 1040, Vienna, Austria. E-mail: ww@ipf.tuwien.ac.at

Reviewer: Prof. Dr. Günter Blöschl  
Institute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology  
Karlsplatz 13/222, A-1040 Vienna, Austria. E-mail: bloeschl@hydro.tuwien.ac.at

*Parts of the present work have already been incorporated in several peer-reviewed journal papers and reports which include:*

ESA CCI Soil Moisture: Comprehensive Error Characterisation Report, Version 0.7, 21 May 2012, Prepared by VUA, TU Wien, GeoVille, ETH Zürich, AWST, FMI, UCC and NILU.

Doubková, M., Van Dijk, A. I. J. M., Sabel, D., Wagner, W., & Blöschl, G. (2012). Evaluation of the predicted error of the soil moisture retrieval from C-band SAR by comparison against modelled soil moisture estimates over Australia. *Remote Sensing of Environment*, 120, 188–196.

*The work has been conducted as part of several research projects:*

SHARE (Soil Moisture For Hydrometeorological Applications In The SADC Region) and SHARE-Extension2, project of the Data User Element Innovator Tiger Initiative of the European Space Agency (ESA ESRIN Contract No. 19420/05/I-EC).

ESA CCI Soil Moisture, project of the ESA Climate Change Initiative, (ESA ESRIN Contract No. 4000104814/11/I-NB).

*and further supported by:*

The Austrian Science Fund (FWF) as part of the Doctoral Programme on Water Resource Systems (DK-plus W1219-N22).

The water information research and development alliance between the Bureau of Meteorology and CSIRO's Water for a Healthy Country Flagship.

## Abstract

To support the operational use of Synthetic Aperture Radar (SAR) earth observation systems, the European Space Agency (ESA) is developing Sentinel-1 radar satellites operating in C-band. Much like its SAR predecessors (Earth Resource Satellite, ENVISAT, and RADARSAT), the Sentinel-1 will operate at a medium spatial resolution (ranging from 5 to 40 m), but with a greatly improved revisit period, especially over Europe (~ 2 days). Given the planned high temporal sampling and the operational configuration Sentinel-1 is expected to be beneficial for operational monitoring of dynamic processes in hydrology and phenology. The benefit of a C-band SAR monitoring service in hydrology has already been demonstrated within the scope of the Soil Moisture for Hydrometeorologic Applications (SHARE) project using data from the Global Mode (GM) of the Advanced Synthetic Aperture Radar (ASAR).

To fully exploit the potential of the SAR soil moisture products, well characterized error needs to be provided with the products. Understanding errors of remotely sensed surface soil moisture (SSM) datasets was indispensable for their application in models, for extractions of blended SSM products, as well as for their usage in evaluation of other soil moisture datasets.

This thesis has several objectives. First, it provides the basics and state of the art methods for evaluating measures of SSM, including both the standard (e.g. Root Mean Square Error, Correlation coefficient) and the advanced (e.g. Error propagation, Triple collocation) evaluation measures. A summary of applications of soil moisture datasets is presented and evaluation measures are suggested for each application according to its requirement on the dataset quality. The evaluation of the Advanced Synthetic Aperture Radar (ASAR) Global Mode (GM) SSM using the standard and advanced evaluation measures comprises a second objective of the work. To achieve the second objective, the data from the Australian Water Assessment System (AWRA-L) hydrological model, OzNET in-situ stations, and several other coarse resolution data sources were used. The results are combined to provide an exhaustive estimate of all qualities of the ASAR GM SSM product. The third objective is to provide guidance on appropriate evaluation methodology applicable to any SSM product. For this purpose the results of the ASAR GM evaluation analyzed are discussed from a general perspective and restructured to answer scientific questions identified in the introductory part of the thesis. These include:

- Can we apply the evaluation requirements from comparable missions such as SMOS and SMAP to ASAR GM SSM?
- How does spatial resolution influence error estimates?
- Is there a single measure to describe the quality of SSM data?
- What is the quality and what are the limitations of ASAR GM SSM?
- Learning from ASAR GM SSM errors for Sentinel-1

The findings and suggestions originating from the discussion are transferable to other satellite-derived soil moisture data. Of special interest is its transfer to data from the planned Sentinel-1 SAR sensor that shares similar technical characteristics but has an improved retrieval error comparable to the ASAR GM sensor. The operationally available medium resolution soil moisture from Sentinel-1 with a well-characterized error is likely to yield benefits for modelling and monitoring of land surface-atmosphere fluxes, crop growth and water balance applications.

## Kurzfassung

Zur Unterstützung für den operationellen Einsatz von Erdbeobachtungssystemen wie Synthetic Aperture Radar (SAR) entwickelt die Europäische Weltraumbehörde ESA den Radar-Satelliten Sentinel-1, der im C-Band arbeitet. Ähnlich seinen SAR-Vorgängern auf den Plattformen ERS, ENVISAT, oder RADARSAT wird der Sensor Sentinel-1 bei einer mittleren räumlicher Auflösung im Bereich von 5 bis 40 m arbeiten, allerdings mit einer vielfach erhöhten Wiederholrate, die über Europa etwa im Bereich von ca. 2 Tagen liegen wird. Aufgrund dieser hohen zeitlichen Auflösung sowie dem operationellen Design wird der Satellit einen großen Beitrag zur Überwachung von dynamischen Prozessen in Hydrologie und Phänologie leisten. Der Nutzen von C-Band SAR-Überwachungssystemen in der Hydrologie wurde in der Vergangenheit bereits im Rahmen des Projects SHARE (Soil Moisture for Hydrometeorologic Applications) gezeigt, bei dem Daten des Instruments ASAR (Advanced Synthetic Aperture Radar) im Global Mode (GM) verwendet wurden (Doubkova et al., 2009). Um das volle Potential von SAR-Produkten im Bereich Bodenfeuchtigkeit auszuschöpfen, ist ein mitgeliefertes Fehlermaß unerlässlich. Das Verständnis dieses Fehlermaßes ist unentbehrlich für die Anwendung von Bodenfeuchteprodukten in Modellen, die Extraktion oder Erstellung von Produkten, sowie den Vergleich mit anderen Bodenfeuchteprodukten geht.

Die vorliegende Arbeit umfasst mehrere Ziele. Erstens werden Grundlagen sowie ein aktueller Stand der Technik im Bereich der Fehlermaße von Bodenfeuchteprodukten dargestellt, wie etwa dem quadratischen mittleren Fehler, Korrelationskoeffizienten oder erweiterten Methoden wie Fehlerfortpflanzung und triple collocation. Des Weiteren wird eine Übersicht der Anwendungsbereiche von Bodenfeuchteprodukten präsentiert und Evaluierungsmethoden je nach Bereich und Qualitätsanforderung vorgeschlagen. Die Evaluierung von ASAR GM Bodenfeuchteprodukten mit eben diesen Methoden stellt ein zweites Ziel dieser Arbeit dar. Um dies zu erreichen wurden Daten des australischen hydrologischen Modells Water Assessment System (AWRA-L), Feldmessungen des australischen Netzwerkes OzNET, sowie weitere grob aufgelöste Fernerkundungsdaten verwendet um die Qualität der Bodenfeuchteprodukte umfassend zu beschreiben. Die dritte Zielsetzung dieser Arbeit ist das Bereitstellen von Richtlinien für eine Evaluierungsmethode, die auf beliebige Bodenfeuchteprodukte angewendet werden kann. Zu diesem Zweck wurden die ASAR GM Ergebnisse vor einem breiteren Hintergrund analysiert um folgende Fragestellungen zu beantworten:

- Ist es möglich die Qualitätsanforderungen von vergleichbaren Missionen wie SMOS oder SMAP auf ASAR GM Bodenfeuchteprodukte zu übertragen?
- Wie beeinflusst die räumliche Auflösung die Fehlerabschätzung?
- Gibt es ein einziges Maß für die Qualitätsbeschreibung von Bodenfeuchteprodukten?
- Wie ist die Qualität von ASAR GM Bodenfeuchtigkeitsprodukten und wo liegen Einschränkungen?
- Wie ist die Qualität von Sentinel-1 Bodenfeuchtigkeitsprodukten und wo liegen Einschränkungen?

Diese Arbeit liefert Antworten und Ergebnisse, die auch auf weitere Satelliten-basierte Bodenfeuchteprodukte angewendet werden können. Besonders die Übertragung auf den geplanten Sentinel-1 Sensor ist von besonderem Interesse da dieser Sensor zwar ähnliche technische Eigenschaften, aber ein verbessertes Fehlermaß im Vergleich zu ASAR GM besitzt. Die operationell verfügbaren Bodenfeuchteprodukte von Sentinel-1 werden wesentlich zur Modellierung und Beobachtung von Land-Atmosphäre Interaktionen, Ernteertrag sowie Anwendungen im Bereich der Wasserbilanz beitragen.

## Acknowledgment

This work would not have happened without the unending support of the team at the Institute of Photogrammetry and Remote Sensing (IPF) at the Vienna University of Technology (TU WIEN). A special thanks goes to Daniel Sabel and Stefan Hasenauer who supported and motivated my work throughout five years at IPF and especially in the last few months when writing this thesis. Further thanks go to Ewelina Rupnik who has, within a short time, become a perfect colleague and friend who always brought energy to the office. Also thanks to Wouter Dorigo for the scientific discussions that spiced up this work, and to Alena Hegyova for the efficient programming support.

Thank you to all who found, even in the last moments, time to proofread this work: Carl Salk, Michael Cwach, Elin Högström, and Mirela Tulbure.

A Big Thank You goes to my supervisors – Professor Wagner and Professor Blöschl – who, even under time pressure, invested their evenings to provide me with comments that significantly improved the quality of this thesis. In addition, thanks to the trust and responsibilities given by Professor Wagner during my stay at IPF that prepared me to master this thesis.

Last but not least, thank you to my family and my partner Robert Permann who have been so understanding during the last busy year.

<b>1. INTRODUCTION</b> .....	<b>1</b>
1.1 EVALUATION OF SOIL MOISTURE DATASETS .....	1
1.2 OBJECTIVE AND STRUCTURE.....	4
<b>2. THEORY</b> .....	<b>7</b>
2.1 PRE-PROCESSING STEPS PRIOR TO DATA EVALUATION.....	7
2.1.1 <i>Statistical terminology</i> .....	7
2.1.2 <i>Preprocessing steps for soil moisture evaluation</i> .....	9
2.2 EVALUATION MEASURES AND METHODS.....	13
2.2.1 <i>Standard evaluation measures</i> .....	13
2.2.2 <i>Advanced evaluation methods</i> .....	18
2.2.3 <i>The next stage: how to evaluate quality of the evaluation measures?</i> .....	22
2.3 THE SELECTION OF APPROPRIATE EVALUATIONS MEASURES AND METHODS.....	25
2.3.1 <i>Sequential data assimilation</i> .....	27
2.3.2 <i>A direct input into Environmental Process models</i> .....	30
2.3.3 <i>Distinguishing between different soil moisture levels</i> .....	31
2.3.4 <i>Anomaly computation</i> .....	32
2.3.5 <i>Conclusion</i> .....	34
<b>3. METHODOLOGY</b> .....	<b>35</b>
3.1 DATASETS.....	35
3.1.1 <i>ASAR GM soil moisture dataset and processing</i> .....	35
3.1.2 <i>The AWRA-L landscape hydrological model</i> .....	37
3.1.3 <i>AMSR-E dataset</i> .....	38
3.1.4 <i>GLDAS-NOAH</i> .....	38
3.1.5 <i>ERA-Interim</i> .....	38
3.1.6 <i>OZNET in-situ soil moisture</i> .....	39
3.1.7 <i>Ancillary data on land cover and roughness</i> .....	41
3.2 METHODS.....	42
3.2.1 <i>Pre-processing</i> .....	43
3.2.2 <i>Evaluation of ASAR GM SSM</i> .....	45
<b>4. RESULTS</b> .....	<b>49</b>
4.1 PREPROCESSING .....	49
4.1.1 <i>Understanding frequency distribution of SSM datasets</i> .....	49
4.1.2 <i>Data transformation</i> .....	52
4.2 STANDARD EVALUATION MEASURES .....	55
4.2.1 <i>Absolute evaluation measures</i> .....	55
4.2.2 <i>Relative evaluation measures</i> .....	65
4.3 ADVANCED EVALUATION METHODS.....	68
4.3.1 <i>Error propagation</i> .....	68
4.3.2 <i>Predicted RMSE</i> .....	69
4.3.3 <i>Triple collocation (TC)</i> .....	73
<b>5. DISCUSSION</b> .....	<b>85</b>
5.1 CAN WE APPLY THE EVALUATION REQUIREMENTS OF SMOS AND SMAP TO ASAR GM SSM? .....	85
5.2 HOW DOES THE SELECTION OF SPATIAL RESOLUTION INFLUENCE ERROR ESTIMATES? .....	87
5.3 IS THERE A BEST COMBINATION OF MEASURES TO DESCRIBE THE QUALITY OF ASAR GM SSM DATASET?.....	88

5.3.1 Absolute evaluation measures .....	89
5.3.2 Relative evaluation measures .....	94
5.3.3 Resume .....	95
5.4 WHAT IS THE QUALITY AND WHAT ARE THE LIMITATIONS OF ASAR GM SSM DATA?.....	96
5.5 LEARNING FROM ASAR GM SSM ERRORS FOR SENTINEL-1.....	98
<b>6. CONCLUSION.....</b>	<b>100</b>
<b>7. REFERENCE.....</b>	<b>103</b>

## List of acronyms

AMSR-E	The Advanced Microwave Scanning Radiometer for EOS (AMSR-E)
ASAR GM	The Advanced Synthetic Aperture Radar (ASAR) Global Mode (GM)
AWRA-L	The Australian Water Resource Assessment modelling system (AWRA-L)
EP	Error propagation
ERA-Interim	The ERA-Interim reanalysis
GLDAS-NOAH	The Global Land Data Assimilation System (GLDAS) NOAH land surface model
OzNET	The Australian monitoring network for soil moisture and micrometeorology (OzNET)
SSM	Surface soil moisture
TC	Triple collocation
RMSE	Root mean square error
$R$	Pearson correlation coefficient
$R_s$	Spearman correlation coefficient
ERS	European Remote Sensing Satellite
SMOS	Soil Moisture and Ocean Salinity
ASCAT	Advanced Scatterometer (Metop)
SMAP	Soil Moisture Active Passive
MAE	Mean absolute error
SSM/I	Special sensor microwave/imager
ESA	European Space Agency
GM	Global Mode
SAR	Synthetic Aperture Radar



# 1. Introduction

## 1.1 Evaluation of soil moisture datasets

Numerous SSM datasets are available on an operational or semi-operational basis from remote sensing platforms at coarse and, recently, also at medium resolutions. These originate from microwave remote sensing instruments such as scatterometers (e.g. ERS) (Naeimi et al., 2009b; Wagner et al., 1999b) radiometers (e.g. AMSR-E) (Kerr et al., 2010; Njoku et al., 2003), and SARs (e.g. ASAR GM) (Pathe et al., 2009b). Soil moisture estimates from remote sensing demonstrated the potential to improve weather forecast capabilities in numerical weather forecast models (Drusch, 2007; Mahfouf, 2010) and water balance in hydrological systems (Brocca et al., 2010c; Matgen et al., 2011), as well as to support estimation of trends and anomalies related to climate change (Liu et al., 2009) or vegetation stratification over areas with water limited vegetation (Sass et al., 2012). The demonstrated benefits of the products motivated an extension of the existing missions (e.g. ASCAT) (Naeimi et al., 2009a) and proposals for new satellite SSM missions (e.g. SMOS and SMAP) (Kerr et al., 2010) and products (Sentinel-1 SSM) (Hornacek et al., 2012).

*Available  
remotely sensed  
SSM datasets*

Understanding errors and differences between remotely sensed SSM datasets was indispensable for their application in models and for an extraction of blended SSM products (Liu et al., 2011). For instance, only a dataset with a good quality error estimation could drive the respective data weights in data assimilation.

The common evaluation methods of remote sensing products were based on their direct comparison with ground-based measurements, where the ground-based measurements were held to be of higher accuracy. For such purpose soil moisture campaigns have been conducted and networks established (e.g. Tarrawarra, REMEDHUS). These guaranteed a collection of ground soil moisture measurements combined with ancillary datasets such as soil and vegetation roughness, vegetation, or meteorological flux measurements.

Early experiments coupling remote sensing with field observations were already performed in 1980 by the Beltsville Agricultural Research Center (BARC) (Wang et al., 1980). Since then the importance of soil moisture has grown rapidly and motivated the establishment of additional soil moisture networks and campaigns. Some examples of such networks include the OzNET (<http://www.oznet.org.au/>, (e.g. Young, R., Walker, J., Yeoh, N., Smith, A., Ellett, K., Merlin & Western, 2008), the SMOSMANIA (Calvet et al., 2007) or the REMEDHUS (Martinez-Fernandez & Ceballos, 2003)) networks that were planned to systematically collect data for several years to decades. In contrast, soil moisture campaigns take usually only few months, investigate the spatio-temporal variations of soil moisture at footprint as well as sub-footprint scale and provide a good testing ground for hydrological models (i.e. Tarrawarra (Western & Grayson, 1998)). Often, short-term campaigns complement satellite missions in that they support algorithm development (NAFE'05 (Panciera et al., 2008)) or evaluate data accuracy (Iowa SMEXo2 (Jacobs, 2004)).

*In-situ networks  
and campaigns*

Recently, a global centralized data hosting facility for data from the in-situ networks was established (Dorigo et al., 2011). This facility allows for a panoramic overview of the globally

existing in-situ stations, but more importantly, supports standardization of the data and simplifies their use for evaluation of the modeled and remotely sensed datasets.

#### *Evaluation with in-situ data*

The role of in-situ observational networks is indispensable as demonstrated by numerous studies (Ceballos et al., 2005; Gruhier et al., 2010; Jackson et al., 2010; Rüdiger et al., 2009; Wagner et al., 2007). The latter studies evaluated one or several remote sensing SSM products (ERS, AMSR-E, METEOSAT and TMI) against the in-situ soil moisture networks in southern Europe, western Africa, Australia, and in the USA.

The unique advantages of the evaluations over the in-situ stations are that a) these are considered to represent the best estimate of the true soil moisture values at the point, b) ancillary information (i.e. soil texture, porosity or local vegetation conditions) are available that may improve understanding on the retrieved statistical results and c) the spatial characteristics of SSM within one satellite footprint may be assessed if many locations are sampled (Ceballos et al., 2005; Jackson et al., 2010). It should, however, be noted that the existence of several measurements within one satellite footprint is common for the soil moisture campaigns but it is exceptional for the soil moisture networks (Dorigo et al., 2011; Miralles et al., 2010).

#### *Sources of spatial errors for in-situ data*

A disadvantage of the evaluation with the in-situ data is that the interpretation of results is hampered by errors of representation. These can originate in the differences in the sensing depths, acquisition times, and scaling. The scaling differences originate in the spatio-temporal distribution of precipitation events (i.e. a convective precipitation events may not be captured by an in-situ station but may still affect a large portion of the corresponding satellite footprint) or in the effects of topography and landcover (i.e. the wetting dynamics measured by an in-situ station in the forest that is situated in the middle of large fields will differ from the wetting dynamics measured by a satellite sensor over a footprint that represents the entire area). Furthermore, the scaling errors showed to be larger than the retrieval error of a single dataset (Martinez-Fernandez & Ceballos, 2005).

The concept of temporal stability is often implemented to understand and mitigate spatial differences (Vachaud et al., 1985). The concept states that soil moisture acts steadily in time (consistently higher, lower or equal) when compared to the spatial mean representing a larger area. If such spatio-temporal behavior is known, one single point can accurately represent the areal mean (Brocca et al., 2010a; Cosh et al., 2004; Famiglietti et al., 1998; Jackson et al., 1999; Jacobs, 2004). This is only possible during days when the study area is governed by its typical characteristic precipitation conditions.

Furthermore, the triple collocation method was recently applied to estimate and mitigate the sampling errors associated with the spatial upscaling of the point measurements (Miralles et al., 2010).

#### *Evaluation with spatial data*

Despite the demonstrated improvements of evaluation studies with in-situ data, these remain restricted to the extent of the networks, where the number of long-term in-situ monitoring networks is still small and mostly restricted to mid-latitude regions (Dorigo et al., 2011). However, to allow for a good evaluation of satellite soil moisture products their global evaluation is needed (Dorigo et al., 2010; Gruhier et al., 2010); mainly because the performance is expected to differ under different land cover, soil type, and climatic conditions.

The first spatial comparisons of in-situ and spatial acquired SSM data were performed in mid 2000's. These comparisons evaluated soil moisture products with other spatial datasets; for instance with remotely sensed precipitation datasets (Bartalis et al., 2008; McCabe et al., 2005) or with aircraft-based soil moisture datasets (Drusch et al., 2004; Mladenova et al., 2010; Panciera et al., 2008). These approaches extended over regions where global soil moisture data could be evaluated. In addition, EP techniques were employed, which allowed evaluation over the entire spatial domain of the data products. EP estimates the error of the soil moisture retrieval based on propagated standard errors of each individual observation (Naeimi et al., 2009b; Parinussa et al., 2011; Pathe et al., 2009a).

Furthermore, numerous evaluation studies were performed using soil moisture outputs of spatial soil water balance and land surface models (Laguardia & Niemeier, 2008; Parajka et al., 2006; Rüdiger et al., 2009; Wooldridge et al., 2003). A thin soil moisture layer has been included into some hydrological models that allowed easy assimilation and evaluation of the shallow (0-5 cm), remotely sensed, soil moisture observations (Brocca et al., 2011; Parajka et al., 2009). The errors of the soil moisture output of modelling system are independent from those of an empirically retrieved remotely-sensed observation and provide additional input to evaluation methods such as triple collocation.

*Evaluation with water balance models*

Commonly, the evaluation approach consisted of a straightforward computation of the correlation coefficient  $R$ , RMSE, and the bias between the remotely sensed and the reference dataset (in-situ or modeled data). A low RMSE value with in-situ data was until now the important soil moisture quality requirement. For instance, the SMOS and SMAP missions' SSM product requirement relies on  $RMSE < 0.04 \text{ m}^3/\text{m}^3$  (Kerr et al., 2010). Several other accuracy requirements for soil moisture provided by WMO are based upon an absolute measure between SSM datasets. Absolute assessments are however largely complicated by differences in represented depths, spatial scaling, and in the exact factors the datasets define. Moreover, they refer to differences in datasets rather than errors, as both of the datasets contribute to the final error estimate with their individual random as well as systematic errors.

*Commonly used RMSE*

The goal to acquire an absolute error estimate largely delayed the development of the soil moisture products from the SAR systems. In fact, the roughness parameterization which has a strong impact on the strength of backscatter is the main source of errors in SAR soil moisture retrievals and its complexity is expected to even increase with the increasing spatial resolution. With the advent of new SAR sensors operating at high spatial resolutions (e.g. TerraSAR-X, RADARSAT-2, and, for 2013 planned, Sentinel-1) the soil moisture product parameterization is expected to remain a major constrain for the SAR soil moisture product development.

*Evaluation of SAR SSM products*

In the last decade a large amount of remotely sensed soil moisture datasets have become (Kerr et al., 2010; Naeimi et al., 2009a; Njoku et al., 2003; Pathe et al., 2009b; Wagner et al., 1999b) that utilize independent algorithms and have an independent error structure. The latest, SMOS, is the first spaceborne mission that was designed specifically for the purpose of soil moisture monitoring over land (Kerr et al., 2010). These products have provided new research opportunities and evaluation activities in the soil moisture domain for coarse resolution sensors.

*Evaluation with other remotely sensed datasets*

First, multi-correlation evaluation techniques were implemented. These assumed that a large number of corresponding datasets signifies that these represent an identical phenomenon (Jeu et al., 2008; Rüdiger et al., 2009; Wagner et al., 2007). Second, a triple collocation method was

*Advanced evaluation measures*

implemented to estimate the error variances by simultaneously solving for systematic differences in the climatologies of three data sources with independent error structures (Dorigo et al., 2010; Scipal et al., 2008b). Third, EP studies were implemented to estimate the random error. A large advantage of the EP and TC method is the fact that these do not require a reference dataset. Fourth, evaluation techniques were developed that judge the quality of soil moisture observations based on the improvements they can bring in a real application. An example of such study is the novel method based on the assimilation of soil moisture retrievals into a simple surface water balance model (Crow, 2007). In this study, the authors analyzed the rainfall errors against filter increments to establish a proxy for the accuracy of the soil moisture retrieval. Another example is a study demonstrating the use of agricultural productivity as an alternative for dataset error evaluation (Champagne et al., 2012). Lastly, a specific case of EP was introduced that evaluates the standard error of the remotely sensed SSM using a soil moisture estimate from a hydrological model (Doubková et al., 2012). The necessary prerequisite of the method is a good understanding on the standard errors of both datasets and their independency.

In conclusion, some historical trends should be highlighted:

- The role of the in-situ observational networks is and will remain indispensable for accurate evaluation of SSM data products
- An increasing number of the coarse resolution soil moisture datasets, retrieved using independent algorithms, and their increasing accuracy allowed for the development of more complex and robust evaluation techniques.
- It is inevitable to complement in-situ evaluation studies with advanced evaluation methods that use a number of soil moisture datasets.
- For several decades, the limited understanding of the effects of the soil roughness and vegetation on the SAR backscatter hampered the development and evaluation of a regional, operationally available, SAR soil moisture product.

## 1.2 Objective and structure

To support the operational use of Synthetic Aperture Radar (SAR) earth observation systems, the European Space Agency (ESA) is developing Sentinel-1, a constellation of two polar-orbiting C-band radar satellites. Much like its SAR predecessors (Earth Resource Satellite, ENVISAT and RADARSAT) the Sentinel-1 will operate at a medium spatial resolution, but with a greatly improved revisit period. Given the planned high temporal sampling and the operational configuration Sentinel-1 is expected to be beneficial for operational monitoring of dynamic processes in hydrology and phenology. The benefit of a C-band SAR monitoring service in hydrology has already been demonstrated within the scope of the Soil Moisture for Hydrometeorologic Applications (SHARE) project (<http://www.ipf.tuwien.ac.at/radar/share/>) (Doubkova et al., 2009). SHARE is one of the ESA's Data User Element (DUE) Tiger Innovator. As part of the project a soil moisture dataset at medium resolution was retrieved from the GM of the ASAR onboard ENVISAT (Pathe et al., 2009b).

*Thesis objective* This thesis was motivated by the need to evaluate the quality of the ASAR GM SSM medium resolution dataset and to provide guidance on appropriate evaluation methodology applicable to any SSM product. The evaluation results are summarized in [chapter 4](#). The guidance on a general evaluation approach is provided in [chapter 5](#).

Prior to the result and discussion section, the state of the art of the SSM evaluation studies is provided in [chapter 1](#), complemented by this section that motivates the objective of this thesis. The essential theoretical background of the work is summarized in [section 1.1](#) introducing currently used evaluation measures in the soil moisture field and providing examples on how the evaluation measures can be selected if an exact application of the data is known and requires a particular data quality. The theoretical part of the work is summarized in [chapter 2](#). Methods and datasets are to be found in [chapter 3](#).

*Intro, theory,  
and  
methodology  
section*

Two main issues hamper the interpretation of the evaluation studies; namely differences in the representation of the spatial extent and sensing depth. Methods exist that mitigate the latter differences; these were summarized in [section 4.1](#). For a long time, error assessment studies were based on an absolute quality requirement of SSM. These were supplemented by relative error assessments and form together a group of evaluation measures that is here referred to as standard evaluation measures. Evaluation of the ASAR GM SSM dataset using standard evaluation measures can be found in [section 4.2](#).

*Results section*

Recently, advanced evaluation methods have emerged in the literature. These rely on statistical analyses of several datasets, or directly on analyzes of filter increments after data assimilation. By doing so these methods take away the long-lasting but unrealistic expectation of the soil moisture community on the existence of one true soil moisture dataset. Given the successful use of advanced evaluation techniques for coarse resolution datasets, their fruitful use to describe the quality of medium resolution SSM products was expected and assessed for the ASAR GM SSM in [section 4.3](#). The new evaluation methods are expected to be beneficial for SSM products from the SAR data. Their development was delayed due to the long lasting relying on standard evaluation measures and requirement of a reference dataset. Such evaluations were often not successful due to the limited understanding of the effects of soil roughness (Verhoest et al., 2008) and vegetation on the SAR backscatter (Wagner et al., 2009).

In the discussion section, the results are analyzed from a general perspective and form answers to questions related to the characteristic problems of data evaluation.

*Discussion  
section*

Given the issue of retrieving absolute soil moisture values from SAR sensors, the requirements of SMOS and SMAP communities on an absolute accuracy appear challenging to apply to the SAR SSM products. This hesitation motivated the first question in the discussion section: “Can we apply the evaluation requirements of SMOS and SMAP to ASAR GM SSM?” ([section 5.1](#)).

Any evaluation study is hampered by the differences in the sensing depth, acquisition times, and spatial scaling. The former differences are commonly experienced (e.g. evaluation between the remotely sensed and in-situ SSM) and were shown to be larger than the retrieval error of a single dataset (Martinez-Fernandez & Ceballos, 2005). Importantly these are expected to also have a large impact on the evaluation studies where medium resolution datasets are evaluated along with coarser resolution products. For this reason [section 5.2](#) reflects on following: “How does spatial resolution influence the error estimates?”

In the results section, evaluation methodologies are summarized based on numerous assumptions. For instance, a simple comparison of two datasets assumes one of the datasets to be close to the “truth”. This is clearly violated because all observation systems contain errors and may introduce a substantial pseudo bias effect (Stoffelen, 1998). Another example is the triple

collocation method. While this method does not rely on one reference dataset, it assumes the errors of the implemented observations to be fully independent. This can be violated given, for instance, the similar physical principles behind all microwave soil moisture datasets. The third discussion question ([section 5.3](#)) addresses the fact that different evaluation measures have different assumptions and may evaluate different qualities of the data and is . In particular, it asks: “Is there a best combination of measures to describe the quality of soil moisture data?” described in.

Finally, the findings on the error characterization of the ASAR GM SSM product retrieved in [chapter 4](#) are collected and summarized to answer the question in [section 5.4](#) “What is the quality and what are the limitations of ASAR GM SSM data?” These findings are transferred to the potential Sentinel-1 SSM product. Changes to the final product due to the sensor characteristics are taken into consideration. Finally, a discussion of the following topic: “Learning from ASAR GM SSM errors for Sentinel” is provided in [section 5.5](#).

*Inovativness of the thesis*

This work provides a critical assessment of standard and advanced evaluation methods for the remotely sensed soil moisture products and applies these to assess the performance of the medium resolution ASAR GM SSM product. While individual evaluation methods have been introduced in journal papers and project reports before, this work is innovative as it provides a concise summary of evaluation measures combined with a demonstration of their shared use. Furthermore, the innovation lies in the transformation of the triple collocation evaluation method to the ASAR GM medium resolution SSM product, which was applied until now only to evaluate coarse resolution (~25 km) datasets (i.e. Dorigo et al., 2010; Scipal et al., 2008).

*Future Sentinel-1 SSM product*

A well-specified error characterization of the medium resolution ASAR GM SSM is provided. The demonstrated evaluation strategies are easily transferable to the future Sentinel-1 SSM product. While data assimilation of the ASAR GM soil moisture estimates may be currently restricted by its poor radiometric resolution, the improved radiometric accuracy of the proposed SSM product from Sentinel-1 combined with at least so good temporal coverage of the ASAR GM SSM suggests a great benefit of the product for flux exchange, crop growth, and water balance modelling.

## 2. Theory

### 2.1 Pre-processing steps prior to data evaluation

This chapter presents the terminology used in this work and highlights steps that are necessary prior to soil moisture data evaluation. These steps consist of understanding and the mitigating differences in the datasets caused by different spatial and temporal resolution, different units and sensing depth.

The chapter is written in general terms and can be applied to soil moisture datasets originating from any source (model, remotely sensed sensor or ground station).

#### 2.1.1 Statistical terminology

The statistical terms presented in this chapter are used in the literature interchangeably. The author's goal is to define their meaning for the scope of this work.

The sources of errors are generally classified into two categories, systematic and random. Systematic errors can be introduced by faulty equipment, faulty calibration of a model or an instrument, or from data unrepresentativeness in space or time. Random errors can originate from statistical fluctuation in the collection of the finite numbers and arise each time the experiment is repeated. *Sources of errors*

For the purpose of this study the systematic errors are further divided into a) time-variant systematic errors (e.g. due to missing parameter), and b) time-invariant systematic errors (indicated as systematic bias throughout this study) that occur due to different mean and range of several soil moisture datasets.

Furthermore, the terms *accuracy*, *precision*, and *error* need to be distinguished as these are interchangeably used in evaluation studies. While *accuracy* is defined as the level of closeness between the measured phenomena and the true value, *precision* refers to the reproducibility of the measurement or to the degree of scatter. According to this, *precision* relates to random processes (Iso, 1994) while *accuracy* involves a combination of the random components as well as systematic errors (Figure 1). *Accuracy* can be assessed with the mean absolute bias and *precision* with the standard deviation computed from the mean value. *Accuracy, precision, and error*

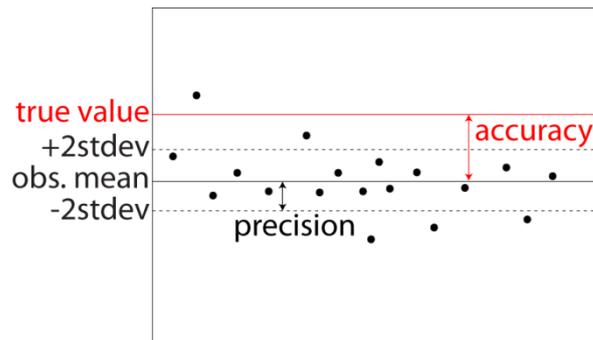


Figure 1. A graphical representation of precision and accuracy.

*Error* is the actual difference between the measured phenomena and the true value. It is what causes values to differ when a measurement is repeated or when a measurement is biased.

### Error and difference

An interesting discussion relates to the understanding of the terms *error* and *difference* (sometimes also *discrepancy*). The term *error* is computed between two datasets where one represents the true measurement (often represented by a ground station). The term *difference* (*discrepancy*) is used for a difference computed between any datasets representing the same quantity, where their 'truth' is not guaranteed. Similar discussion relates to the difference between the terms Root Mean Square Error (RMSE) and the Root Mean Square Difference (RMSD). The RMSE is used if the representation of the "truth" is known or it can be assumed (commonly, observations from ground station). On the contrary, the RMSD is used in cases when no assumption about the true dataset can be made.

It is, nevertheless, a difficult task to decide which measurement represents the "truth", what is the needed level of accuracy and what are the defining criteria to decide about the "acceptance" of the model. For that reason and for the sake of simplification the terms *error* and RMSE (instead of the commonly and more correct "difference" and RMSD) will be used throughout this work.

### Evaluation versus validation

Also the terms *model evaluation* and *model validation* are used in the literature interchangeably (Prisley & Mortimer, 2004) for studies that compare two or several soil moisture datasets. In general terms, *model validation* is achieved if the model accurately predicts the observed phenomena within a certain *precision*. The *model evaluation* refers to any assessment of quality of the remotely sensed data. In this work we refer to the means of *evaluation* as an assessment of errors without consideration whether these fall within certain margin of error.

The relevance of the usage of the term *validation* has been widely discussed. The satellite soil moisture missions have frequently defined validation activities to verify that retrievals meet the required margin of RMSE. It is, nevertheless, a difficult task to decide what represents the "truth", what is the needed level of accuracy and what are the defining criteria to decide about the "acceptance" of the model. In fact, the term *validation* is controversial and used as a term to denote model assessment (Bellocchi et al., 2011). Finally, it is argued that the term *validation* puts pressure on modelers going beyond the degree to which they feel comfortable when discussing the strength of the model (Oreskes, 1998).

While there are no common procedures widely accepted to perform evaluation tasks (Cheng et al., 1991), some suggested norms exist (Jakeman et al., 2006; Van Dijk & Warren, 2010). These suggest that an evaluation should judge the acceptability of the performance of a model or an algorithm for a particular purpose (e.g. has the soil moisture dataset a high accuracy to improve performance of a river-runoff model?). Typically, it is claimed that a model evaluation includes any action in which the quality of a model or an algorithm is established (Jakeman et al., 2006).

### 2.1.2 Preprocessing steps for soil moisture evaluation

Soil moisture products can be derived from remotely sensed data, can be measured on the ground or modeled with hydrological or land surface models. The remotely sensed soil moisture implies observations that are directly related to soil moisture in upper few centimeters. The models (i.e. hydrological, landscape-hydrological, atmospheric), on the other hand, rely on simplified and theoretical relationships. Often the satisfaction of the assumptions on the different soil moisture flows is sometimes more important than the output variable itself. Furthermore, remote sensing data have a spatial character while models are often forced by point-based precipitation inputs. In addition, models often represent soil layers of the upper 10 cm (e.g. AWRA-L represents 5-10 cm) or deeper whereas C-band microwave products retrieve signal from less than 5cm centimeters.

The resulting retrievals and models often represent different soil moisture depths and spatial extents and can be expressed in different units. Furthermore, even if representing the same depth and scale the retrievals may differ due to their varying soil moisture sensitivity. The latter differences prevent measuring an absolute agreement between the time-series (Brocca et al., 2011) and inclusion and assimilation of the datasets into models (Brocca et al., 2010c).

Transformation measures are therefore needed to be applied prior to data evaluation. Four commonly used transformation measures are here summarized. These include:

- the transformation into common soil moisture units
- the linear regression
- the rescaling using mean and standard deviation
- the CDF matching approach
- modelling of vertical soil moisture distribution (e.g. the Soil Water Index (SWI), or the Richards equation)

Table 1 suggests which of the latter methods should be used to remove systematic differences caused by spatial scale, the differences in depth, and the differences in units.

*Table 1. The reasons for the systematic differences between soil moisture datasets and suggested methods for their removal.*

<b>Cause of the systematic difference</b>	<b>Recommended method for its removal</b>
Differences in spatial representation	CDF, linear regression, or rescaling using mean and standard deviation
Differences in depth	modelling of vertical soil moisture distribution, CDF, linear regression, or rescaling using mean and standard deviation
Differences in units	Unit transformation

*Consequences of data transformation* Several effects due to the transformations should be considered. Firstly, all pre-requisites of the transformation techniques need to be fulfilled (e.g. normality of data, sufficient data samples). These are discussed in detail later in this section. Secondly, the higher-order transformation may overfit data to the reference dataset and introduce further bias. Performing transformation on varying window sizes (temporal as well as spatial) may serve as sensitivity analyses on the stability of the calibration constants. Lastly, the dataset transformation alters the absolute values of soil moisture observations and therefore the evaluation results. Such effect can be neglected given that all applied studies of soil moisture require relative rather than absolute estimates of soil moisture parameters (section 2.3).

#### *2.1.2.1 Unit transformation*

*Existing units* The soil moisture products are expressed in a variety of units. The most commonly used soil moisture units used in a large number of soil moisture networks (i.e. OzNet, REMEDHUS or the AMMA) and satellite soil moisture products (e.g. from AMSR-E or SSM/I) are the volumetric units. These were referred to throughout this thesis as vol % or  $\text{m}^3/\text{m}^3$  and express the ratio between the volume of water and the volume of soil holding the water in a given soil depth [ $\text{m}^3$  water per  $\text{m}^3$  of soil]. In other words it represents the fraction to which the pores are filled with water. The pores usually occupy soil fraction lower than 0.6. As a result, the volumetric fraction ranges between  $0.0 \text{ m}^3/\text{m}^3$  (completely dry) and  $0.6 \text{ m}^3/\text{m}^3$  (full saturation).

Soil moisture datasets are also often expressed in relative units. The relative units are commonly used for microwave satellite soil moisture products and measure the change of the retrieved signal relative to its maximum dynamic range (0–1 or 0–100 %). An example of such dataset are the ERS the ASCAT, or the ASAR GM soil moisture products. The backscatter measurements are converted to soil moisture estimates by applying the TU Wien soil moisture retrieval algorithm (Wagner et al., 1999c). To transform the volumetric to the relative soil moisture it simply needs to be divided by porosity  $P$ .

A detailed information of these and other existing soil moisture units (e.g. gravimetric, plant available water) and their conversion measures were presented elsewhere (Dorigo et al., 2011).

Unit conversion measures were not the focus of this thesis due to the following reasons which include: given the linear character between the volumetric and the relative soil moisture units (the only used in this thesis) it was expected that other transformation methods (e.g. linear regression or CDF, see following sections for detailed discussion) can well replace the unit transformation methods. In addition, other transformation methods account for the shortcomings of the ancillary data (i.e. texture, porosity, and organic matter content), and for the differences in spatial representation and depth of the different measurements. These could not be accounted for by the unit transformation technique.

#### *2.1.2.2 Linear normalization*

Two approaches are commonly used for linear regression of soil moisture datasets. Both require an assumption of linear relationship between the datasets. One is based on the application of a simple regression equation between two evaluated datasets, minimizes the RMSE between the compared datasets, and removes the differences in the mean (Jackson et al., 2010). The latter approach can also be performed iteratively. In iterative regression the individual RMSEs are

*Linear regression*

initially assumed to be equal. Consequently, the calibration constants and errors are iteratively altered until their convergence is achieved. This method is of a great benefit if more than two datasets are used in regression.

The second approach removes the differences in the standard deviation and the mean between two datasets (Brocca et al., 2010b; Draper et al., 2009) and as such requires these datasets to have a normal distribution. In particular, the matched dataset  $y$  is computed using two SSM datasets  $x$  and  $y$  as follows:

*Linear rescaling using mean and standard deviation*

$$x_i = Bx_{or,i} + A, \quad 2-1$$

where  $i = 1, \dots, N$ ,  $N$  is the total number of soil moisture acquisitions and the local coefficients  $A$  and  $B$  are defined as:

$$A = \bar{y} - \frac{s(y_i)}{s(x_{or,i})} \bar{x} \quad 2-2$$

and

$$B = \frac{s(y_i)}{s(x_{or,i})}, \quad 2-3$$

where  $\bar{x}_{or}$  is the mean of all  $x_{or,i}$ ,  $\bar{y}$  mean of all  $y_i$  and  $x$  represents the rescaled  $x_{or}$ .

Here, the parameter  $B$  mirrors the difference in the variability of individual SSM datasets; the parameter  $A$  reflects combination of differences of both the variability and the mean. Implicitly, these parameters also refer to different soil types, land cover, and climate (Scipal et al., 2008a). If the expectation on normal distribution is fulfilled the parameters linear fitting parameters and the parameters  $A$  and  $B$  in linear rescaling using mean and standard deviation should be equal.

Importantly, while the linear regression approach allows evaluated time-series to have different mean, the linear rescaling approach allows dataset to have different mean and variance.

### 2.1.2.3 Cumulative Distribution Function (CDF)

To remove differences in higher order moments a non-linear CDF is recommended (section 5.2.3). Application of the CDF is commonly performed in data assimilation studies. As mentioned at the beginning of this chapter caution should be given not to over fit the evaluated datasets as this may introduce unnecessary bias. For the latter reason only datasets with similar data distribution should be transformed. Significantly different data distributions and dynamics may signify that the datasets represent different phenomena.

*CDF*

The CDF is performed by matching cumulative distribution functions of two datasets by using linear or polynomial fitting. Depending on the order of the fitted polynomial, equivalent number of moments is mitigated. For instance, a 3<sup>rd</sup> order polynomial could correct differences in the first four moments (the mean, the variance, the skewness and the kurtosis) (Drusch et al., 2005).

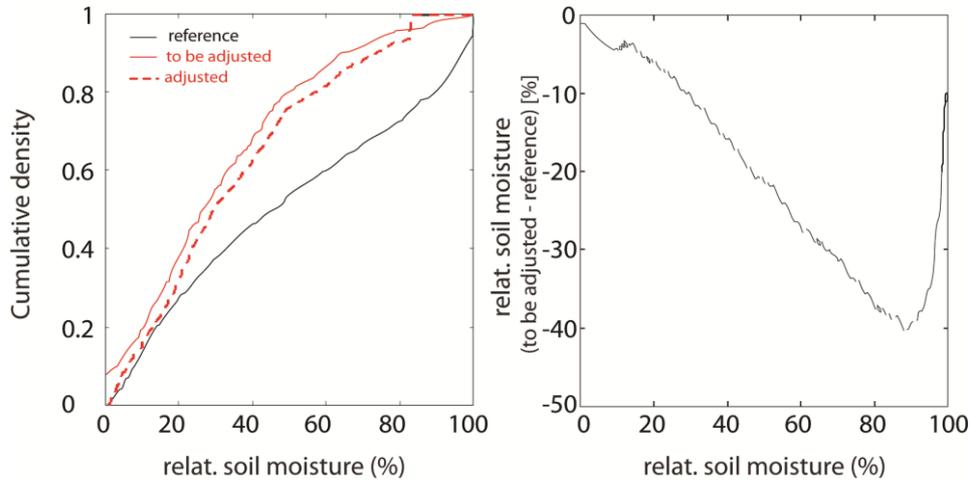


Figure 2. The CDF for a point location in southeastern Australia for two soil moisture datasets (left) and the corresponding differences in soil moisture for each rank (computed as a difference reference – to be adjusted) (right). The operators were computed using a 6<sup>th</sup> order polynomial fit.

The actual computation of the CDF function is performed in three separate steps. Firstly, the datasets are ranked. Secondly, the differences in soil moisture between the corresponding ranks of the two datasets are computed (Figure 2, left). Lastly, the observation operators are computed as a polynomial fit between the computed differences and the ranked observed soil moisture (Drusch et al., 2005) (Figure 2, right). These remove the systematic differences between both datasets. The observation operators are defined by the type of the observations; in particular, by their specific statistical properties and distributions (Drusch et al., 2005).

#### 2.1.2.4 Exponential filter

#### Exponential filter

The exponential filter can be used to remove the difference in depth of soil moisture measurements. It simulates the profile soil moisture over a deeper soil layer based on the acquisitions of the shallow soil moisture as in Wagner et al. (Wagner et al., 1999b). In this study the version of the SWI according to (Albergel et al., 2010) is introduced:

$$SWI_n = SWI_{n-1} + K_n [SSM(t_n) - SWI_{n-1}], \quad 2-4$$

with the gain  $K_n$  at time  $t_n$  given by:

$$K_n = \frac{K_{n-1}}{K_{n-1} + e^{-\left(\frac{t_n - t_{n-1}}{T}\right)}}, \quad 2-5$$

where  $T$  is a characteristic time length that characterizes the temporal variation of soil moisture within the root-zone profile and the gain  $K_n$  ranges between 0 and 1. For the initialization of this filter,  $K_0=1$  and  $SWI_0 = SSM(t_0)$ .

## 2.2 Evaluation measures and methods

The interest in evaluation of climatological and environmental datasets has grown rapidly (Willmott & Matsuura, 2005). Likewise, the need to address the error of remotely sensed soil moisture datasets is evident (e.g. Matgen et al., 2011; Scipal, Holmes, De Jeu, Naeimi, & Wagner, 2008). Interest is mounting which statistical measures are the most suitable and how their selection differs based on an application. The commonly used evaluation measures in soil moisture campaigns were the correlation coefficient ( $R$ ) (Brocca et al., 2010b) and the root mean square error (RMSE) (Jackson et al., 2010). Regrettably, it is RMSE that is also the most misinterpreted error measure (Willmott & Matsuura, 2005).

This chapter summarizes standard (section 2.2.1) as well as advanced evaluation measures (section 2.2.2) that may complement the commonly used RMSE and  $R$ . Furthermore, several recommendations about when to use these evaluation measures or how to combine several of them are provided. Detailed recommendations on application-relevant evaluation measures can be found in section 2.3. The last chapter (2.2.3) introduces the future direction in soil moisture evaluation studies: the evaluation of evaluation measures.

### 2.2.1 Standard evaluation measures

This section introduces and interprets the commonly used evaluation measures. These include a measure of the absolute agreement (RMSE, MAE, and bias) and the relative agreement ( $R$ ,  $R^2$  and  $R_s$ ) between two or more soil moisture datasets. While the absolute measures assess the effect of random and/or systematic errors, the relative measures inspect the evolution in time of the separate datasets. The relative measures are often yielded by dividing the absolute measure by the dataset itself or by its variance or standard deviation. Such measures are spatially comparable and independent on the absolute magnitude.

All standard evaluation measures are based on a comparison of two observation systems, where both systems contain errors. As such, final evaluation results only represent differences and should not be considered as errors related to the true observations. Such assumption may introduce a substantial pseudobias effect (Stoffelen, 1998).

#### 2.2.1.1 Measures of absolute agreement

Measures of absolute agreement refer to the positive magnitude or mean of two variables dissimilarity. The measures outlined here provide a summary of dissimilarities of comparative soil moisture datasets.

The measures of absolute agreement should be interpreted carefully as these are influenced by the mean and variance of the datasets. For instance, the increasing MAE and RMSE can be explained by the increasing error in the datasets as well as by the increasing mean or variance. For the latter reason, the normalized versions of the absolute measures are often computed.

Furthermore, the absolute evaluation measures are performed on already transformed datasets (see chapter 4). Applying evaluation measures on the original datasets would assess random errors as well as errors due to differences in spatial scale, units or soil moisture dynamics. The final evaluation measure is expressed in the units of the datasets to which the original data were transformed. A brief summary of the current measures of the absolute agreement is provided below.

### RMSE

RMSE is currently the most commonly used measure of precision and, if evaluated with a ground observations, also of accuracy. It has been widely used in evaluation studies of soil moisture datasets (i.e. (Brocca et al., 2010b, 2011; Doubkova et al., 2012; Jackson et al., 2010; Mladenova et al., 2010)) and plays an important role in the assessment of performance criteria for the SMOS and SMAP missions (Miralles et al., 2010). RMSE signifies the closeness of two datasets representing the same phenomena and is defined for two samples of variables  $x_i$  and  $y_i$  as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{N}}, \quad 2-6$$

where  $i=1\dots, N$  and  $N$  is the number of measurements. By its computation RMSE alters the magnitude of each difference by its squaring and rooting. The squaring is performed to remove the potential negative value. However, this has a potentially negative consequence of quadratically penalizing the residuals between parameters.

It should be noted that RMSE reflects not only the average error but also the variance in the error and the number of data points (Willmott & Matsuura, 2005).

### MAE

Further measures of consistency between two datasets are the Mean Absolute Error (MAE) and Bias. Both are absolute measures of error. Importantly, bias is computed on non-transformed datasets. The Mean Absolute Error (MAE) of a sample of  $n$  measurements is defined as:

$$MAE = \frac{\sum_{i=1}^n |x_i - y_i|}{N} \quad 2-7$$

, where  $x_i$  and  $y_i$  are two continuous variables with  $y_i$  being the representation of the true value. The measure returns the average absolute magnitude of each difference and represents so the typical error magnitude.

### MAE versus RMSE

Unlike RMSE, MAE doesn't quadratically penalize errors, nor does it reflect their variance. For these reasons, MAE has been recommended by several studies as a more suitable measure of average error than RMSE (Mielke & Berry, 2007; Willmott & Matsuura, 2005).

In particular, large errors have a relatively greater influence on the total square error than do smaller errors. In other words, RMSE increases if the total error is concentrated within a small number of increasingly large individual errors. As such, RMSE reflects total error magnitude (MAE) as well as variability of error magnitudes and it is impossible to distinguish between them.

If the positive or negative nature of the error is required the Bias can be computed. It is *Bias* calculated using the expression:

$$Bias = \frac{\sum_{i=1}^n (x_i - y_i)}{N} = \bar{x} - \bar{y}, \quad 2-8$$

where  $x_i$  and  $y_i$  are two continuous variables. Bias should be interpreted cautiously since it indicates the average model bias and for two datasets with the same bias can approach 0. This is explained by the cancelation of the independent negative and positive errors. Furthermore, bias measures reflect only on accuracy measures while RMSE and MAE reflect on both accuracy and precision.

On the contrary to all above listed measures, nRMSE assess a scaleless performance. It is here *Normalized root mean square error (nRMSE)* provided for the sake of completeness. nRMSE normalizes RMSE with  $\bar{y}$ , the mean of  $y_i$ 's, as follows:

$$nRMSE_m = \frac{RMSE}{\bar{y}}. \quad 2-9$$

The final measure gives an estimate of the averaged, quadratically penalized, difference between two datasets normalized by their mean. The  $nRMSE_m$  allows for a spatial comparison as it is not affected by the dataset variance. An identical normalization can be performed using the standard deviation.

### 2.2.1.2 Measures of relative agreement

In many cases, information pertaining to the nature of the association between two variables, and not solely the nature of their dissimilarity, is required. To extract such information, we use measures of relative agreement. Relative agreement refers to the potential existence and strength of an association between two variables. Outlined here are correlation measures which serve to convey information about such associations.

The *Pearson correlation coefficient*  $R$  measures linear relationship between two variables. It is *Pearson correlation coefficient* retrieved by a division of the covariance by the estimators of the standard deviations:

$$R = \frac{Cov(x, y)}{s(x)s(y)}, \quad 2-10$$

where  $x^*$  and  $y^*$  represents respectively the standard normal random variable for which the  $\bar{x^*} = 0$  and  $s$  stands for the estimator of the standard deviations. The same applies for  $y$ . This is written in full using (2-10) and the definition of the standard deviation of the variables  $x_i$  and  $y_i$ , as:

$$R = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2}} \quad 2-11$$

where  $\bar{x}$  is the mean of all  $x_i$ 's and  $\bar{y}$  is the mean of all  $y_i$ 's. The achieved values range between 1 and -1. These maxima indicate a perfect positive and negative correlation respectively.  $R$  effectively provides a measure of how well the two datasets are associated in their evolution in time.

*Coefficient of determination*

To obtain an easier to understand meaning of  $R$  it is recommended to compute its square ( $R^2$ ). Known as the *coefficient of determination*,  $R^2$  represents the proportion of the total variation in  $y_i$  that can be attributed to the linear relationship with corresponding values in  $x_i$ . In a perfect correlation (where  $R = \pm 1$ ) a variation in one of the variables is exactly matched by a corresponding variation in the other. The parameter  $1 - R^2$  indicates to what extent other factors (outside  $x_i, y_i$ ) influence  $x_i$  and  $y_i$ .

*Significance level of R*

To assess significance level of  $R$  and  $R^2$  auxiliary tests need to be performed. For that purpose the t-statistic test can be used:

$$t = R \sqrt{\frac{N - 2}{1 - R^2}} \quad 2-12$$

where  $N-2$  signifies the degrees of freedom.

For small number of samples and for cases when data follow bi-normal or two-dimensional Gaussian distribution the  $t$  has Student's t-distribution. The null hypothesis claims that there is no correlation between data. This is tested by comparing the value of  $t$  with  $t$ -table tail probabilities for a given significance level. If  $t$  is larger than the  $t$ -table tail probability, the null hypothesis is rejected and the significance of the correlation is demonstrated.

$R_S$

The next metric is the Spearman nonparametric correlation ( $R_S$ ). Its key advantage is that it does not require any assumption about the nature of the relationship between evaluated datasets and measures only monotonic relationship between datasets. The computation relies on the replacement of the values with its rank among all  $x_i$  in the sample. This is useful in cases where the raw soil moisture data values or estimate values are not normally distributed or contain outliers. The Spearman rank correlation is defined as:

$$R_S = \frac{\sum_{i=1}^N (S_i - \bar{S})(T_i - \bar{T})}{\sqrt{\sum_{i=1}^N (S_i - \bar{S})^2 \sum_{i=1}^N (T_i - \bar{T})^2}} \quad 2-13$$

where  $S_i$  and  $T_i$  are the rank of  $x_i$  and  $y_i$ , where  $i=1, \dots, N$  and  $N$  is the number of measurements. Similarly to  $R$  the significance of the correlation can be computed according to (2-12). Importantly,  $R_s$  is independent on the distribution of the data. As that it is less affected by variance of data and allows for spatial comparisons with different evaluation results.

In case of soil moisture,  $R$  computed on absolute values may be artificially enhanced by the effect of seasonality. This is given by the mathematical formulation of  $R$  that incorporates covariance and standard deviations; these both increase with increasing magnitude of the absolute values.  $R_s$  is computed on ranked datasets and doesn't weight to the actual values of soil moisture. As such it reflects solely the quality of the data to depict the order of the values in an ordered sequence.

Last but not least, correlation does not prove causation; that can only come from knowledge of underlying behavior.

Many environmental datasets exist as a sequence in time or space where a single measurement relates to neighboring measurements in time or space. Such tendency is called autocorrelation. Autocorrelation causes errors to be non-random and can thus violate the basic assumption for computation of correlation coefficients  $R$  and RMSE. It also alters the computed significance level of a correlation and may artificially increase the variance of the data. By selecting the evaluated acquisitions randomly provides a solution to avoid autocorrelation. Often, the Durbin-Watson statistic is used to examine the possible autocorrelation effects by studying independency of the residuals.

*Autocorrelation and Durbin Watson test*

The Durbin-Watson test examines whether the first order autocorrelation parameter  $\rho$  is 0. If this is the case, the residuals are independent and the autocorrelation is dismissed. The statistic is computed as:

$$D = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n (e_i)^2}, \quad 2-14$$

where  $e_i$  are the residuals determined by fitting a model using least square, where  $i=1, \dots, N$  and  $N$  is the number of measurements. Durbin and Watson obtained the approximate upper and lower bounds ( $d_U$  and  $d_L$ ) on the statistic  $D$ . If  $d_L \leq D \leq d_U$ , the test is inconclusive. However,  $D > d_U$  concludes that  $\rho = 0$ ; and  $D < d_L$  signifies that  $\rho > 0$ .

The way how to avoid autocorrelation in data is to conduct experiments on random selection rather than on the consequent measurements.

### 2.2.1.3 Evaluation of multiple datasets

The above mentioned statistical measures can be computed between more than two datasets. When several datasets are evaluated the statistical measures are computed between each dataset pair (Rüdiger et al., 2009; Wagner et al., 2007). The results can be then averaged or interpreted independently.

When correlations are computed between several datasets it is important to ensure that the errors of the datasets are independent and that they capture the same physical phenomenon. Only then would high individual cross-correlations signalize a high probability that these represent an identical phenomenon that is closely linked to soil moisture. The probability increases with the increasing number of datasets with high correlation values.

Also, a simple averaging of  $R$  and RMSE computed among several soil moisture datasets can provide valuable results (Wagner et al., 2007). While the mean  $R$  gives an estimate of a cross-association between several datasets the mean RMSE represents the cross-difference of several datasets. As mentioned in chapter 6.1.1 the absolute evaluation measures are performed on already transformed datasets to allow RMSE to include errors due to different spatial scaling or sensing depth of the evaluated datasets.

### 2.2.2 Advanced evaluation methods

An increasing number of coarse resolution soil moisture datasets and their improving accuracy supported the development of advanced evaluation methods. These are introduced in this chapter along with a discussion on their ability to assess random and systematic errors.

The advanced evaluation methods are performed on transformed datasets (see chapter 4) to avoid effects of errors due to a different spatial or temporal scaling or a different sensing depth.

#### 2.2.2.1 Error propagation (EP)

The EP technique determines the standard deviation of the error of a model by assessing the impact of the standard deviations of the errors in the individual input model parameters (Taylor, 1997). The final measure comprises random errors but doesn't encompass the effect of systematic errors.

The solution of the EP can be described as follows. Firstly, several assumptions about the different sources of errors are made. Secondly, the detected standard deviation of the errors are propagated through the model while it is assumed that all approximations are negligible and that the model itself is correct. Finally, the numerical or analytical solutions of the EP are solved. An example of a numerical method is the Monte Carlo simulation (Metropolis & Ulam, 1949), an example of an analytical method is the method of moments (Wong, 1985).

The ASAR GM error estimated according to Pathe et al. (Pathe et al., 2009a) was used in this study that implements the method of moments. For detailed information on the application of EP methods for the ASAR GM SSM dataset refer to Pathe et al. (Pathe et al., 2009a).

The commonly used formulation for the method of moments is the first-order second-moment error analyses that estimates the spread of  $x$  based on a first-order approximation to  $f$ , where  $x=f(u)$ . The standard deviation of the error of  $x$  is thus approximated as:

$$s_x \triangleq \left| \frac{dx}{du} \right| s_u, \quad 2-15$$

where  $s_x$  and  $s_u$  are the standard deviations of the errors of the dataset  $x$  and  $u$ . In other words, the standard deviation of the error in  $x$  depends on the standard deviation of the error of  $u$  and how sensitive is change in  $x$  to change in  $u$  (expressed as  $dx/du$  in (2-15)).

Considering a function of two variables  $x=f(u,v)$  the problem expands as follows. Taking  $\frac{\partial x}{\partial u}$  and  $\frac{\partial x}{\partial v}$  as the slopes in the  $u$  and  $v$  dimension, the error variance of  $x$  is expressed as:

$$s_x^2 \approx s_u^2 \left(\frac{\partial x}{\partial u}\right)^2 + s_v^2 \left(\frac{\partial x}{\partial v}\right)^2 + 2s_u s_v \left(\frac{\partial x}{\partial u}\right) \left(\frac{\partial x}{\partial v}\right), \quad 2-16$$

where  $s_x$ ,  $s_u$ , and  $s_v$  are the standard deviations of the errors of the datasets  $x$ ,  $u$ , and  $v$ . The final standard deviation of the error can thus be expressed as:

$$s_x \approx \sqrt{\left(\frac{\partial x}{\partial u} s_u\right)^2 + \left(\frac{\partial x}{\partial v} s_v\right)^2 + 2s_u s_v \left(\frac{\partial x}{\partial u}\right) \left(\frac{\partial x}{\partial v}\right)}. \quad 2-17$$

The parameter  $s_x$  will be greater if  $u$  and  $v$  are positively correlated. In an opposite case, when  $u$  and  $v$  are uncorrelated, the entire equation gets simpler as the third term can be omitted. The resulting relationship is denoted as the Gaussian EP method in which the standard deviation of the error in  $x$  depends on the standard deviations of the errors in  $u$  and  $v$  and on their sensitivity to  $x$  as follows:

$$s_x \approx \sqrt{\left(\frac{\partial x}{\partial u} s_u\right)^2 + \left(\frac{\partial x}{\partial v} s_v\right)^2}. \quad 2-18$$

An important assumption of the EP is the fact that the separate errors are small compared to the partial derivatives.

A successful application of the Gaussian EP has been used to assess the standard deviation of the error of satellite (Doubková et al., 2012; Naeimi et al., 2009b; Parinussa et al., 2011; Pathe et al., 2009b) as well as modeled (Mölders et al., 2005) soil moisture datasets. It has been concluded that the Gaussian EP is indispensable for analyses of parameterized soil processes (Mölders et al., 2005). Other study concluded that the Gaussian EP method can achieve comparable results to, the computationally much more expensive, Monte Carlo method (Parinussa et al., 2011).

#### 2.2.2.2 Triple collocation (TC)

The triple collocation (TC) method estimates errors (random and systematic error alternating in time (e.g. due to missing parameter)) of three or more calibrated soil moisture products by multiplying differences between three, to a common dataset calibrated, soil moisture products

with independent error structures. The assumptions on the statistical characteristics of the datasets and errors are crucial for the validity of the method and assume that:

- the residual errors are uncorrelated
- the different datasets observe the same physical phenomenon
- a sufficient number of triplets is available

Three datasets  $x$ ,  $y$  and  $z$  with independent error structures are needed that represent an identical phenomenon. These are expected to relate to the true soil moisture dataset  $t$  in a linear fashion. Their transformation can be performed using a simple linear fitting, linear normalisation by using standard deviation and mean or iterative linear least square approximation (Scipal et al., 2008b). The selection of an appropriate matching technique should be considered careful based on the distribution of the input data. See section 8.1 for detailed discussion. Importantly, in contrast to other evaluation measures the triple collocation doesn't require a dataset representing the "truth".

The TC method is here demonstrated with a linear calibration:

$$\begin{aligned}x_i &= a_x + b_x t_i + \varepsilon_{x,i}, \\y_i &= a_y + b_y t_i + \varepsilon_{y,i}, \\z_i &= a_z + b_z t_i + \varepsilon_{z,i},\end{aligned}\tag{2-19}$$

where  $i = 1, \dots, N$  and  $N$  is the total number of elements,  $\varepsilon_x$ ,  $\varepsilon_y$  and  $\varepsilon_z$  are the residual errors of the datasets  $x$ ,  $y$  and  $z$  and  $a$  and  $b$  are the linear scaling coefficients with subscript corresponding to the actual dataset.

The aim is to estimate the variance of the residual errors of each dataset. To do so we first need to simplify the equation (2-19) by eliminating the calibration constants  $a$  and  $b$ . This can be achieved by expressing  $x_i^* = \frac{x_i - a_x}{b_x}$  and  $\varepsilon_{x,i}^* = \frac{\varepsilon_{x,i}}{b_x}$ . The analogous relationships apply also for the other two equations in (6.2-5) and result in

$$\begin{aligned}x_i^* &= t_i + \varepsilon_{x,i}^*, \\y_i^* &= t_i + \varepsilon_{y,i}^*, \\z_i^* &= t_i + \varepsilon_{z,i}^*,\end{aligned}\tag{2-20}$$

where the flagged variables demonstrate the fact that we refer to the transformed datasets  $x$ ,  $y$ , and  $z$  and to the transformed error estimates  $\varepsilon^*$ .

To eliminate the unknown "truth" the difference between the datasets  $x_i^*$ ,  $y_i^*$  and  $z_i^*$  is computed as

$$\begin{aligned}
x_i^* - y_i^* &= \varepsilon_{x,i}^* - \varepsilon_{y,i}^*, \\
x_i^* - z_i^* &= \varepsilon_{x,i}^* - \varepsilon_{z,i}^*, \\
y_i^* - z_i^* &= \varepsilon_{y,i}^* - \varepsilon_{z,i}^*.
\end{aligned}
\tag{2-21}$$

Finally, the variance of the residual errors  $\langle \varepsilon_x^{*2} \rangle = \tau_x^{*2}$  can be retrieved by the cross-multiplications of the equations (2-21) assuming a sufficiently large population (signalized by the angle brackets) (Zwieback et al., 2012). If the errors are uncorrelated their covariances  $\langle \varepsilon_x^* \cdot \varepsilon_y^* \rangle$ ,  $\langle \varepsilon_x^* \cdot \varepsilon_z^* \rangle$  and  $\langle \varepsilon_y^* \cdot \varepsilon_z^* \rangle$  are equal 0 and the residual errors can be expressed as:

$$\begin{aligned}
\tau_x^{*2} &= \langle (x_i^* - y_i^*)(x_i^* - z_i^*) \rangle \\
\tau_y^{*2} &= \langle (x_i^* - y_i^*)(y_i^* - z_i^*) \rangle \\
\tau_z^{*2} &= \langle (x_i^* - z_i^*)(y_i^* - z_i^*) \rangle.
\end{aligned}
\tag{2-22}$$

The scaling coefficients  $a$  and  $b$  cannot be solved since the true soil moisture value  $t$  is never known. We therefore select one of the datasets,  $x$  in this case, as a reference assuming the scaling coefficient  $a$  and  $b$  equal to 0 and 1, respectively. This allows for solving of the other calibration coefficients  $a_y$ ,  $b_y$ ,  $a_z$  and  $b_z$  and for solving of the equations (2-22).

The absolute magnitudes of the residual errors will reflect the climatology of the reference dataset according to  $\tau_{x,i}^* = \frac{\tau_{x,i}}{b_x}$ . Since the calibration coefficients influence the computation of residual errors an iterative scheme for its derivation is recommended that takes into account errors in both datasets (Scipal et al., 2008b). If all prerequisites are met the relative patterns of the residual errors should remain stable for any reference dataset (Dorigo et al., 2010).

As already mentioned the final errors are expressed in the climatology of the reference dataset. To be able to compare the results it is recommended to keep the computed errors in the dynamics of the reference dataset (Dorigo et al., 2010). If needed, the inversion to the original dynamics is possible using the equation (2-19).

Note, while more than 500 samples were recommended to achieve the relative uncertainty of 10% (Zwieback et al., 2012), evaluation studies usually adopt the threshold of 100 triplets (Dorigo et al., 2010; Scipal et al., 2008b).

The introduced TC technique can be computed using the absolute values of soil moisture dataset (Stoffelen, 1998) as well as soil moisture anomalies (Dorigo et al., 2010; Miralles et al., 2010). The results of such studies differ in their interpretation. The first refers to the capability of the product to estimate the overall seasonality of the soil moisture product (e.g. wet and dry season),

while the anomaly-based approach gives an estimate on the ability of datasets to capture single drying and wetting events (Dorigo et al., 2010).

### 2.2.3 The next stage: how to evaluate quality of the evaluation measures?

An increasing usage of diverse evaluation measures for soil moisture datasets has been demonstrated in the last decade (Dorigo et al., 2010; Naeimi et al., 2009b; Parinussa et al., 2011; Scipal et al., 2008b). To understand divergences and similarities of the errors resulting from the different evaluation measures these need to be qualitatively or quantitatively compared. This chapter introduces a quantitative method to assess quality of the EP method (section 2.2.3.1 and a qualitative comparison of several methods estimating random and systematic errors (section 2.2.3.2).

#### 2.2.3.1 Predicted RMSE

This method estimates the RMSE from the error characteristics of two datasets  $x$  and  $y$  with independent error structures (Taylor, 1997) and as thus represents addition of random errors of both datasets. The method combines the Gaussian EP method with the RMSE estimate (equations (2-6) and (2-18)). Given the independence of the observed and predicted RMSE, their high correspondence is in this thesis used to assess quality of the individual error estimates estimated by the EP method.

The error of  $x_i$  is equal to  $\varepsilon_i = x_i - y_i$ , where  $x$  represents the evaluated dataset and  $y$  the reference dataset, and  $i = 1, \dots, N$  and  $N$  is the total number of elements. RMSE is then equal to:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2}. \quad 2-23$$

The critical assumption of the predicted RMSE is that  $\sum_{i=1}^N (\varepsilon_i)^2 / n = \langle \varepsilon^2 \rangle$ , where the angle brackets represent the mean over time. RMSE can therefore be expressed as:

$$RMSE = \sqrt{\langle \varepsilon^2 \rangle}. \quad 2-24$$

Given that errors in  $x$  and  $y$  datasets are independent and constant, the Gaussian EP method (equation (2-18)) can be applied to estimate the error of  $\varepsilon = f(x, y)$  as:

$$\langle \varepsilon^2 \rangle = s_x^2 + s_y^2, \quad 2-25$$

where the variables  $s_x$  and  $s_y$  represent the standard errors of the datasets  $x$  and  $y$ , respectively. The partial derivatives of the parameters  $x$  and  $y$  are equal to  $\frac{\partial \varepsilon}{\partial x} = \frac{\partial \varepsilon}{\partial y} = 1$  and are neglected in equation (2-25).

By a combination of the equations (2-24) and (2-25) one receives that:

$$RMSE(x, y) = \sqrt{s_x^2 + s_y^2}. \quad 2-26$$

The equation (2-26) demonstrates that evaluated as well as reference datasets contain errors. This finding has been prior highlighted in another study (Stoffelen, 1998). For a good estimate of RMSE the errors of the independent datasets need to be known.

The equation (2-26) was applied to estimate the RMSE between two coarse resolution datasets (Wagner et al., 2007) or to assess the impact of scaling error introduced by point in-situ measurements (Miralles et al., 2010).

#### 2.2.3.2 Qualitative comparison of RMSE, EP and TC

A large variety of evaluation methods have been introduced in section 2.2. Interestingly, a large portion of these assessed absolute errors (predicted RMSE, observed RMSE, EP, and TC). Their absolute values are however expect to differ due to the different weight they give to systematic and random errors. This section assesses such differences and provides their possible causes. Understanding the similarities and differences between evaluation measures and methodologies allows for an efficient selection of an adequate combination of evaluation methods and helps to avoid redundancy in evaluation studies.

Importantly, in the discussion below a good quality of the evaluated algorithm is assumed.

The observed RMSE (equation 2-6) refer to a combination of time-variant systematic and random errors of two evaluated datasets whereas the predicted RMSE (2-26) only takes into account random errors of both datasets (this is given by the fact that the separate standard errors are estimated using the EP scheme). The latter assumes that the time-invariant systematic error (bias) to be removed. If all conditions are fulfilled for the predicted RMSE and if the time-variant systematic error is minimal the expectation is that both RMSEs are equal. The predicted RMSE introduces strict simplifications on the individual errors  $\varepsilon_i$ , namely that  $\sum_{i=1}^N (\varepsilon_i)^2 / n = \langle \varepsilon^2 \rangle$ . This assumption is expected to be violated in areas where the individual parameters  $x$  and  $y$  are highly variant (Willmott & Matsuura, 2005). This may result in predicted RMSE < observed RMSE.

Importantly, the absolute value of predicted RMSE is dependent on the estimation of errors of individual datasets; these are estimated using the EP method. As a result, the absolute range of is highly influenced by the quality of the estimated EP error ( $s$ ). Importantly, the  $s$  error estimates need to be performed only after data transformation to account for a possible systematic bias between the datasets.

### *EP and RMSE*

The  $s$  assesses random errors solely; it doesn't account for the effect of systematic errors. This is on the contrary to RMSE that reflect a) systematic error (the time-variant portion of it, see section 2.1.1 for explanation of time-variant error), and b) random errors of both datasets. As a result, it is expected that  $s < \text{RMSE}$ . The spatial differences between RMSE and  $s$  may be accounted a) to the time-variant systematic error of the second dataset, or b) to the missing or wrongly propagated parameter in the EP method. Importantly, the absolute value of  $s$  is dependent on the quality of each of the detected errors.

### *TC and EP*

The TC method estimates the combined effect of random error and systematic error alternating in time (e.g. due to missing parameter) of the second and third evaluated dataset. As such the relative TC error  $\tau^*$  is expected to differ from  $s$  mainly due to the assessment of time-variant systematic error.

### *TC and RMSE*

The essential difference between  $\tau^*$  and RMSE is the usage of an additional dataset in the TC computation. While RMSE relies on a second power of difference between two datasets (6.1-1), the TC error uses a multiplication of differences between three datasets (6.1-10). As such RMSE reflect a combination of errors of both datasets. On the contrary, the effects of a second and a third dataset are mitigated (cancel each other) in the  $\tau^*$  computation due to their independency. The latter explains why exchanging one dataset in the TC method doesn't change the relative patterns of the residual errors (Dorigo et al., 2010).

Further, the advantage of the TC method is that it assess the relative quality of several products simultaneously (Dorigo et al., 2010). It does so by assuming that the closer together are the products the higher is the probability that these represent the studied phenomenon. This is only possible if all errors of the datasets are independent. Violation to the latter may introduce pseudo biases (Zwieback et al., 2012). In particular, the estimated  $\tau$  of the dependent datasets would decrease while the error of the independent datasets would increase. Increasing the number of datasets may reduce such pseudo bias. Further, studying changes introduced by switching one dataset in the TC computation is also a good way how to investigate dataset independency.

### *Effect of scale*

In evaluation studies, datasets with different spatial resolutions need to be implemented, often because no other dataset with comparable scale is available. Given the high radiometric quality of the coarse resolution sensors (e.g. AMSR-E or ERS) also the SSM products derived from these sensors are expected to have lower errors comparable to the medium (1-5 km) resolution datasets (e.g. ASAR GM). Second reason for the low errors of the coarser resolution datasets may be the fact that the parameterization of the models becomes simpler when moving to coarser scales suggesting existence of so called "effective states" (Settin et al., 2007). Third, the medium resolution datasets may exhibit a shift or small geocoding mismatches that may cause that the forcing effect influencing one dataset counteract the forcing effects of the other. This is not expected to influence coarse resolution datasets that are effected by large-scale atmospheric forcings (Vinnikov et al., 1999) and correlate to local soil moisture patterns according to concept of temporal stability (Cosh et al., 2004).

Given all the above reasons the effect of scale needs to be carefully considered when interpreting evaluation results and was therefore given a special attention in separate section (section 5.2).

Below, the findings of this chapter are summarized (these are graphically incorporated in Table 2):

- The possible sources of discrepancies between the relative spatial patterns of  $s$  and  $\tau^*$  can be accounted to the fact that: a) the residual errors in  $\tau^*$  are not independent, or b) the estimates of the input variable errors  $s$  in the EP method are not correct or missing.
- The possible spatial differences between RMSE and  $s$  can be accounted either to the time-variant systematic error of the second dataset used in RMSE or to the missing or wrongly propagated parameter in the EP method.
- The possible spatial differences between RMSE and  $\tau^*$  can be accounted either to a) the time-variant systematic error of the second dataset used in RMSE, or b) to the fact that the residual errors in the TC method are not independent.
- The absolute values of the four methods are expected to compare as follows:  $s < \tau^* < \text{RMSE}$ . This order is caused by the fact that RMSEs include time-variant systematic error of the second evaluation dataset,  $s$  doesn't assess the impact of systematic error of interpretation (correctness of the model), and the error of the second and third dataset in  $\tau$  is mitigated due to their independency.
- The absolute value of  $s$  is highly dependent on the quality of each of the detected errors. This can rapidly change the position of  $s$  in the above ranking.
- The assumptions on the statistical characteristics of the datasets and errors are crucial for the validity of the method.
- The effect of different spatial resolution needs to be considered when interpreting results of RMSE, EP, and TC.

Table 2. *The list of reasons for possible discrepancies between advanced evaluation methods.*

The possible differences can be explained by the:	$\tau^*$	$s$
RMSE	time-variant systematic error error of the second dataset in RMSE lacking independency of the residual errors in the TC method	time-variant systematic error error of the second dataset in RMSE estimates of the input variables in the EP method are not correct missing variable in the EP method
$\tau^*$		lacking independency of the residual errors $\varepsilon^*$ in the TC method estimates of the input variables in the EP method are not correct missing variable in the EP method

## 2.3 The selection of appropriate evaluations measures and methods

This chapter a) provides the state of the art in the application of soil moisture datasets, b) identifies (from the existing literature) the dataset qualities impacting the success or the failure of each application, and c) provides a set of evaluation measures that are expected to address these qualities (see section 2.2 for the implication of the evaluation measures). The evaluation measures are summarized in Table 3.

Importantly, while in this section it was assumed that the actual application of the SSM data is known, this is often not the case because: a) the dataset is only an experimental product or b) numerous applications of the dataset are expected (e.g. planning a new satellite mission). In such situations another evaluation approach is needed (see section 5.3).

No soil moisture estimate, neither from remote sensing, nor from a model, nor from in-situ, represents the absolute “truth”. As a result, the performance of soil moisture in models is influenced by the quality of a remotely sensed soil moisture dataset as well as by the quality of a model and the scaling errors of the in-situ measurements (Dharssi et al. 2011; van Dijk and Renszullo 2011). This chapter summarizes efforts made to evaluate quality of soil moisture products; it only marginally discusses limitations of models or a usefulness of soil moisture in models.

In the long history of evaluation studies of soil moisture datasets, two main trends dominated; the first stresses the importance of the absolute correspondence of soil moisture acquisitions [provided usually in  $\text{m}^3/\text{m}^3$ ] (e.g. Jackson et al. 2010), while the second stresses the ability of the product to capture the relative dynamics (e.g. Drusch 2007; Entekhabi et al. 2010).

*Evaluation using absolute measures* The main goal of the evaluation using absolute measures is to retrieve RMSE through a direct comparison of the remotely sensed data with in-situ or modelled data. An example may be the requirement on the SMOS level 3 product setting up a condition of  $\text{RMSE} < 0.04 \text{ m}^3/\text{m}^3$  (Kerr et al., 2010). Numerous other requirements for soil moisture provided by WMO are provided as an absolute error measure (WMO, 2012). Moreover, these rely on only one number without specifying land cover for which this number should be applied. However, the meaning of the absolute difference is uncertain as it is a combined effect of variances of the datasets, variable scales of hydrologic processes, as well as the depth that these represent. Also, the poor quality of the soil maps may enlarge these absolute differences (Koster et al., 2009). For the latter reason transformation measures need to be applied prior to data evaluation. Such measures serve to mitigate effects of different scales, depths, units or soil maps and were in detail summarized in section 2.1.

*Evaluation using relative measures* Contrary to RMSE, the correlation coefficient between modelled, in-situ and remotely sensed data should in general agree across different spatial scales because temporal changes in soil moisture are driven by atmospheric processes identical across several scales (Entin et al., 2000). This obviously holds only over areas where soil moisture varies. Such an effect is referred to as a “temporal stability” concept and is the reason why coarse resolution datasets and point measurements correlate.

The temporal stability concept formed the basis for the second evaluation trend stressing the relative dynamics of a product. This approach is based on the principle that as long as the products represent the relative dynamics well they can be biased in their mean and dynamic range and still be useful (Entekhabi et al., 2010).

*Evaluation based on application* Recently, an evaluation trend combining the latter two strategies emerged (Bellocchi et al., 2011; Crow et al., 2012; Entekhabi et al., 2010). This recommends selecting the most appropriate evaluation measure according to the technique used in the application (e.g. data assimilation, a direct input). This approach relies on the fact that no single metric can capture all attributes of environmental variables. For instance, some techniques may require a good ability of the soil

moisture product to capture the relative evolution in time of soil moisture (e.g. data assimilation), others may require a linear relationship with the “truth” or ability to distinguish between two soil moisture stages (e.g. soil moisture evaporation).

Some first evaluation studies performed evaluation according to the application needs (i.e. Brocca et al. 2010; Brocca, Moramarco, et al. 2011; C. S. Draper et al. 2012). In particular, the latter studies implied a correlation coefficient to judge the relevance of the dataset for data assimilation. Other studies addressed the diverse application needs by providing a combined assessment of both, absolute and relative evaluation measures (Draper et al., 2009; Entekhabi et al., 2010; Verstraeten et al., 2010).

This chapter discusses and further evolves the ‘evaluation based on application’ (EbA) approach *The EbA strategy* in evaluation studies (Entekhabi et al., 2010). Motivated by Entekhabi’s study, a set of necessary steps for the selection of the most appropriate evaluation measure is recommended:

- a) specify how soil moisture data are used in an application
- b) define the soil moisture data requirements of the method (i.e. a low absolute difference, a good relative correspondence, a good quality of soil moisture within/below/above certain threshold)
- c) select an evaluation measure(s) that describes the above relationship
- d) compare the results of the selected measure for several soil moisture data sources
- e) select the best performing dataset

Even though EbA method seems quite demanding, the usage of it in practice should be possible given that few techniques are used to implement datasets into models. Contra versa, it may be hampered by the not clearly definable requirements of some applications on the soil moisture dataset. For a demonstrational purpose Table 3 summarizes qualities needed for four major SSM applications and suggests set of evaluation measures.

### 2.3.1 Sequential data assimilation

Data assimilation is an analysis technique commonly used to apply remotely sensed soil moisture *Assimilation schemes* into numerical weather prediction (NWP), seasonal forecast (SF), river-runoff (RR) or crop yield models. It is here defined as a technique that jointly estimates a variable from observed data and model. The modelling of land-surface processes requires certain simplifications that often lead to systematic errors in the modeled soil moisture fields. An example of such a simplification is the computation of the precipitation field as an average of several in-situ acquisitions. Data assimilation helps to overcome the latter weaknesses by introducing remotely sensed datasets that, comparable to the models, represent a more direct measure of surface soil moisture and operate at corresponding resolutions.

The assimilation techniques differ from very simple weighting of differences (i.e. nudging) to complex sequential assimilations (e.g. the Extended Kalman filter). The nudging scheme was successfully applied mainly in NWP and RR models. It adjusts the modelled measurement towards the remotely sensed measurement by adding a weighted ( $K$ ) difference between the two estimates. The  $K$  is usually temporally and spatially stable and, in case of optimal nudging, can be determined by minimizing some cost function (i.e. 4D-VAR). Nudging doesn’t include any potential random errors of the satellite data, and in this respect the scheme assumes that the

observations are highly accurate with respect to the model (Drusch, 2007). On the contrary, the Kalman filter and extended Kalman filter predict the soil moisture state sequentially for every time step and correct each prediction according to the remotely sensed estimate and its initial error.

### *2.3.1.1 Studies*

There are numerous studies that assimilate active and passive microwave surface soil moisture products in numerical weather predictions models, river-runoff models and recently also, for instance, in crop yield models. The studies implement nudging schemes (Dharssi et al., 2011; Drusch, 2007; Pauwels et al., 2002; Scipal et al., 2008a), simple Kalman Filter (Mahfouf, 2010), and Ensemble Kalman Filters (Reichle et al., 2008a, 2008b) methods. Data assimilation techniques are also commonly applied to improve rainfall-runoff estimates (Aubert et al., 2003; Draper et al., 2011; Francois et al., 2003; Pauwels et al., 2001), crop yield modelling (Verstraeten et al., 2010), or improving rainfall estimates (Crow et al., 2009).

#### *NWP and RR modelling*

In synthetic experiments, the skill of assimilated products exceeded the skill of the model acting alone in the majority of cases (Reichle et al., 2008b). Nevertheless, the positive impacts of the assimilated product in real experiments differed according to time of the year (Francois et al., 2003), land cover type, and topography (Matgen et al., 2011) or climatic conditions (Wooldridge et al., 2003). In particular, the improvements were demonstrated over sparsely-vegetated and shallow-rooted vegetation. On the contrary, the limited success of soil moisture assimilation was found over densely vegetated areas and areas with deeply rooted vegetation. A common result in initial RR modelling was an improvement of surface soil moisture but no or small improvement of river runoff (Parajka et al., 2006, 2009).

Several reasons may have caused the assimilation of real data to fail in contrast to synthetic data: a) the real errors are usually not Gaussian, b) can be expected to change with time, and to start with c) the fact that the error characteristics are often unknown. The latter provides a further evidence about the criticality of error assessment studies such as this thesis.

#### *RR modelling*

Contrary to NWP there seems to be no obvious explanation in RR studies that would clarify under which conditions an improvement can be achieved. Recently, several studies demonstrated positive impact when assimilation soil water index (SWI) (Wagner et al., 1999b) (Brocca et al., 2010a, 2010c; Matgen et al., 2011; Meier et al., 2011). SWI represents the profile soil moisture in the root zone which is the hydrologically most important zone in terms of runoff generation (Parajka et al., 2006).

#### *Vegetation modelling*

A positive effect of assimilation was also demonstrated in vegetation and carbon modelling. The ability of SWI to determine the limitation of available water and correctly present the low soil moisture values is especially important since these limit soil respiration and photosynthesis (Verstraeten et al., 2010; de Wit & van Diepen, 2007)(Verstraeten et al., 2010). Furthermore, a recent study (Van der Velde et al., 2011) highlighted the importance of detection of excessive soil moisture by remote sensing products as this limits crop development due to increased pest and disease development.

Still, there seem to be a need for improvements in both hydrological models as well as in remotely sensed datasets (Brocca, Moramarco, et al. 2011; Bronstert et al. 2012). The critical

aspects to be addressed include a) the selection of the adequate data assimilation techniques, b) a proper structure of the model (in particular a strong coupling of the surface and the deeper soil layer (Kumar et al., 2009)), c) the high accuracy and temporal resolution of the remotely sensed data and d) the correspondence of the depth of the assimilated products.

### 2.3.1.2 Evaluation measures and methods

The correspondence of the relative dynamics between the model and remotely sensed data is more important than the correspondence of absolute values (Drusch, 2007; Reichle et al., 2008b). The absolute accuracy of the soil moisture data in data assimilation is of lesser importance mainly because the absolute values are calibrated before use by the models. The calibration is performed using the wilting point and field capacity (Drusch, 2007) and/or methods such as linear regression or non-linear matching (Brocca et al., 2011). *The relative correspondence of datasets*

To assess the relative correspondence between datasets the combination of Spearman correlation coefficients ( $R_s$ ) and MAE can be used.  $R_s$  is preferred to the Pearson's coefficient ( $R$ ) because it doesn't require linearity on datasets and is not affected by the absolute magnitude of soil moisture dataset. As such, it refers to the ability of datasets to capture anomalies rather than absolute seasonality which is needed for short-term forecasting (Reichle et al., 2008b). It is anticipated, that similar effects can be achieved by computing Person's correlation coefficient on anomalies. The MAE should be computed on transformed datasets and is a good complement to  $R_s$  as it assess independent quality of the dataset - dataset precision.

As it assesses both, relative and absolute error, combination of  $R$  and MAE is an evaluation strategy common to several application methods. This is commonly supplemented by additional measures or limited to certain data ranges (see following sections).

Knowledge of random errors is required in several assimilation techniques. The traditional method to estimate observational error examines MAE between evaluated data and a reference dataset. The latter is assumed to represent or to be close to the 'truth'. Nevertheless, assumptions about what is true may introduce a substantial pseudobias effect (Stoffelen, 1998). An improved method that solves initial error is the adaptive filter (Reichle et al., 2008b) that estimates and continuously updates the observations error parameters according to the innovative functions. Nevertheless, implementation of the triple collocation technique appears computationally more feasible (Dorigo et al., 2010; Scipal et al., 2008b). *Initial error understanding*

The results of assimilation studies in hydrological models were significantly better when profile soil moisture data were implemented (Brocca et al., 2012). There are several possible reasons for the latter. First, in hydrological modelling the profile soil moisture is generally better understood than shallow soil moisture parameter. Secondly, the remotely sensed surface soil moisture represents an extremely shallow layer (< 5 cm) that is problematic to simulate in hydrological models and is also difficult to acquire with in-situ instruments. Third, as recently suggested by Brocca et al. (Brocca et al., 2012), the improved assimilation results could be caused by the reduced noise of the profile soil moisture. Fourth, the profile measurements better represent the areal mean as the spatial scaling errors are smaller. *Assimilate corresponding layers*

In general, a good correspondence of the depths of the assimilated and modelled dataset seems essential to achieve improvements in parameters of the hydrological models. Different soil

moisture layers were demonstrated to have characteristic reaction time and lag-time that could be measured using a simple correlation  $R$  as well as autocorrelation function (Rebel et al., 2012). Autocorrelation analyses were less sensitive to outliers and appear better suited to evaluate the correspondence of soil depth layers.

*Problems of CDF matching* The absolute accuracy is of limited importance for data assimilation. As a result, data transformation became a required pre-processing step for the data assimilation studies. The CDF matching technique is commonly applied in assimilation studies (Drusch et al., 2005; Reichle & Koster, 2004) because it has the ability to assign together data with different soil moisture dynamics as well as with different data distributions. Importantly, modellers should stay away from matching data with significantly different distributions as this may indicate that these represent different phenomena. Also, it is important not to overfit data which may introduce additional bias to the models.

*Independency of model parameters* The main justification in using satellite acquisitions in data assimilation is the fact that the error structures of the model and remotely sensed dataset are hoped to be independent and that their combination may result in less biased parameter estimation and so exhibit less random errors. The majority of studies however only assume but do not quantify such independency.

## 2.3.2 A direct input into Environmental Process models

### 2.3.2.1 Studies

*SSM to replace a hydrological parameter* Brocca (Brocca et al., 2009) and Beck et al. (Beck et al., 2009) investigated the ability of remotely sensed soil moisture to simulate the maximum retention parameter ( $S$ ) in the rainfall-runoff hydrological model over Italy's Tibera river and numerous Australian catchments, respectively. Ignoring other factors,  $S$  value refers to the infiltration rate and this is expected to be closely related to the antecedent wetness condition in the catchment. Brocca et al. found SWI derived from ASCAT scatterometer highly correlated with  $S$  and antecedent wetness over the Tibera basin and suggested that both can be reliably estimated using SWI parameter. Similarly, Beck et al. investigated the potential to improve  $S$  estimates using AMSR-E radiometer data and demonstrated an improved performance of the model that was as large as 0.25. Importantly, the SWI must have been calibrated prior to the inclusion to the model.

Better results were achieved in dry catchments with low topographical relief and over areas where the precipitation datasets were of low quality. The original method of derivation of  $S$  uses the actual runoff for the parameter calibration. Larger runoff variations are therefore expected to exhibit better prediction of  $S$ . The correlation between SWI and  $S$  was demonstrated as a relevant proxy for improving runoff estimates.

*SWI to predict NDVI* Zribi (Zribi et al., 2010) employed soil moisture from the ERS scatterometer data to propose a simple linear model for prediction of the NDVI values one month in advance at large spatial scales. Such empirical relationships could be very useful for vegetation forecast development, without the need for complex physical models. Importantly, such relationships only occur over regions with water limited vegetation types.

### *2.3.2.2 Evaluation measures and methods*

The latter studies used SWI directly. In particular, they calibrated the modelled parameter  $S$  (maximum retention parameter) and NDVI using SWI.

Given the assumption of linear regression between the remotely sensed SWI and the retention parameter  $S$  and NDVI the  $R_s$  and MAE appear as an appropriate combination of evaluation measures. The linearity can be evaluated by studying the difference between Spearman and Pearson correlation coefficients. In particular, a Spearman correlation coefficient higher than the Pearson coefficient is a sign of non-linearity. The opposite behaviour can be explained by the influence of seasonality on Pearson's correlation coefficient.

Very low  $R_s$  would signalize a low correspondence between the parameters  $S$ /NDVI and the SSM dataset and demonstrates a low likelihood of remotely sensed SSM to improve  $S$  and NDVI estimates. On the contrary, very high  $R_s$  values can be expected to have very limited space for an improvement of the parameters  $S$  and NDVI. Setting up the  $R_s$  threshold of possible improvement is probably the most challenging task. Inverse law applies on the MAE measure.

### **2.3.3 Distinguishing between different soil moisture levels**

#### *2.3.3.1 Studies*

Several crop yield and carbon flux monitoring studies highlighted the importance of SWI dataset to capture dynamics in lower soil moisture ranges (i.e. de Wit and van Diepen 2007; Entekhabi et al. 2010; Verstraeten et al. 2010). In these applications the limited water in the soil is essential as it limits the rates of photosynthesis, crop yield, evapotranspiration and frequently also the net ecosystem production (NEP). For evaluation purposes for such applications the ability of soil moisture to capture the relative dynamics at very low (< 30%) levels is essential whereas the ability to capture soil moisture in the near-saturated portion of the range is of limited importance.

Matgen (Matgen et al., 2011) presented an approach to improve RR models that is linked only to certain soil moisture levels. In particular, the effective field capacity is defined at which any additional soil water contributes to the generation of rapid water flow. The ASCAT SWI could reliably distinguish between the two main states under and above the effective field capacity and demonstrated so a good ability for assimilation into rainfall-runoff models.

Similarly, Sass et al. (Sass et al., 2012) demonstrated the ability of SAR soil moisture estimates to detect different levels of species richness of vascular plants. In particular, the soil moisture dataset was divided into three essential classes for species richness - unsaturated, saturated and inundated.

#### *2.3.3.2 Evaluation measures and methods*

Simple statistics such as 'missed alarms', 'false alerts' or probability of detection (Dinku et al., 2008) can be used to define the ability of dataset to fall into a specific range.

Furthermore, to capture the aspired quantities within a defined range, the  $R_s$  combined with MAE seem as appropriate evaluation measures. These should be computed on the values lying

within required ranges according to each application. The Spearman correlation is recommended as it is not affected by a dataset magnitude and seasonality.

### 2.3.4 Anomaly computation

#### 2.3.4.1 Studies

The ability to detect soil moisture anomalies can be a strong asset in seasonal rainfall forecasts (Koster et al., 2004). A close link between soil moisture anomalies and subsequent precipitation distribution and magnitude has been demonstrated (Kim & Wang, 2007; Koster et al., 2004; Kuenzer et al., 2009).

In particular, Koster et al. (Koster et al., 2004) demonstrated that the main links between soil moisture and precipitation are in the transition zones between wet and dry climates. Kim et al. (Kim & Wang, 2007) demonstrated in different parts of the world that precipitation patterns are a) only dependent on soil moisture during certain times of the year, b) dependent on soil moisture linearly but only within a certain range, c) responsive for longer to dry than to wet soil moisture anomalies, and d) more persistently responsive to soil moisture anomalies at coarser scale than to those at smaller scales. Lastly, Kuenzer et al. (Kuenzer et al., 2009) showed that ENSO related floods and droughts can be depicted with the scatterometer-retrieved SWI anomaly maps derived from this time series in different geographical regions.

#### *SWI to predict vegetation dynamics*

Soil moisture anomalies affect also the vegetation stage. Gouveia (Gouveia et al., 2009) studied the impact of soil moisture on vegetation dynamics by analyzing monthly anomalies of SWI from ERS scatterometer and by studying the annual cycle of SWI versus NDVI. The impact of the soil moisture limitation was demonstrated over arable land and forest, with higher impact on arable land. As expected, the impacts of SWI were also related to the timing and duration of the events.

#### 2.3.4.2 Evaluation measures and methods

In the presented applications the ability of datasets to detect anomaly (below or above normal stage) is essential. The recommended evaluation measures for anomaly detection are 'missed alarms', 'false alerts' or probability of detection.

The actual magnitude of the anomaly was addressed only as secondary factor. For its evaluation the  $R$  is preferred to  $R_s$  as it is computed directly on the anomalies (not on ranks) and address the effect of their magnitudes. A higher correspondence, means a higher probability that these detect the anomalies correctly.

Lastly, the detection of soil moisture dry anomalies seems more important than the detection of wet anomalies. Also, coarse resolution products may be more appropriate for anomaly detection because only coarse resolution scale anomalies led to persistent effects on precipitation. Lastly, if certain applications demonstrated benefit for precipitation or vegetation monitoring only within certain bounds the above recommended evaluation measures should only be computed within this range.

Table 3. A list of applications and their recommended evaluation measures and methods.

APPLICATION'S METHOD	APPLICATION	REQUIREMENT	EVALUATION MEASURES and METHODS
Data assimilation	NWP, SC, RR models, crop yield and carbon monitoring, rainfall improvements	A good relative correspondence of remotely sensed and modelled soil moisture datasets	<ul style="list-style-type: none"> <li>• <math>R_S</math></li> <li>• MAE</li> </ul>
		Background and observation errors	<ul style="list-style-type: none"> <li>• TC</li> <li>• EP</li> </ul>
		An assurance that corresponding depth layers are compared, (i.e. corresponding reaction times and lag-times)	<ul style="list-style-type: none"> <li>• Autocorrelation studies</li> </ul>
		Check the independency of error structures of the model and remote sensing dataset	<ul style="list-style-type: none"> <li>• Check on independency of the errors (using EP method)</li> </ul>
A direct input into environmental models	RR models	A good relative correspondence and linearity achieved between the remotely sensed and required parameters	<ul style="list-style-type: none"> <li>• <math>R_S</math></li> <li>• Limited difference between <math>R</math> and <math>R_S</math></li> <li>• MAE</li> </ul>
Distinguishing between different soil moisture levels	RR models, crop yield and carbon models, vegetation monitoring	A good relative correspondence of remotely sensed and modelled soil moisture datasets within a certain range	<ul style="list-style-type: none"> <li>• <math>R_S</math> computed within a certain range of data</li> <li>• MAE computed within a certain range of data</li> </ul>
		An ability to distinguish stages below and above a certain threshold	<ul style="list-style-type: none"> <li>• 'missed alarms', 'false alerts' or probability of detection</li> </ul>
Anomalies as input to models	Drought assessment, land-atmosphere interaction	A good relative correspondence of remotely sensed and modelled soil moisture datasets within a certain range	<ul style="list-style-type: none"> <li>• <math>R</math> computed on anomalies</li> <li>• MAE computed on anomalies</li> </ul>
		A detection of anomalies	<ul style="list-style-type: none"> <li>• 'missed alarms', 'false alerts' or probability of detection</li> </ul>

### 2.3.5 Conclusion

A state of the art in the application studies of soil moisture was provided. A variety of needed qualities of the SSM data was demonstrated and supplemented with the evaluation measures that well describe these qualities. A majority of demonstrated applications considered essential the ability of the remotely sensed soil moisture datasets to capture the relative dynamics while maintaining a high dataset precision. For some applications, this quality applies only to a certain soil moisture range. Other applications required a dataset to fall below or above a certain threshold. Interestingly, none of the applications required solely an understanding of absolute soil moisture values.

The EbA strategy was recommended that supports selection of the evaluation methods for each particular application (see introduction in section 2.3). In step d) of EbA, spatial maps were retrieved that map the result of each evaluation measure. Importantly, these maps need to be interpreted with a consideration of a probability density function (PDF). For instance, soil moisture measurements will need to be more precise for dry regions comparable to regions with high soil moisture variation. A generation of such PDF maps (describing, for example, soil moisture mean and standard deviation) has been proposed by Entekhabi et al. (Entekhabi et al., 2010). A full discussion of weighting the PDF function and an evaluation measure is beyond the scope of this work.

The actual selection of the best performing dataset is to be accomplished by the data users themselves based on the consideration of a) the spatial maps demonstrating results of the appropriate evaluation measures and methods, b) the PDF function, and c) the biogeophysical processes influencing the usefulness of the soil moisture in the particular region.

## 3. Methodology

### 3.1 Datasets

#### 3.1.1 ASAR GM soil moisture dataset and processing

The ASAR GM SSM represents the relative surface soil moisture in upper > 3-5 cm of soil retrieved from active radar sensor at the 1 km spatial resolution. The ASAR GM dataset was retrieved using a change detection algorithm (Pathe et al. 2009) assuming a sufficiently long time series cover a full range of soil moisture values from wilting point to saturation (Pathe et al. 2009; Wagner et al. 1999a). The algorithm was initially developed for a derivation of soil moisture product for the ERS and ASCAT scatterometers (Wagner et al., 1999a). Its transformation to ASAR GM was possible due to the identical operating wavelengths and frequencies of coverage of the ASAR GM radar and the ERS and ASCAT scatterometer instruments. Conversely, the algorithm transformation was challenged due to differences in spatial and radiometric resolution, polarization as well as differences in image retrieval geometry. The units of the resulting product is % of saturated soil moisture.

The ASAR GM change detection model defined (Pathe et al. 2009) as:

$$\theta_s = \frac{\sigma^0(\Theta, t) - \sigma_{dry}^0(30) - \beta(\Theta - 30)}{\sigma_{wet}^0(30) - \sigma_{dry}^0(30)}, \quad 3-1$$

where  $\sigma_{dry}^0(30)$ ,  $\sigma_{wet}^0(30)$ ,  $\beta$  and  $S$  are considered constant in time and represent respectively the dry and wet reference at medium incidence angle  $30^\circ$  ( $\Theta$ ), the slope, and the sensitivity of the ASAR GM backscatter to soil moisture. The slope quantifies the dependence of sigma nought on the local incidence angle. Importantly, the slope parameter has been used, for instance, for normalization of the backscatter acquisitions to  $30$  degree incidence angle what is required for computation of the dry and wet references (see explanation below). The  $\sigma^0(\Theta, t)$  stands for the backscatter values at an incidence angle  $\Theta$  in time  $t$ .

The dry and wet references have been derived using a similar methodology to that of Pathe et al. (Pathe et al., 2009b). Respectively, the references are computed as an average of measurements taken during wet and dry conditions, respectively as:

$$\sigma_{dry}^0(30) = \frac{1}{N_{dry}} \sum_{i=1}^{N_{dry}} \sigma_t^0(30), \quad 3-2$$

and

$$\sigma_{wet}^0(30) = \frac{1}{N_{wet}} \sum_{i=N-N_{wet}}^{N_{dry}} \sigma_t^0(30), \quad 3-3$$

where  $N_{dry}$  and  $N_{wet}$  is the number of measurements removed from the ranked ASAR GM measurements to retrieve the dry and wet reference and  $t$  refers to backscatter acquisitions in time. These were expected to correspond to data outliers.

The algorithm differs from that introduced by Pathe et al. (Pathe et al., 2009b) in how it retrieves  $N_{dry}$  and  $N_{wet}$ . While Pathe based his selection on the ratio computed between the number of ERS extreme (lowest and highest 5%, respectively) soil moisture acquisition to the total number of ERS soil moisture acquisition, in this study,  $N_{dry}$  and  $N_{wet}$  are computed as follows:

$$N_{dry} = N_{GM} \frac{N_{ERS} \in interval_{dry}}{N_{ERS}}, \quad 3-4$$

and

$$N_{wet} = N_{GM} \frac{N_{ERS} \in interval_{wet}}{N_{ERS}}, \quad 3-5$$

where  $N_{GM}$  and  $N_{ERS}$  represent the total amount of ASAR GM and ERS backscatter measurements, respectively. The intervals for the dry and wet reference are defined as:

$$interval_{dry} = [\min(\sigma_t^0(\theta_{dry})), \min(\sigma_t^0(\theta_{dry})) + noise\ of\ \sigma_t^0(\theta_{dry}) * 2], \quad 3-6$$

and

$$interval_{wet} = [\max(\sigma_t^0(\theta_{wet})) - noise\ of\ \sigma_t^0(\theta_{wet}) * 2, \max(\sigma_t^0(\theta_{wet}))], \quad 3-7$$

where  $\sigma^0(\theta_{dry})$  and  $\sigma^0(\theta_{wet})$  represent the backscatter measurements normalized to dry and wet crossover angle. The noise and crossover angle was described by Naeimi et al. (Naeimi et al., 2009b).

A number of simplifications needed to be adopted to the ASAR GM algorithm. These included: a) neglecting the seasonal effects on the ASAR GM backscatter, b) assuming linear relationship between the ASAR GM backscatter signal and soil moisture, and c) assuming similar signal to noise ratio of the ASAR GM and the ERS backscatter coefficients in equations (3-4) and (3-5).

The last assumption is unlikely to be true and may cause a systematic bias by removing an insufficient number of ASAR GM acquisitions to retrieve dry and wet references. Nevertheless, systematic biases are easily removable and were demonstrated to have negligible impact on applied studies of soil moisture (section 2.3).

The effect of vegetation on the ASAR WS SSM was demonstrated negligible (Van Doninck et al., 2012). An identical result is expected for ASAR GM given even lower radiometric resolution of the product.

### 3.1.2 The AWRA-L landscape hydrological model

The AWRA-L SSM represents the relative surface soil moisture in upper 5 – 10 cm of soil at the 5 km spatial resolution. The AWRA (Van Dijk & Warren, 2010) consists of a selection of models that estimate all water balance terms associated with the vegetation, soil, groundwater and surface water balance. The models operate at moderate to high resolution across the Australian continent. With a view to assimilate satellite-derived soil moisture observations, a gridded landscape hydrology model (AWRA-L) was built as a sub-model of the AWRA system (Van Dijk, 2010) to explicitly model soil surface moisture dynamics.

The AWRA-L landscape hydrological model estimates the soil water balance at a daily step for four different layers: a) the surface top soil, b) the shallow root zone, c) the deep root zone and d) the saturated ground water store. These are defined by their extractable water storage capacity that depends on the pore size distribution, soil porosity and storage capacity (Van Dijk and Warren 2010). The conceptual differences are that the surface soil layer loses water through direct evaporation; the shallow root zone is accessed by all vegetation; and the deep root zone can be accessed only by deep-rooted (usually perennial) vegetation. The model relies on a number of assumptions summarized by Van Dijk and Warren (Van Dijk & Warren, 2010). The top soil moisture water balance is estimated by:

$$S_0(t + 1) = S_0(t) + I(t) - E_S(t) - D_0(t), \quad 3-8$$

where  $S_0$  is top soil water storage,  $I$  is infiltration,  $E_S$  soil evaporation, and  $D_0$  top soil drainage integrated over a time step (Van Dijk & Marvanek, 2010) (all in mm). The model is based on energy and mass balance equations and uses empirical relationships to estimate the fluxes. The evaporation part of the model is critically important for soil moisture estimates. It accounts for rainfall interception evaporation, soil evaporation and transpiration; the latter two used the Penman-Monteith equation model. AWRA-L parameters were derived from the literature and analysis of streamflow data from several hundred Australian catchments. Full details on the model and its implementation can be found in Van Dijk (2010).

A soil moisture estimate comparable to the relative satellite-derived soil moisture product was calculated as :

$$\theta_M = \frac{S_0}{S_{0,FC}}, \quad 3-9$$

where  $S_{0,FC}$  is the top soil water storage capacity between field capacity and the point at which evaporation ceases (wilting point).  $S_{0,FC}$  was estimated at 30 mm across the continent, corresponding to 0- z cm of the top soil layer, where z ranges between 5 to 10 cm. While this differs from the depth represented by the ASAR GM (< 3 cm), high correlations are expected between the two layers due to their hydraulic coupling. Potentially, portion of the bias removed during the normalization may also be induced by the difference in the depth represented by the ASAR GM and AWRA-L soil moisture products.

The AWRA-L soil moisture is estimated at 0.05° spatial resolution and daily time step. Errors in AWRA-L soil moisture estimates arise from a) the model structure, b) the model parameters, and c) the data used to force the model (Van Dijk & Warren, 2010). The errors in the model structure

are caused by the inevitable simplification of the processes regulating soil moisture dynamics. The errors of the model parameters are dominated by the inability to obtain optimal spatial parameter sets across large areas. The errors in input originate mainly in station measurement and interpolation. Precipitation errors in particular have been shown to strongly affect the agreement in satellite and model soil moisture (Crow et al., 2009; Draper et al., 2009; McCabe et al., 2008).

The AWRA-L SSM daily data over entire continent of Australia were kindly provided from Albert van Dijk and the team at CSIRO, Australia (<http://www.clw.csiro.au/>).

### 3.1.3 AMSR-E dataset

The AMSR-E SSM represents the volumetric soil moisture in upper > 3–5 cm of soil retrieved from radiometer at 0.25 degree spatial resolution. The brightness temperatures measured by the AMSR-E were converted to volumetric soil moisture using the Land parameter Retrieval Model (Owe et al., 2008). The method uses a forward modelling optimization procedure to solve a radiative transfer equation for both soil moisture and vegetation optical depth.

The AMSR-E soil moisture derived from the C-band was used in this study as this is expected to represent soil depth corresponding to that represented by the ASAR GM soil moisture product. The original resolution of the footprint measurements at C-band is 56 km; this was resampled to 0.25 degree grid. Only night-time acquisitions were used as these were better suited for retrieving soil moisture than day-time observations (Jeu et al., 2008).

### 3.1.4 GLDAS-NOAH

The GLDAS-NOAH SSM represents simulations of soil moisture in upper approximately 0 – 10 cm of soil. Since 2000 the Global Land Data Assimilation System (GLDAS) uses the NOAH land surface model to provide soil moisture and other atmospheric and land surface variables at a 3-h time interval. The parameter implemented in this work is the GLDAS-NOAH gravimetric soil moisture measure simulated at 0-10 cm depth. The data are provided at spatial resolution of 0.25° (Rodell et al., 2004). The model is forced by a combination of NOAA/GDAS atmospheric analysis fields, spatially and temporally disaggregated NOAA Climate Prediction Center Merged Analysis of Precipitation (CMAP) fields, and observation-based downward shortwave and longwave radiation fields derived using the method of the Air Force Weather Agency's Agricultural Meteorological system (Rodell et al., 2004).

The lower boundary of each layer is at 0.10, 0.40, 1.00, and 2.00 m, respectively. The Noah model uses the same soil property datasets LPRM (<http://ldas.gsfc.nasa.gov/gldas/GLDASsoils.php>), which is based on the Food and Agriculture Organization (FAO) Soil Map of the World linked to a global database of over 1300 soil samples. Soil moisture and other fields are taken 8 times per day (00:00, 03:00, 06:00, 09:00, 12:00, 15:00, 18:00, and 21:00 UTC). Data generated by the GLDAS-NOAH model are publicly available from <ftp://agdisc.gsfc.nasa.gov/data/s4pa/>.

### 3.1.5 ERA-Interim

The ERA-Interim SSM represents simulations of soil moisture in upper approximately 0–7 cm of soil. The ERA-Interim reanalysis dataset contains atmosphere and surface parameters for the

period from 1989 to present based on the ECMWF Integrated Forecast System (IFS) model (Simmons et al., 2007). The data can be downloaded at low resolution from [http://data-portal.ecmwf.int/data/d/interim\\_daily/](http://data-portal.ecmwf.int/data/d/interim_daily/).

The ERA-Interim reanalysis is produced with a sequential data assimilation scheme. In each cycle, available observations are combined with prior information from a forecast model to estimate the evolving state of the global atmosphere and its underlying surface (Dee et al., 2011). The variety of satellite and ground based measurements assimilated include, among others, clear-sky radiance, rain-affected SSM/I radiances, or recalibrated ERS-1 and ERS-2 surface wind data).

In the IFS, land surface processes are described by the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL) (Viterbo & Beljaars, 1995). In TESSEL, soil processes are calculated in four layers. The lower boundary of each layer is at 0.07, 0.28, 1.0 and 2.68 m depth, respectively. For simplicity, TESSEL uses a globally uniform soil type with fixed soil hydraulic parameters. Saturation is prescribed with a value of  $0.472 \text{ m}^3 \text{ m}^{-3}$ , field capacity with  $0.323 \text{ m}^3 \text{ m}^{-3}$  and the wilting point with  $0.171 \text{ m}^3 \text{ m}^{-3}$ . Soil moisture estimates are provided four times a day at approximately 80 km spatial resolution.

### 3.1.6 OZNET in-situ soil moisture

The Australian monitoring network for soil moisture and micrometeorology (OzNET) offers great resource for the validation of remotely sensed soil moisture data over a variety of land cover types (Figure 3). Each of the 38 stations measures soil moisture at four different depths: 0-5 cm, 0-30, 30-60, and 60-90 cm using the technology of CS616 and CS615 water reflectometers (Merlin et al. 2007). Only the upper most layer is considered in this study. The CS616 reflectometer implements the time-domain measurement method in order to measure the volumetric water content. The variety of soil moisture conditions is demonstrated in Figure 4. The stations were selected as they represent different land cover types within the network.

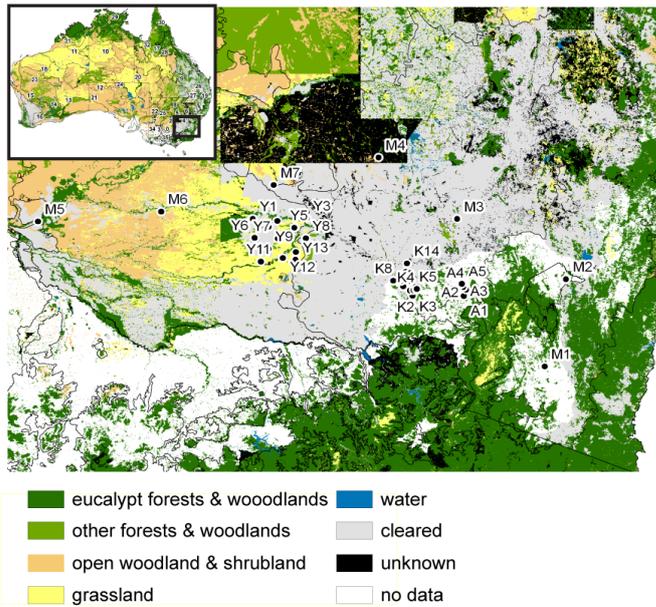


Figure 3. Location of the Oznet stations in south-eastern Australia. The background map displays the present major vegetation groups (Australian Government Department of the Environment and Water Resources, 2005).

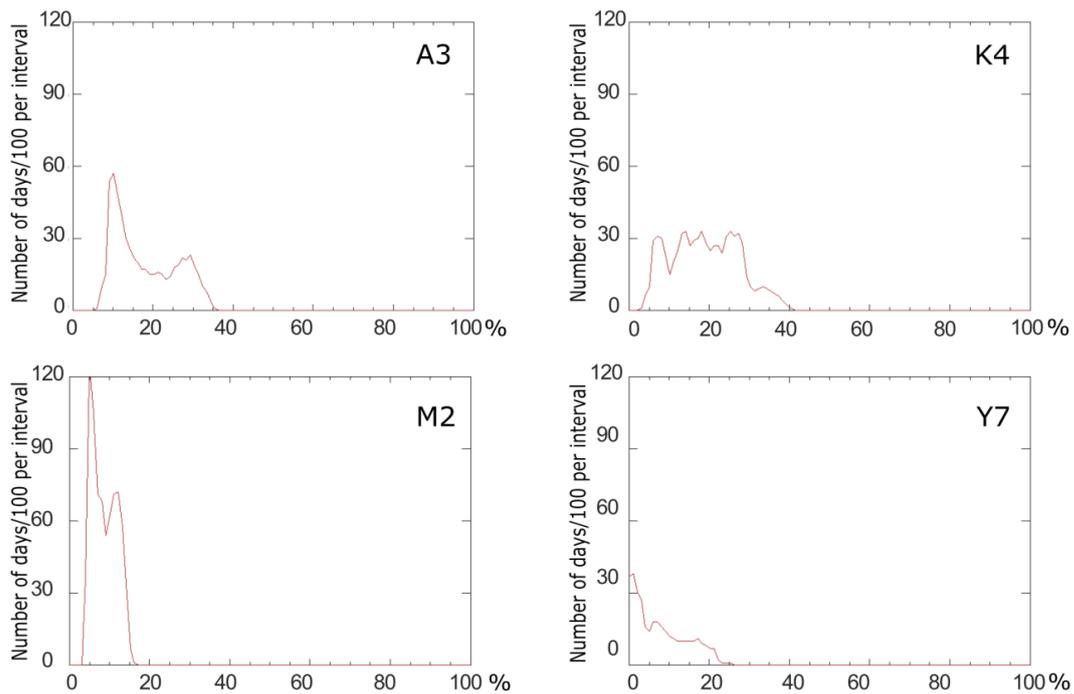


Figure 4. Frequency distribution of soil moisture datasets from four OzNET stations (A3, K4, M2 and Y7). The units are volumetric %.

The six soil moisture datasets and their original units are summarized in Table 4. These were transformed into the dynamics of ASAR GM SSM and later in this work provided in percentage units of saturated soil moisture (see section 4.1.2).

Table 4. List of all applied SSM datasets and their original units prior to dataset transformation.

SSM dataset	Original units	Represented depth	Spatial resolution	Temporal resolution over Australia
ASAR GM	% of saturated soil moisture	> 3 cm	1 km	Approx. every 2-3 day
AWRA-L	% of saturated soil moisture	5 –10 cm	5 km	Daily
AMSR-E	volumetric %	> 3 cm	0.25 degree	Daily
ERA-Interim	volumetric %	Approx. 7 cm	86 km	Daily
GLDAS -Noah	volumetric %	Approx. 10 cm	0.25 degree	Daily
OzNET	volumetric %	0 – 5 cm	point data	Daily

### 3.1.7 Ancillary data on land cover and roughness

Set of ancillary datasets were used to provide an understanding on biogeographic conditions of Australia. These were:

The Interim Biogeographic Regionalisation (IBRA) (Thackway & Creswell, 1995), which was developed by the Australian Government as a planning tool for land conservation (Figure 5). The IBRA classifies Australia into 89 large, geographically distinct, bioregions based on common climate, geology, landform, native vegetation, and species information. Given the definition of the bioregions these are expected to exhibit distinct soil moisture conditions.

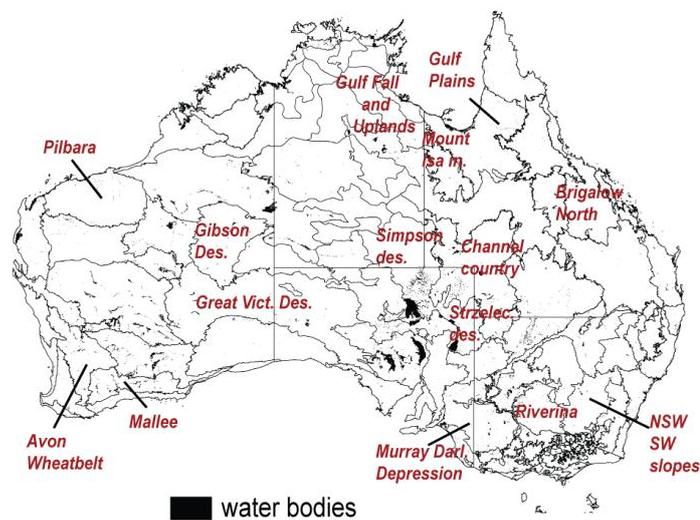


Figure 5. The Interim Biogeographic Regionalization dataset for Australia (IBRA) with selected regions (Thackway & Creswell, 1995).

Furthermore, a map of the Australian soil classification (Isbell, 1996) and a map of the major vegetation groups (Australian Government Department of the Environment and Water Resources, 2005) supported discussions on effects of soil groups and effects of vegetation on soil moisture signal throughout this study (Figure 6, on the left).

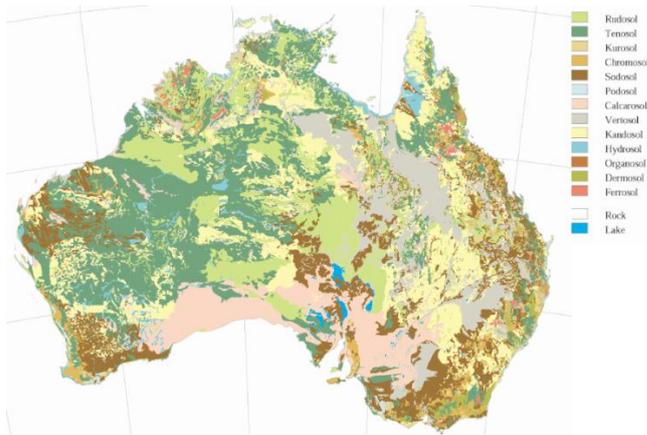


Figure 6. The Australian soil classification (Isbell, 1996) from 1996 (source: CSIRO, Division of Soils).

National land use dataset from the Bureau of Rural Sciences (BRS) represents the land use of Australia from (Smart et al., 2006). It was produced by the Bureau of Rural Sciences (BRS) as a product of the Australian Collaborative Land Use and Management Program (ACLUMP).

Finally, the mean annual precipitation layer, acquired from the Bureau of Meteorology (Figure 6, on the right) was used to support the discussion of the evaluation results.

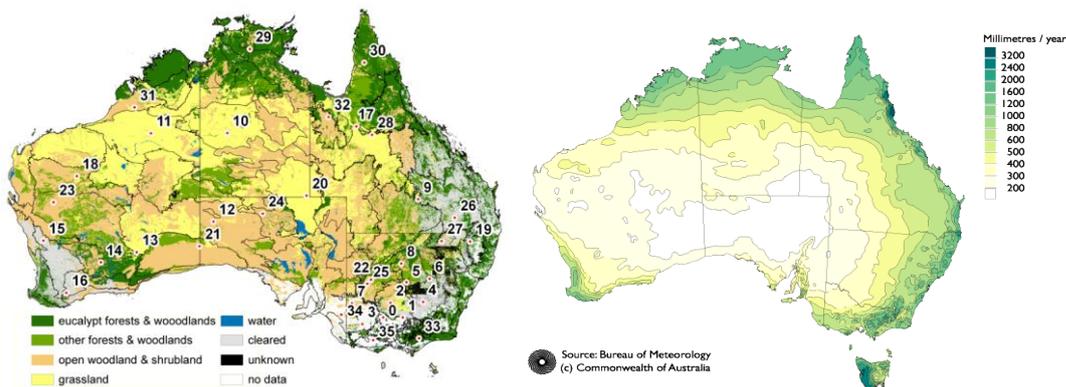


Figure 7. Mean Annual Precipitation for 1900 to 2005 (source: Bureau of Meteorology) (left) and the present major vegetation groups (Australian Government Department of the Environment and Water Resources, 2005), overlaid with the Interim Biogeographic Regionalization dataset for Australia – IBRA (Thackway & Creswell, 1995). Points 0, 12, and 34 refer to regions 1, 2, and 3; these will be referred to in section 4.

### 3.2 Methods

The goal of this thesis was to evaluate the ASAR GM SSM dataset using an exhaustive list of evaluation measures and to answer questions related to the general concept of evaluation of soil moisture dataset. These questions were motivated in the introduction section of this thesis. The steps performed to answer the above goals are summarized below and are divided into a) the pre-processing steps (section 3.2.1) and b) the evaluation steps (section 3.2.2). Section 3.2.1 explains the ASAR GM SSM processing chain applied within the SHARE project over continents of Australia and Africa (Doubkova et al., 2009, 2012). Furthermore, the pre-processing of the five

ancillary SSM datasets (section 3.1), their transformation and sampling is described. Section 3.2.2 provides a concise overview of evaluation measures used to evaluation the ASAR GM SSM dataset. These are introduced in the order as these appear in the text. The exact formulations, assumptions, and limitations of each measure are also listed.

### 3.2.1 Pre-processing

Data from the multiple modes of the side-looking SAR onboard ENVISAT are available since December 2004. The ASAR Global Monitoring Mode (GM) is activated by default when no data from other modes are requested. The ASAR GM 1-km resolution sensor thus offers higher temporal sampling over certain regions when compared to other modes and is suitable for monitoring of dynamical processes such as soil moisture (Pathe et al., 2009b) or inundation (Bartsch et al., 2009).

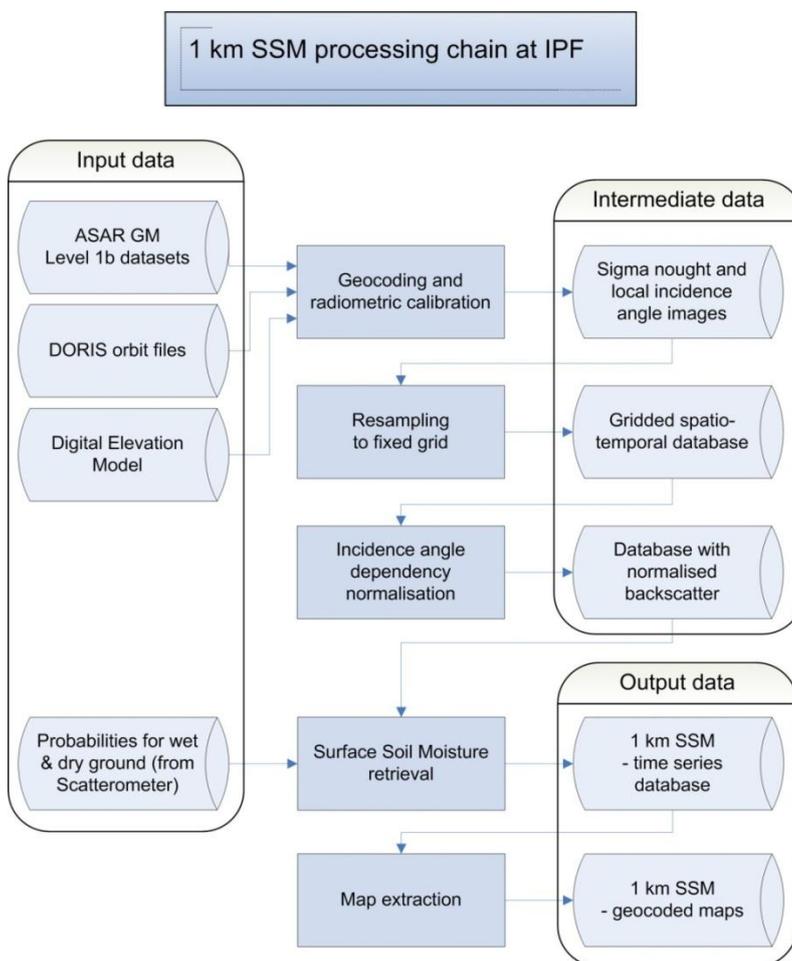


Figure 8. The processing chain of the ASAR GM data at the TU WIEN (Sabel et al. 2010).

The algorithm used to retrieve soil moisture from the ASAR GM observations was derived from the soil moisture algorithm for the Earth Resource Satellite (ERS) scatterometer as in detail described in section 3.1.1. For the soil moisture product generation a processing chain has been setup at the Vienna University of Technology (TU WIEN) (Sabel et al., 2012). The processing consists of several steps including geocoding, radiometric correction, resampling, normalisation

and soil moisture retrieval (Figure 8). For the purpose of this study over 7000 ASAR GM scenes over Australia were processed (Figure 9).

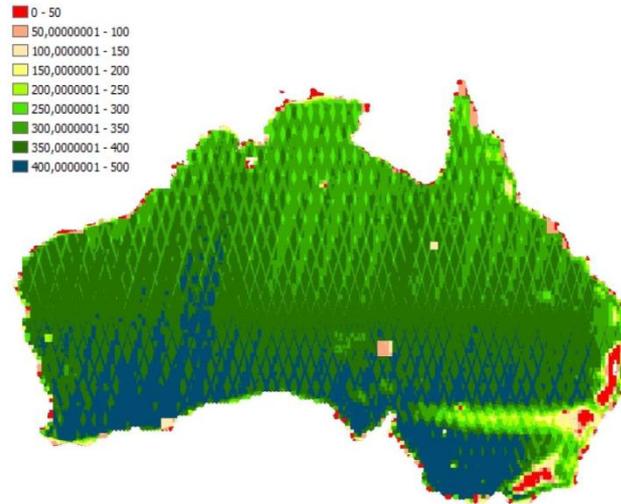


Figure 9. Number of ASAR GM SSM scenes per pixel used in the evaluation studies.

#### Temporal sampling

The ASAR GM SSM and all other SSM products (AWRA-L, GLDAS-NOAH, ERA-Interim, OzNET SSM) were processed over time period during which all observations were available for all data sources – 1<sup>st</sup> of January 2005 until 31<sup>st</sup> of January 2010. The actual number of acquisitions was determined by the dataset with the least frequent temporal sampling – the ASAR GM SSM (Table 4). As such, the number of acquisitions used in the evaluation can be summarized by the Figure 9.

#### Spatial sampling

A spatial aggregation was recommended to reduce the noise of the ASAR GM SSM when used in applied studies (Pathe et al., 2009). The evaluation analyses in this thesis were therefore performed at 5 km scale. To demonstrate the decreasing ASAR GM SSM noise with the coarsening spatial scale, RMSE evaluation measures were performed at both, 1 and 5 km scale. The AMSR-E, ERA-Interim, and GLDAS-NOAH datasets were oversampled to the 0.05 degree grid using the nearest neighbor technique. The ASAR GM data were averaged over the corresponding AWRA-L footprint.

#### Assumption of zero bias

Remotely sensed and modeled SSM data often exhibit bias. This is because these refer to different soil moisture depths, and represent different spatial extents. Furthermore, these often represent different aspects. For instance, the ASAR GM and AWRA-L datasets are expressed in relative units ranging from 0-100% while the AMSR-E dataset is expressed in volumetric units. For these reasons transformation measures needed to be applied prior to data evaluation (section 2.1). Because the goal of this study was to evaluate the quality of the existing ASAR GM SSM product, AMSR-E, and the modelled data were transformed to the dynamics of the ASAR GM SSM (for data assimilation studies, an inverse approach is more logical; Reichle and Koster 2004) using a simple regression equation. The selection of the linear regression is justified in section 4.1.2.

An exception to the latter rule was the TC method for which the transformation between three SSM datasets is needed that is addressed with an iterative regression scheme. Importantly, no unit transformation was performed as the differences in units were accounted for during the linear rescaling.

### 3.2.2 Evaluation of ASAR GM SSM

All below described computations were performed on a continental grid at 5 km spatial resolution. The evaluation was performed over close to 500 000 pixels using a maximum available number of acquisitions per pixel (Figure 9). The large number of per pixel analyzed acquisitions increased the validity of the results and formed an important part of this thesis.

First, absolute and relative standard evaluation measures were computed as introduced in *Standard evaluation measures* section 2.2.1. The measures computed between the ASAR GM and five other SSM datasets are summarized in section 4.2. The emphasis was given on the understanding of spatial distribution of each of the resulting maps as well as on their comparison.

The spatial patterns were assessed using the knowledge of the scattering and emitting behavior of the active and passive systems. Especially, the effects limiting the sensitivity of the ASAR C-band backscatter and AMSR-E C-band temperature brightness to soil moisture were discussed. Also, the effects possible limiting the quality of models were listed. The correspondence of the measures to geophysical parameters was assessed using bioregion maps from the Biogeographic Regionalization dataset for Australia (IBRA) (Thackway & Creswell, 1995) (Figure 5) and the mean annual precipitation (Figure 7).

Second, the ASAR GM SSM was assessed using advanced evaluation methods – the EP and TC *Advanced evaluation measures* methods (introduced in sections 2.2.2 and 2.2.3, respectively). The results are summarized in section 4.3. The EP method was computed according to the equation 3-10. In addition, a method was introduced – predicted RMSE – that assessed the quality of the EP method. The methodologies used for computation of EP, predicted RMSE, and TC methods are introduced below.

The maximum ASAR GM SSM propagated error ( $s_{AS}$ ) at the 1 km spatial scale was estimated *Error propagation* following Pathe's et al. method (Pathe et al., 2009b). The method applies a Gaussian EP scheme (section 2.2.2.1) to propagate the backscatter noise and the standard model parameter errors. In particular, it propagates the noise of the ASAR GM SSM (1.25 dB) and the standard deviation of errors incurred during a) the determination of the dry ( $\sigma_{dry}^0$ ) and wet ( $\sigma_{wet}^0$ ) references, and b) the normalization of the influences of the local incidence angle using the slope ( $\beta$ ). All three parameters are explained below in detail.

The noise is a critical parameter for data users as it defines how well different surfaces can be classified. Noise can be produced by numerous factors including thermal effects, sensor saturation, quantization errors, and transmission errors (Corner et al., 2003). In radar images both, additive and multiplicative noise (speckle) can be present. The multiplicative noise originates from the coherent superposition of spatially random multiple scatterers within the footprint of the radar.

The exact errors are unknown and were estimated based on the understanding of the potential error sources that, importantly, were assumed to be independent. Out of four, by Pathe et al. (Pathe et al., 2009b) listed errors sources, two are expected to impact the ASAR GM SSM acquisitions over Australia. These are namely, the statistical methods used for calculating of the ASAR GM SSM parameters ( $\beta$ ,  $\sigma_{dry}^0$  and  $\sigma_{wet}^0$ ) and the neglecting of seasonal vegetation cover effects, and were both incorporated in the Pathe's EP scheme:

$$s_{AS} = \sqrt{\left(\frac{1.2}{S}\right)^2 + \left(\frac{\beta}{S}\right)^2 + (0.1)^2}. \quad 3-10$$

The algorithm could be easily transformed to the 5 km scale by decreasing the original noise to 0.21 dB (corresponds to the noise of the 5 km product; see section 4.2.1.6). The model parameters  $\Delta\sigma_{dry}^0$  and  $\Delta\sigma_{wet}^0$  were assumed to be 10 % of saturated soil moisture (Pathe et al., 2009b). Assuming that this number reflects, among others, the differences in signal to noise ratio of the ASAR GM and the ERS backscatter coefficients (see discussion about the ASAR GM SSM algorithm in section 3.1.1) it should decrease with decreasing spatial resolution of the ASAR GM SSM product. The resulting number is computed by dividing and rounding the original 10 % of saturated soil moisture by the ratio between the original noise (1.2 dB) and the noise of the product at the 5 km scale (0.21 dB) as in:

$$\Delta\sigma_{dry}^0 = \Delta\sigma_{wet}^0 = \text{Round}\left(0.1 * \frac{0.21}{1.2}\right) = 0.02. \quad 3-11$$

The final standard retrieval error at the 5 km scale then equals to:

$$s_{AS} = \sqrt{\left(\frac{0.21}{S}\right)^2 + \left(\frac{\beta}{S}\right)^2 + (0.02)^2}. \quad 3-12$$

### Triple collocation

The triple collocation method was implemented using ASAR GM, AWRA-L, and a model (either AWRA-L, GLDAS-NOAH, or ERA-Interim). The datasets were rescaled using the iterative linear regression (section 2.1.2.2). First, the significance of the relationship between the three soil moisture datasets was tested that is a prerequisite for the triple collocation method. The one-tailed T test was implemented (section 4.1). Next, the errors  $\tau_{AS}^*$ ,  $\tau_{AM}^*$ , and  $\tau_{AW}^*$  were computed according to the equations 3-13. The ASAR GM SSM data were used as a reference. The selection of the reference was not expected to influence the relative patterns of the residual errors (Dorigo et al., 2010). Finally, the trends in  $\tau_{AS}^*$  related to the choice of the third reference dataset were assessed. For this purpose the triple collocation is repeated replacing AMSR-E with two different globally available model reanalysis dataset (ERA-Interim and GLDAS-NOAH).

$$\tau_{AS}^{*2} = \langle (ASAR^* - AWRAL^*)(ASAR^* - AMSRE^*) \rangle \quad 3-13$$

$$\tau_{AM}^{*2} = \langle (ASAR^* - AWRAL^*)(AWRAL^* - AMSRE^*) \rangle$$

$$\tau_{AW}^{*2} = \langle (ASAR^* - z^*)(AWRAL^* - AMSRE^*) \rangle$$

The method relies on the following fact: if datasets are alike in the way they evolve in time, their triple collocation errors decrease and the probability that these represent the true soil moisture increases. Reversely, if a dataset largely differs from both of the others also its estimated error will be large.

The TC method has rather strict assumptions that were presented as critical for the success of the method (Zwieback et al., 2012). These are listed below and assessed for ASAR GM, AWRA-L, and AMSR-E SSM:

- Selected datasets must represent the same phenomenon (assumption of identity)
- There is no bias between the separate datasets (assumption of zero bias) or this has been removed prior to data analyses
- Errors of the datasets are uncorrelated (assumption of zero crosscorrelation)
- Errors are not correlated in time (assumption of zero autocorrelation)
- There is a sufficient number of triplets (assumption of sufficient triplets)

The errors of the datasets were assumed to be independent given the large differences between the retrieval strategies of the three datasets (section 3.1). The second assumption has been in detail assessed in sections 3.2.1 and 4.1 and is especially important as bias is commonly found between remotely sensed and modeled data. The third assumption of uncorrelated errors is required by both, TC, and predicted RMSE measure and is discussed later in this section. Forth, the selection of the triplets (corresponding acquisitions of all three datasets) was limited by the availability of ASAR GM SSM – the shortest and the less frequent dataset. The ASAR GM SSM has a revisit period over Australia approximately every 3 to 4 days. Given the fast infiltration and evaporation rates over vast portions of Australia the autocorrelation level after three days can be assumed close to zero. Finally, the number of triplets used in the analyses range from 300 (northeastern Australia) to 550 (southeastern Australia) with an exception of < 200 triplets in southeastern Australia. According to Zwieback et al. (Zwieback et al., 2012) this corresponds to the relative error of TC of 13% (300 triplets) to 9.5% (550 triplets).

The predicted RMSE was computed according to the equation 2-26. This measure evaluates other evaluation measures and, hereby, represents a new class of evaluation measures. In the equation 2-26, the parameters  $s_y$  and  $s_x$  stand for the independent errors of ASAR GM SSM and AWRA-L SSM, respectively. The computation is complicated by the limited knowledge of the AWRA-L dataset error ( $s_y$ ) (Van Dijk & Warren, 2010). In a first approximation,  $s_y$  was assumed to be constant and equal to 15% of the soil moisture content at field capacity (30 mm). Given the top soil water storage of 30 mm  $s_y$  of 15% accounts for approximately 4.5 mm. Considering that this corresponds to 0.05–0.10 m of the top soil layer, 4.5 mm corresponds to 0.045–0.09 m<sup>3</sup>/m<sup>3</sup>. This seems as a realistic error estimate for an uncalibrated model (Crawford et al., 2000). The assumption of a constant behaviour of  $s_y$  is unlikely to be accurate, either spatially or temporally, but was necessary due to the lack of spatial estimates other than ASAR GM that could be used for the independent evaluation of the AWRA-L SSM dataset. Where possible the difference between

the RMSE and predicted RMSE is qualitatively assigned to the satellite or to the modelled data in interpretation.

The predicted RMSE measure was applied to assess the ability of the ASAR GM propagated error ( $s_{AS}$ ) to predict the RMSE between the ASAR GM and AWRA-L SSM. A model was used that relates the RMSE to the individual errors of each dataset according to the equation 2-26. The RMSE is calculated from the observations according to the equation 2-6. Given the independence of the two methods, a high correspondence between the RMSE and predicted RMSE would suggest a high quality of the RMSE model and  $s_{AS}$  estimate.

The assumptions of the predicted RMSE and TC methods require the respective errors to be independent. The assumption of errors is realistic as the main input data to the AWRA-L, GLDAS-NOAH, and ERA-Interim datasets (daily precipitation, incoming shortwave radiation and temperature) are independent of the ASAR GM backscatter.

The errors of the two remotely sensed datasets can be also assumed independent given the differences in the algorithms (Owe et al., 2008; Pathe et al., 2009b) and the resulting differences in the sensitivity of the ASAR GM backscatter and the AMSR-E brightness temperature to soil moisture (Parinussa et al., 2011; Pathe et al., 2009b).

#### *Discussion*

Finally, in the discussion section, all results were summarized to answer six scientific questions of the thesis presented in the introduction chapter.

Simple visual interpretations of the error characteristics were supplemented with regression analyses performed between the separate error maps. The divergences between maps were assigned either to random error, systematic error, to a failure in fulfilling conditions of selected evaluation measures, or to an unknown error resulting. Also, for an easier understanding of the evaluation measures, these were divided into three groups according to the quality described: a) random and systematic error analyses, b) data evolution in time, and c) mean bias. One measure per group was selected that well described the entire group and could be used as a standard evaluation approach for any SSM dataset.

## 4. Results

### 4.1 Preprocessing

To avoid systematic bias during data evaluation SSM datasets need to be transformed into a common soil moisture dynamics. Biases exist between different acquisitions of SSM due to different sensing depths and spatial extents represented. This chapter assesses biases and differences in the distributions of four SSM datasets implemented in this study (AWRA-L SSM, ERA-Interim SSM, GLDAS-NOAH SSM, and OzNET SSM). Understanding the biases and different distributions of the data was necessary to select a method to be able for data transformation.

#### 4.1.1 Understanding frequency distribution of SSM datasets

In this chapter the frequency distribution of each of dataset is provided and the likely origin of discrepancies is explained.

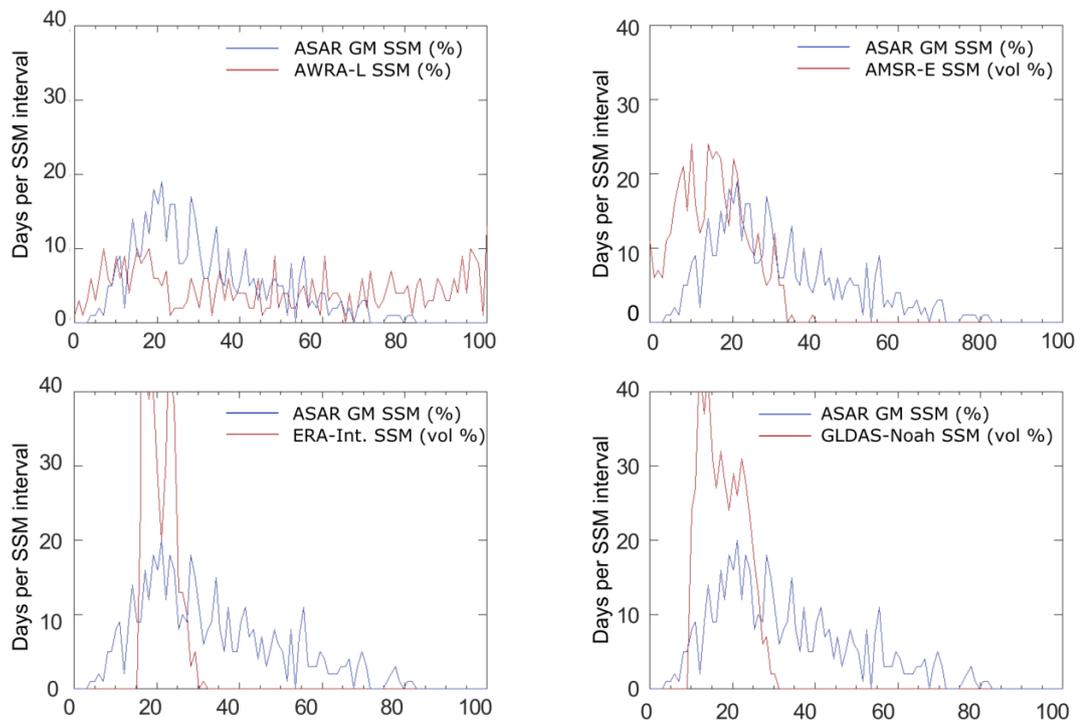


Figure 10. Frequency distribution of ASAR GM SSM plotted along with a) AWRA-L SSM, b) AMSR-E SSM, c) ERA-Interim SSM, and d) GLDAS NOAH SSM over region 1 with a centre coordinate 144 E, 35.9 S (agriculture fields). The data are plotted in the original units (Table 4) representing areas of 5x5 km (AWRA-L and ASAR GM SSM), 25x25 km (AMSR-E, GLDAS), and 86x86 km (ERA-Interim).

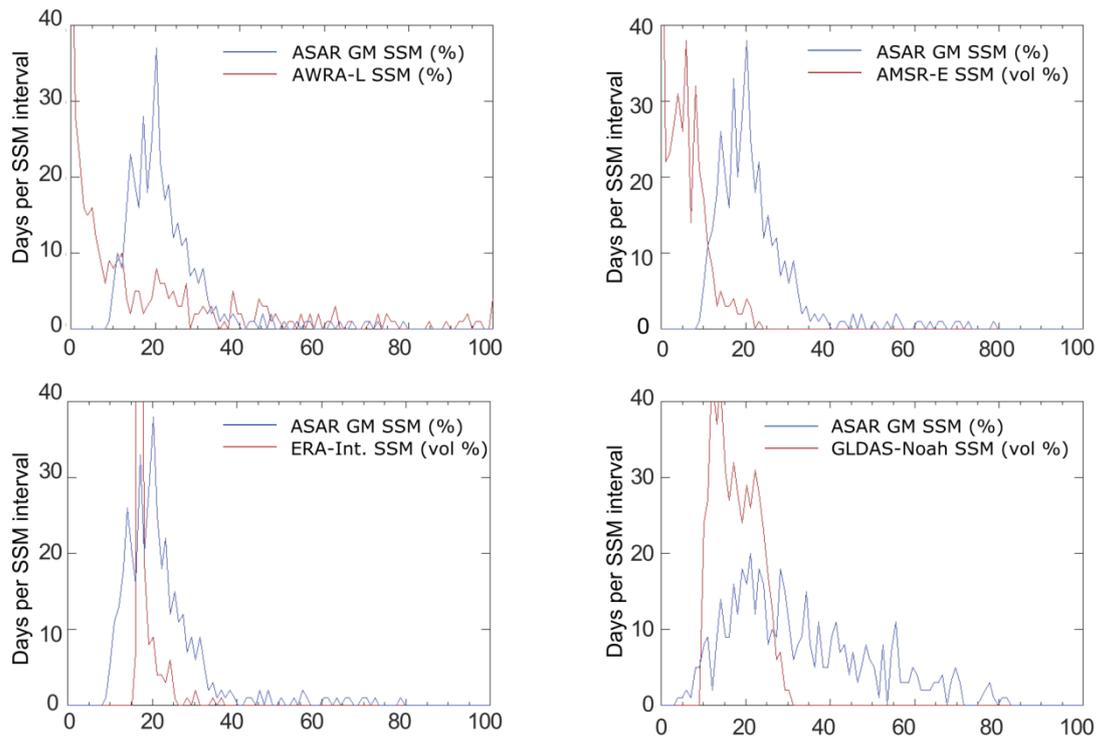


Figure 11. Frequency distribution of ASAR GM SSM plotted along with a) AWRA-L SSM, b) AMSR-E SSM, c) ERA-Interim SSM, and d) GLDAS NOAH SSM over region 2 with a centre coordinate 130.1 E and 28.0 S (shrublands and rare woodlands). The data are plotted in the original units (Table 4) representing areas of 5x5 km (AWRA-L and ASAR GM SSM), 25x25 km (AMSR-E, GLDAS), and 86x86 km (ERA-Interim).

Figure 10 and Figure 11 represent frequency distributions of four SSM datasets along with the ASAR GM SSM (used as a reference dataset and displayed in blue) for two locations, representing wet climate with agriculture fields (region 1) and dry climate with rare woodlands and shrublands (region 2) respectively. The corresponding OzNET stations to region 1 are K4 and A3 (Figure 4). Unfortunately, no corresponding stations are available for region 2. Importantly, the data are plotted in their original units (% for ASAR GM and AWRA-L, volumetric % for ERA-Interim, AMSR-E, and GLDAS-NOAH). Furthermore, the spatial resolutions of the datasets differ (Table 4) which may cause differences in dataset mean and the standard deviation.

#### Assumptions

#### Frequency distribution

The frequency distributions over the two regions differ as they are controlled by different underlying rainfall and climate conditions, as well as vegetation and soil conditions (Settin et al., 2007). Nevertheless, a common characteristic of the majority of displayed frequency distributions is the positive skewness (ASAR GM SSM, ERA-Interim, and GLDAS-NOAH SSM), which reflects log-normal distribution. This corresponds to the expected PDFs over regions where lower boundary soil moisture values are approached (Famiglietti et al., 1999; Western et al., 2002). Given the low mean precipitation conditions over vast portions of Australia (Figure 7) frequency distribution similar to that in Figure 10 and Figure 11 can be expected over the entire continent with an exception of tropical northern and temperate southeastern and southwestern Australia. Positively skewed lognormal distribution was also demonstrated over OzNET stations (Figure 4).

On the contrary, the AWRA-L SSM dataset demonstrate a whole range of distributions, including uniform, exponential (Figure 11), as well as bimodal distribution (Figure 10).

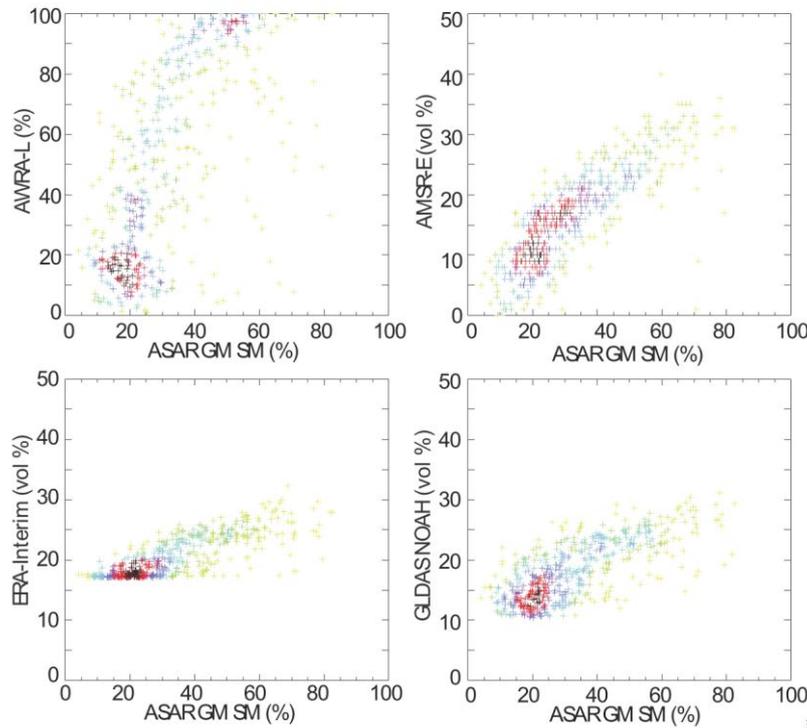


Figure 12. Relationship between the ASAR GM SSM and AWRA SSM (a), ASAR GM and AMSR-E SSM (b), ASAR GM and ERA-Interim SSM (c), and ASAR GM SSM and GLDAS NOAH SSM (d) over region 1 with a centre coordinate 144 E, 35.9 S (agriculture fields). The data are plotted in the original units representing areas of 5x5 km (AWRA-L and ASAR GM SM), 25x25 km (AMSR-E, GLDAS), and 86x86 km (ERA-Interim). The color represents the density of points ranging from high (black) to very low (yellow).

AWRA-L SSM exhibit bimodal distribution over vast portions of Australia with precipitation *Frequency distribution of AWRA-L* exceeding 400 mm (well demonstrated in Figure 10 and Figure 12). In particular, it corresponds with ASAR GM SSM in low soil moisture ranges, but was significantly higher in higher (> 40 %) soil moisture ranges. This is rather surprising as bimodal distribution is usually only encounter over areas with distinguishable seasonal behavior (e.g. dry and wet seasons) which is typical only for the northern tropical regions. Moreover, the second peak occurs at range close to saturation and often exceeds the peak in lower ranges. This is not encountered in any other dataset. The AWRA-L distribution diverge from other datasets, also over very dry regions exhibiting exponential behavior (Figure 11). The latter may be caused by simplifications in the AWRA-L model. For instance, an assumption of identical soil water storage of the AWRA-L model (approximately 30 mm) across the continent may keep soils with a fast infiltration rate saturated for too long and shift the soil moisture values towards higher ranges comparable to other SSM datasets. Another reason for the bimodal behavior may be an underestimation of evapotranspiration or runoff rate during the wet period which would also keep soil wetter than it actually is.

The minimum value of the ERA-Interim dataset never decreases below 17.1 (Figure 12 and Figure *The sources of bias* 11). This was expected and is explained by the scaling of the ERA-Interim data between the

saturation ( $0.323 \text{ m}^3 \text{ m}^{-3}$ ) and the wilting point ( $0.171 \text{ m}^3 \text{ m}^{-3}$ ). Similarly, the GLDAS-NOAH appears to be cut off at the wilting point (Figure 11 and Figure 12). Such differences are to be removed during dataset transformation.

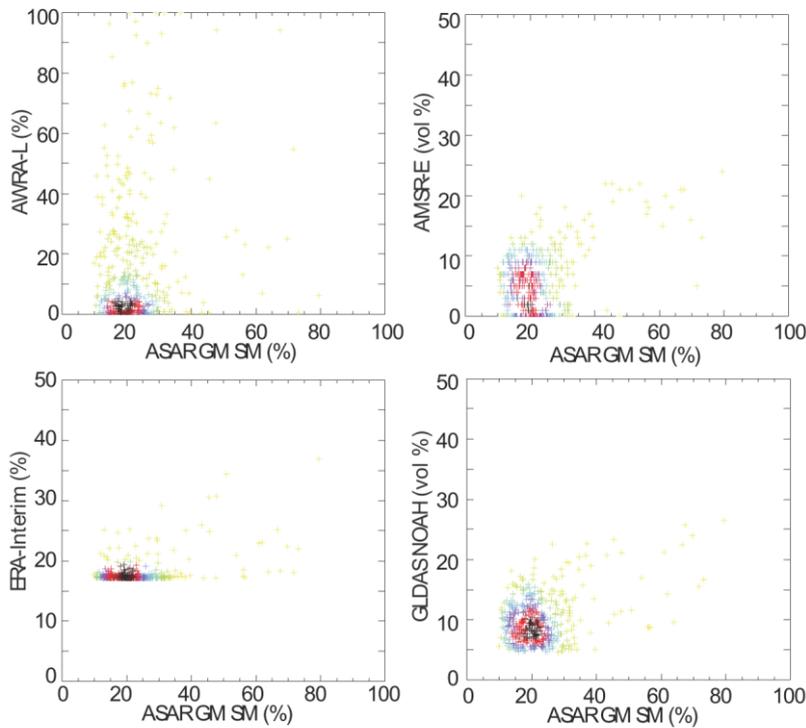


Figure 13. Relationship between the ASAR GM SSM and AWRA SSM (a), ASAR GM and AMSR-E SSM (b), ASAR GM and ERA-Interim SSM (c), and ASAR GM SSM and GLDAS NOAH SSM (d) over region 2 with a centre coordinate 130.1 E and 28.0 S (shrublands and rare woodlands). The data are plotted in the original units representing areas of 5x5 km (AWRA-L and ASAR GM SM), 25x25 km (AMSR-E, GLDAS), and 86x86 km (ERA-Interim). The color represents the density of points ranging from high (black) to very low (yellow).

#### 4.1.2 Data transformation

To allow absolute comparison between SSM datasets the discrepancies described in chapter 4.1.1 needed to be mitigated by transforming the original values to a reference soil moisture dynamics value. In this chapter the most appropriate transformation method was selected.

The transformation technique was selected using the RMSE measure as a criterion. RMSE was computed using two sampling sizes (all triplets and only triplets corresponding to the evening AMSR-E acquisitions) to investigate if the performance of the transformation technique is influenced by the temporal sampling. ASAR GM SSM is selected as a reference given the aim of this study to characterize the errors of this dataset.

It is important to note that no unit transformation was performed in this thesis. The reason is that the transformation from relative to volumetric units (the only units seen in this thesis) has a linear character and could be equally well as part of other transformation measures. As such, the

result of the transformation performs both; the mitigation of the differences due to different units as well as differences in different spatial representations and depths.

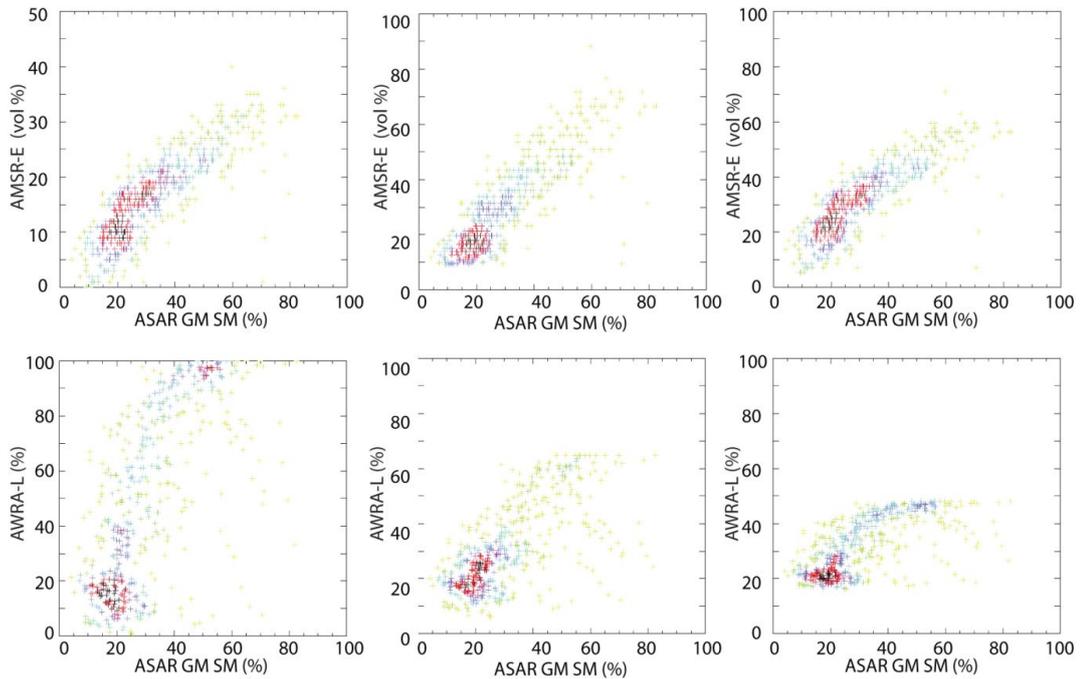


Figure 14. Relationship between ASAR GM SSM and AWRA-L (lower line), and ASAR GM and AMSR-E SSM (upper line). The relationship is displayed prior to data transformation (left), after applying CDF matching (third order) (middle), and after applying linear regression (right) techniques over region 1 with a centre coordinate 144 E, 35.9 S. The color represents the density of points ranging from high (black) to very low (yellow).

The first method, linear rescaling using standard deviation and mean, requires datasets to be normally distributed. The method was rejected because none of the investigated datasets followed normal distribution. This is demonstrated on the example of regions 1 and 2 (Figure 9 to Figure 12).

The second method, CDF matching technique (Drusch et al., 2005), is commonly applied prior to data assimilation. It assigns data distribution of a reference to an analyzed dataset and thus changes the data distribution. Given the complexity of the CDF method it was impossible to analytically describe the actual transformation of the dataset. Furthermore, there is a lack of understanding in the literature about the potential overfitting of the transformed to the reference dataset (especially if higher order fitting is used) that may alter information contained in the dataset. Given the above doubts, and the demonstrated higher RMSE comparable to other transformation methods (Figure 15 and Figure 16 left panel) this method was also rejected.

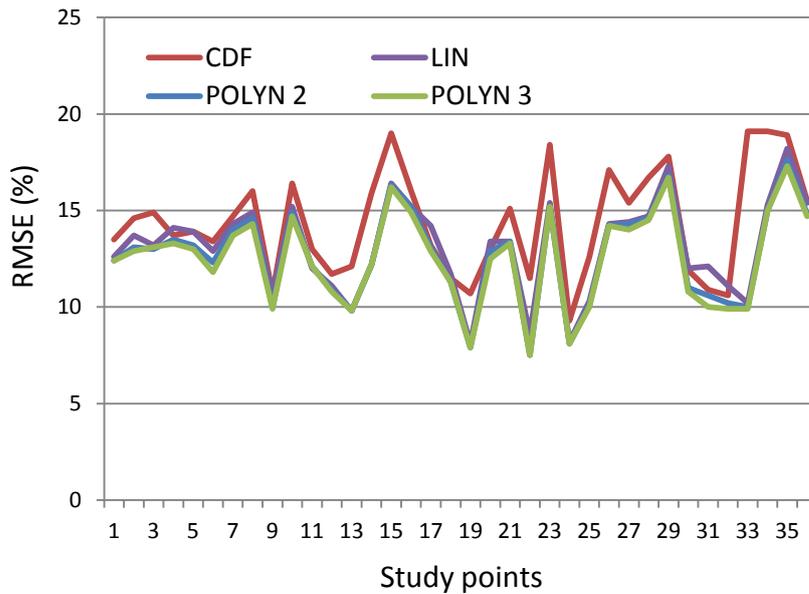


Figure 15. RMSE between ASAR and AWRA-L SSM computed over 36 study points (randomly distributed over Australia, for distribution see Figure 16) on transformed datasets using a) CDF matching (third order), b) simple linear fit, c) second order polynomial fit, and d) third order polynomial fit. The analyses were performed at the 5 km spatial resolution.

The simple regression equation demonstrated the lowest RMSE values (Figure 15) and preserved the original data distribution (Figure 14). Furthermore, using simple regression significantly eases the transformation and potential back-transformation of the datasets in the TC method. Higher order polynomial fit (Figure 15) only minimally (1%) improved RMSE, while this was possibly influenced by the increase in the number of model terms.

As a result, a simple regression equation was chosen for data transformation. To assure that such selection is not influenced by temporal resolution an experiment was run investigating differences in RMSE when all and only night acquisitions were used. The results (Figure 16 right panel) demonstrated that the change in the number of used scenes used for data transformation caused changes to RMSE of > 3%, which was considered negligible for the purpose of this study.

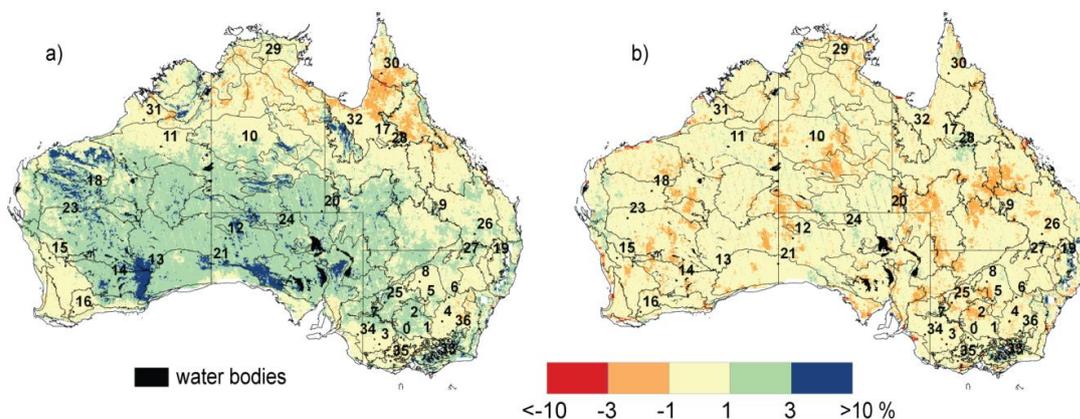


Figure 16.  $\Delta$ RMSE (between ASAR GM and AWRA-L SM) computed as a difference between RMSE using CDF matched datasets minus RMSE using linearly matched datasets (left),  $\Delta$ RMSE (between

*ASAR GM and AWRA-L SM) computed as a difference of RMSE using all images minus RMSE using only night acquisitions (right). The analyses were performed at the 5 km spatial resolution.*

## 4.2 Standard evaluation measures

The standard evaluation measures (as introduced in section 2.2.1) computed between SSM data from the ASAR GM sensor, three separate models, the AMSR-E passive microwave sensor, and in-situ OzNET SSM network are introduced in this section. The models are represented by the AWRA-L landscape hydrological model and the ERA-Interim and GLDAS-NOAH reanalyses. The analyses were performed over the entire continent of Australia on previously transformed datasets (with the exception of untransformed RMSE in section 4.2.1.5). The transformation was performed according to rules presented in section 4.1.2. Important to note is that real absolute values, rather than anomalies, were used in the evaluation.

### 4.2.1 Absolute evaluation measures

The absolute measures used to evaluate ASAR GM SSM include RMSE, nRMSE, MAE, and bias. The results are provided as maps; for each measure several maps (2-5) were produced of the difference between ASAR GM SSS and the selected auxiliary soil moisture datasets. The measures are not displayed for all datasets, given the sometimes very similar results. Results of each evaluation measure are discussed independently as well as compared with other absolute measures. Two additional chapters discuss the important effect of spatial scale on the computation of the absolute evaluation measures (section 4.2.1.6) and demonstrate the misrepresentation that can be caused when computing RMSE on datasets with different soil moisture ranges, depths, and spatial resolutions without performing dataset transformation (section 4.2.1.5).

#### 4.2.1.1 RMSE

The RMSE maps for all gridded data are displayed in Figure 17 while Figure 18 displays RMSEs between ASAR GM and OzNET in-situ stations. The RMSEs computed using coarse resolution products were lower comparable to  $RMSE_{AW}$  computed between ASAR GM SSM and the 5-km AWRA-L product (Figure 17). This can be explained by small geocoding mismatches between the medium resolution datasets. Mismatches can be assumed minimal at coarser resolution scales. Higher  $RMSE_{AW}$  may also be influenced by the data distribution (unique or bimodal) of AWRA-L that significantly differed from other SSM datasets. As such, AWRA-L corresponds with other datasets at very low soil moisture and saturation stage; it however diverges at mid-range values. Another reason for high  $RMSE_{AW}$  values may be the errors of AWRA-L (e.g. errors in precipitation forcing, model structure, the limited parameters available over large areas) that were presented in detail in section 3.1.2. The latter errors, and especially the errors in precipitation forcing, get mitigated at coarser resolutions due to the averaging of in-situ precipitation datasets from a larger number of stations. Furthermore, the differences may also stem from the different represented depth (Table 4) that may not have been removed entirely during the dataset transformation. Lastly, the low error of the ERA-Interim dataset is partly due to the low dynamic range between wilting point and saturation of the ERA-Interim soil moisture dataset (Balsamo et al., 2009).

**Correspondences** The relative RMSE patterns of all maps coincide and demonstrate high errors over areas with high annual precipitation and high soil moisture variation (southeast, southwest and eastern inland areas) and low errors in central and western Australia. The errors estimated over deserts in central Australia (e.g. the Simpson' Desert) remain high due to the infrequent but severe rains that increase the variability of soil moisture. Such RMSE patterns are not surprising given that RMSE is an absolute measure and as such is strongly influenced by the mean and variance of a dataset.

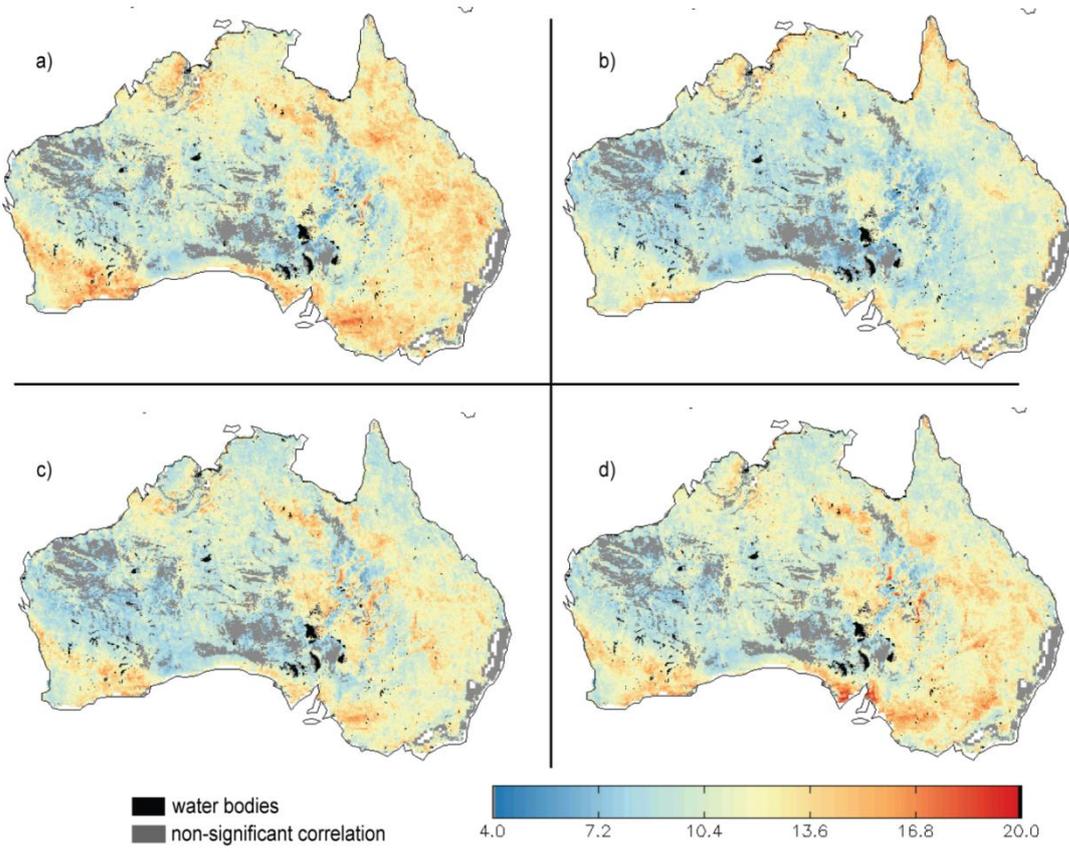


Figure 17. RMSE computed between ASAR GM and a) AWRA-L SSM ( $RMSE_{AW}$ ), b) AMSR-E SSM ( $RMSE_{AM}$ ), c) ERA-Interim SSM ( $RMSE_{AE}$ ), and d) GLDAS-NOAH SSM ( $RMSE_{AO}$ ). The grey areas display the non-significant correlation values ( $p > 0.05$ ). The analyses were performed at the 5 km scale. The units are percentage (%) of saturated soil moisture.

The highs and lows in  $RMSE_{AO}$  and  $RMSE_{AW}$  values are similar (Figure 18).  $RMSE_{AO}$  ranges between 14-18% over stations dominated by croplands (indexed with 'K'), between 10-14% over stations dominated by a mixture of cropland and grassland (indexed with 'Y' and 'A') and is rather random over stations corresponding to urban areas (indexed with 'M').

**Discrepancies** Nevertheless, a detailed look into Figure 17 reveals some differences. In particular, high errors (> 14%) in models ( $RMSE_{AW}$ ,  $RMSE_{AG}$ , and  $RMSE_{AE}$ ) appear over the entire eastern and southwestern inland areas (more than about 200 km from the coast). High values of (> 14%) in  $RMSE_{AM}$  are only evident over the southeast and southwest, keeping the  $RMSE_{AM}$  over eastern Australia rather low (< 10%).

The regions in the southwest and southeast are dominated by dense vegetation (e.g. Mallee bushes and woodlands in the Murray Darling Depression or eastern Mallee). This decreased the sensitivity of the ASAR GM C-band backscatter to soil moisture and was probably the main reason for high RMSEs. The quality of modeled data over these areas remains unexplained.

Vice versa, eastern inland areas are covered with relatively sparse vegetation (grasslands, or cleared areas among sparse woodlands). It allows a good penetration of the ASAR GM backscatter signal and introduces a limited effect of vegetation optical depth on the AMSR-E brightness temperature. The higher values of the RMSE maps computed between ASAR GM SSM and the models may be explained by the different depth of soil moisture of each of the products (Table 4), uncertainties in rainfall forcing, and the simplified manner in which the models simulate evaporation and infiltration parts of the model.

Furthermore, eastern inland areas are dominated by vertisols (Figure 6) that are known to form deep cracks during extended dry periods. The low  $RMSE_{AM}$  over these areas suggests that the ASAR GM backscatter and AMSR-E brightness temperatures detect the process of drying and cracking of soils in a similar manner. On the contrary, the high RMSE values computed between ASAR GM SSM and models portray the simplifying nature of hydrological and landscape models that is not able to capture the cracking characteristics such as: a) drying via evapotranspiration beyond wilting point, b) large infiltration and c) change of porosity during the year (Liu et al., 2010). *Effect of vertisols*

Lastly, high  $RMSE_{AM}$  values are found over coastal northern Australia with mangrove vegetation. These can be explained by the decreasing sensitivity of AMSR-E brightness temperature to soil moisture with increasing open water ratio.

Given the non-existing global error estimates of the three soil moisture simulations (the common parameters used to assess the quality of the ERA-Interim dataset include humidity, wind speed, temperature, precipitation, or water vapour (Dee et al., 2011), only assumptions on their errors and contribution could be provided above. These, however, correspond to the findings recently published by Dorigo et al. (Dorigo et al., 2010) about TC errors of the ASCAT, ERA-Interim, and AMSR-E soil moisture datasets. In particular, the very low values of  $RMSE_{AE}$ , as well as the in eastern inland enhanced errors of  $RMSE_{AE}$ , correspond to these findings.

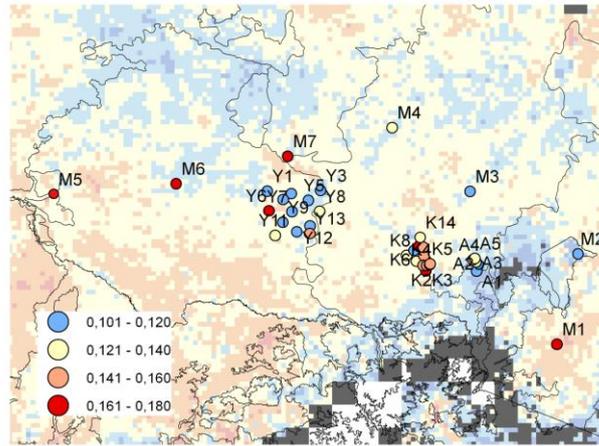


Figure 18. RMSE between ASAR GM SSM and SSM from OzNET in-situ stations (colored dots); at the background RMSE between ASAR GM and AWRA-L SSM. The grey areas in the correlation map display the non-significant correlation values ( $p=0.05\%$ ). The units are percentage (%) of saturated soil moisture.

*Relationship between different RMSEs*

The above discussion suggests that the RMSE maps demonstrate two sources of variations: the first, and much stronger, acts at coarser resolution and reflects the mean and variation of soil moisture dynamics (and corresponds to the frequency distribution); the second acts mainly at medium scale and reflects the soil moisture behavior as impacted by different soil and vegetation types.

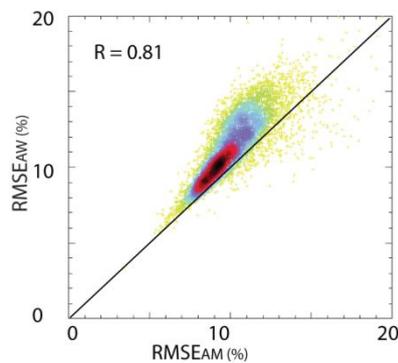


Figure 19. The relationship between  $RMSE_{AW}$  and  $RMSE_{AM}$ . The analyses were performed at 5 km scale.

As mentioned above, all RMSE maps (Figure 17) seem to correspond closely. This correspondence is quantified in Figure 19 using an example of  $RMSE_{AW}$  and  $RMSE_{AM}$ . The high correspondence of RMSEs demonstrated that a larger proportion (64%, originates from Figure 19 and the computed  $R^2=0.64$ ) of its values can be explained by a) the overall mean and variance of soil moisture data and b) medium-scale effects influencing ASAR GM SSM error (land cover and topography). The medium-scale effects did not impact the AWRA-L model nor do they impact AMSR-E coarse resolution SSM data. On the contrary, a smaller percentage of variation (36%) originated in the differences between AWRA-L and AMSR-E. The prior applied matching/rescaling is not expected to increase correspondence in Figure 19 because the linear fit only removes the deterministic

component of the relationship between SSM datasets (slope and intercept) but cannot affect the relative spatial patterns of residuals.

#### 4.2.1.2 nRMSE

The necessity to normalize RMSE statistics arose from the desire to compare performance of RMSE in different areas where the average soil moisture differs.

Not surprisingly, the spatial patterns differ between Figure 20 and Figure 17. The largest difference is illustrated by the relative decrease of the values in eastern and southwestern Australia and the increase in central and southern Australia. This can be explained by the fact that the high RMSEs in areas with high soil moisture mean and variance had lower relative impact on retrieved soil moisture values than the same RMSE values over dry areas. Interestingly however, the values in densely vegetated areas (i.g. Mallee bushes and woodlands in the Murray-Darling Depression or in the eastern Mallee) remain high ( $> 0.6$ ), accentuating the decrease of sensitivity of ASAR GM SSM over dense vegetation.

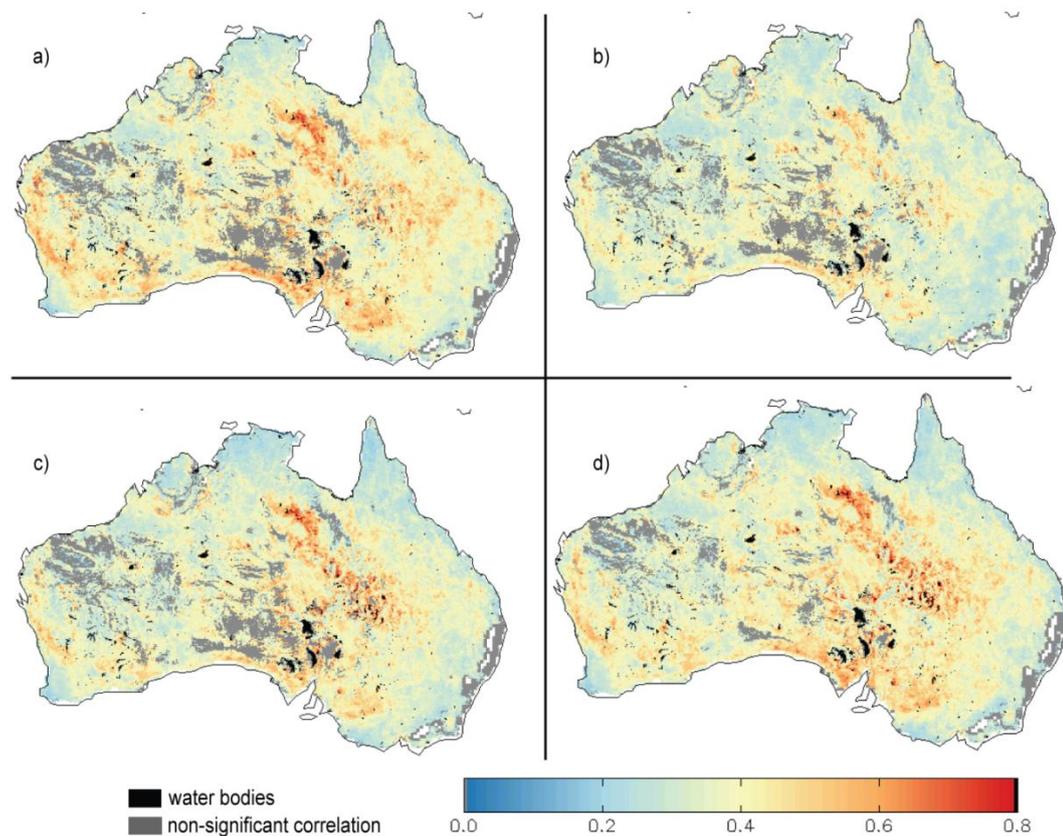


Figure 20. nRMSE computed between ASAR GM and a) AWRA-L SSM ( $RMSE_{AW}$ ), b) AMSR-E SSM ( $RMSE_{AM}$ ), c) ERA-Interim SSM ( $RMSE_{AE}$ ) and d) GLDAS-NOAH SSM ( $RMSE_{AO}$ ). The grey areas display the non-significant correlation values ( $p > 0.05$ ). The analyses were performed at the 5 km scale.

Interesting are the large ( $> 0.8$ ) values of nRMSE in central North (south of Mount Isa Inlier and Channel County bioregions). These may be influenced by the effect of clay cracking soils (Figure 6) that hamper modelling of soil moisture and may also have a limited impact on  $RMSE_{AM}$  since

the algorithm assumes a time-invariant soil porosity (Liu et al., 2010). Likewise, large nRMSE values are evident in the Channel country bioregion that is affected by cracking soils during the dry season and by extensive floods in the wet season; these leave water in the clay pans for several weeks. As expected, such effects were easier detectable by more direct acquisitions of AMSR-E and were difficult to capture in models, without assimilating empirical datasets.

#### 4.2.1.3 MAE

The MAE is an evaluation measure that doesn't quadratically penalize errors of datasets and therefore minimizes the effect of dataset variance. As such, MAE has, in contrast to RMSE, a consistent functional relationship with an absolute error.

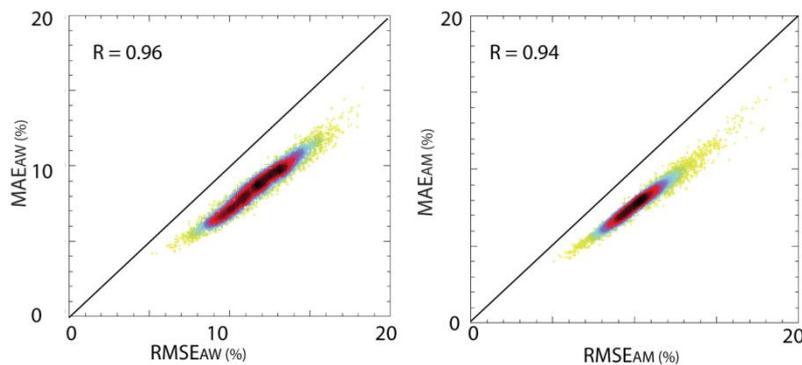


Figure 21. *The relationship between RMSE and MAE for two combinations of datasets – ASAR GM SSM and AWRA SSM (left) and ASAR GM SSM and AMSR-E (right). The analyses were performed at the 5 km scale. The color represents the density of points ranging from high (black) to very low (yellow).*

The higher values of RMSE comparable to MAE in Figure 21 demonstrate the above mentioned effect of quadratically penalized RMSE. Furthermore,  $RMSE_{AW}$  reach considerably larger errors than  $RMSE_{AM}$ , which has been amply demonstrated (Figure 17) and discussed. The difference between MAE and RMSE increase with increasing RMSE values demonstrating the increasing penalization of large residuals with higher dataset variance.

#### 4.2.1.4 Bias

Bias between soil moisture datasets has only minor importance for the majority of soil moisture applications (section 2.3) and was therefore removed prior to analyses performed in this work (section 4.1.2). However, it plays an important role in preliminary dataset evaluations that are commonly performed by a simple visual comparison of absolute values of soil moisture maps. A large and routine bias may cause distrust of the evaluated product. This section demonstrates biases computed between SSM datasets in their original units.

The analyses of ASAR GM SSM bias reveal large differences in the mean values when compared to other studied datasets. These differences varied spatially. The resulting values provide a spatial demonstration of areas where ASAR GM SSM observations are below or above an average bias between pairs of datasets (Figure 22).

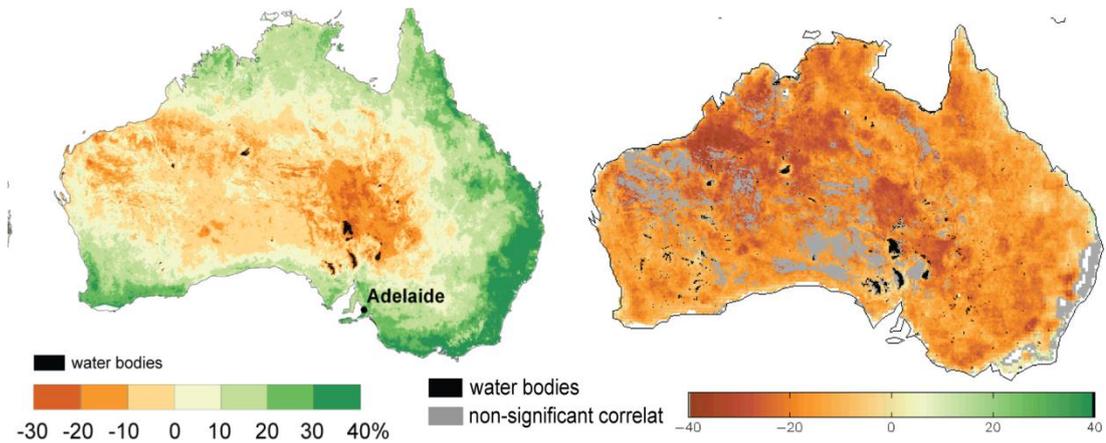


Figure 22. Bias computed between AWRA-L SSM and ASAR GM SSM ( $BIAS_{AW}$ ) (left) and ASAR GM SSM and AMSR-E SSM ( $BIAS_{AM}$ ) (right). Areas with non-significant correlations ( $p>0.05$ ) are grey. The analyses were performed at the 5 km scale. The units are percentage (%) of saturated soil moisture.

These discrepancies can be explained by differences between the physical principles of the satellite retrieval models and the design of AWRA-L hydrological models. The relative ASAR GM soil moisture values are closely linked to the dielectric properties of the surface and rely on a selection of appropriate scaling references to retrieve the saturation level (Pathe et al. 2009). The models simulate the movement of stored water in the top soil as a result of interactions between rainfall, evaporation, infiltration and drainage. The absolute values depend on the weighting of the discrete model components including wilting point and field capacity. AMSR-E applies the Land parameter Retrieval model (Owe et al., 2008) to solve a radiative transfer equation for both soil moisture and vegetation optical depth. The vegetation optical depth and water content within a pixel is known to strongly impact the absolute value of volumetric soil moisture. All three approaches rely on simplifying assumptions and introduce systematic errors (both detection and interpretation). These inevitably alter the absolute values and dynamics of the final soil moisture observations. *Bias due to physical principles*

The spatial patterns of bias are almost identical to the spatial patterns of mean annual precipitation (MAP) (Figure 6). While the AWRA-L soil moisture were considerably higher than ASAR GM over areas with large MAP, the opposite applied over areas with low MAP. Different patterns are evident in  $BIAS_{AM}$ ; these remain negative due to the considerable lower range achieved by AMSR-E data (usually 0 to 0.6). The  $BIAS_{AM}$  is very low over central and western Australia, low over inland areas and high over coastal vegetated areas.

A large portion of positive bias in both maps over arid areas may be explained by the systematic error of interpretation when assuming the ASAR GM dry reference parameter (Pathe et al. 2009). As referred to in section 3.1.1 on pre-processing, the reference retrieval was transformed from the ERS data without further consideration of the pronounced roughness effects at 1 km scale and the large noise of the ASAR GM data. This could have caused an underestimation of the dry reference and thereby an overestimation of average soil moisture values. *ASAR GM references*

On the contrary, the high AWRA-L simulations over the eastern coast may be explained by the poor knowledge of the soil water evaporation parameter in the AWRA-L model. Lastly, the highs

*Bias due to  
AWRA-L and  
AMSR-E*

in AMSR-E may originate from the high vegetation optical depth that increases emission from vegetation while decreasing emission from soil by absorption or scattering.

#### 4.2.1.5 RMSE without dataset transformation

This and the preceding chapters are closely related in that both reflect on dataset bias. Importantly, the original units of the datasets were used (AWRA-L in % and AMSR-E in volumetric units). The theoretical range remains between 0 and 1.

Figure 23 demonstrates RMSE measure computed on SSM datasets without dataset transformation. The spatial patterns differ substantially from those in Figure 17 and correspond to those in Figure 22. This is not surprising given that RMSE computed on original data, before transformation, reflect the absolute dataset error as well as bias between datasets. Bias reflects detection and interpretation errors, especially where the interpretation error (how the measurements are transformed to soil moisture variables) seems to differ substantially between the three datasets. In particular, all three datasets should range between 0 and 1, but in practice, AMSR-E does not exceed 0.6.

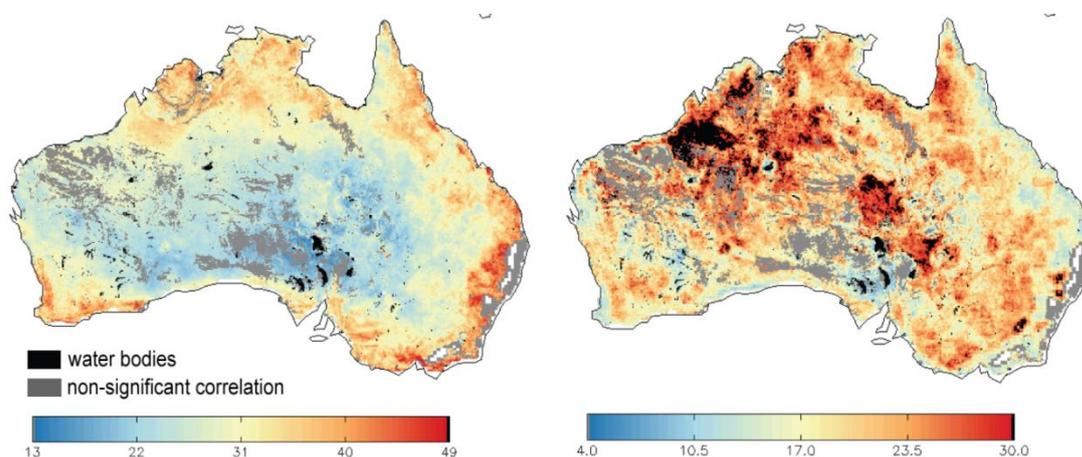


Figure 23. The RMSE computed between ASAR GM and a) AWRA-L SSM ( $RMSE_{AW}$ ) and b) AMSR-E SSM ( $RMSE_{AM}$ ). The grey areas have non-significant correlation values ( $p > 0.05$ ). The analyses were performed at the 5 km scale. The units are percentage (%) of saturated soil moisture.

Figure 17 and Figure 23 demonstrate that RMSE is spatially variable and reflects variation in absolute error as well as in mean and variance. Furthermore, Figure 23 illustrates that without dataset transformation RMSEs computed between different datasets cannot be compared even if they share identical units.

#### 4.2.1.6 Effect of spatial scale on absolute evaluation measures

The standard evaluation methods were computed at the 5 km spatial scale. Noteworthy, however, are the differences between RMSE computed at 1 and 5 km scale. These are essential for understanding section 4.3 which incorporates analyses at two spatial scales.

*RMSE at 1 and  
5 km scale*

The RMSE computed at 1 and 5 km spatial scale are shown in Figure 24 (note the dynamic legend). It should be noted that the datasets were aggregated prior to data transformation. A

decrease of RMSE values with decreasing spatial resolution is evident. This is an expected result given the effects of spatial averaging; namely a) the improvement in the radiometric resolution of ASAR GM SSM with spatial averaging and b) the minimizing effect of spatial mismatch and increasing probability that the products react to identical atmospheric forcing.

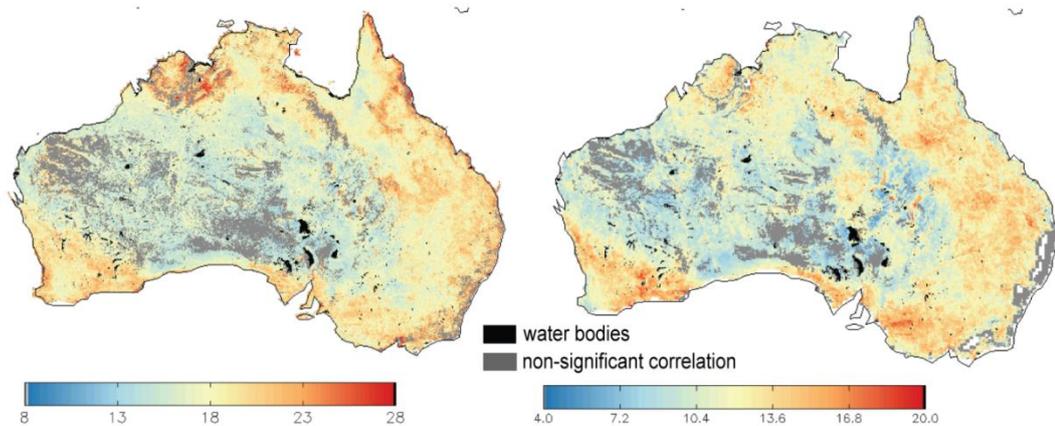


Figure 24. RMSE computed between ASAR GM and AWRA-L SSM ( $RMSE_{AW}$ ) at 1 km (left) and 5 km (right) spatial scale. The areas with non-significant correlation ( $p>0.05$ ) are grey. The analyses were performed after data transformation to ASAR GM dynamics. The units are percentage (%) of saturated soil moisture.

Noteworthy, however, is that the decrease in RMSE is spatially variable (Figure 25, left) and appears more evident over areas with high values of 1km  $RMSE_{AW}$  (e.g. Central Kimberley, Pilbara, or Channel country bioregions). Large decrease of RMSE is also evident over Mount Isa and Channel country bioregions. On the contrary, small decreases of RMSE with decreasing spatial resolution seem to dominate over areas with low values of 1 km  $RMSE_{AW}$ .

The result suggests that RMSE computed at 1 km scale is strongly affected by the spatial patterns *Reasons for differences* introduced by medium-scale (1 km) features (land cover, topography) that impact the error of ASAR GM SSM. Atmospheric forcing acts at larger scales (> 25 km) and impact both the error of ASAR GM SSM and AWRA-L SSM. This forcing seems to appear only as a secondary effect in the RMSE maps. A portion of the AWRA-L SSM error are expected to act directly at AWRA-L resolution (5 km) addressing different land cover and soil types (e.g. poor estimates over rock outcrops, salt lakes or clay soils). Nevertheless, its effect appears small in the 1 km RMSE map, probably due to its much smaller magnitude comparable to the large noise of the ASAR GM SSM.

At the 5 km scale, the effect of small scale patterns is averaged out and the noise of the ASAR GM *Radiometric noise* SSM decreases. As a result, the patterns change; this especially applies at scales > 5 km. A good demonstration of such change is the Pilbara region, where RMSE decreases on average from 25% to about 15%. Here, the soil moisture mean and variance remain very low, and most of the RMSE variation is attributed to the noise of ASAR GM SSM. When decreasing resolution to 5 km, the expected decrease of noise is large and can be described as:

$$\rho_{5km} = \frac{1}{\sqrt{N}} \rho_{1km}, \quad 4-1$$

where  $\rho$  stands for the noise and  $N$  is the number of independent measurements used in the aggregation. For a footprint with a 5 km diameter approximately 144 backscatter measurements with a sampling distance of 0.417 m were used. Assuming that only every fourth is independent (only non-overlapping measurements) the resulting  $N$  is 36.

The radiometric noise can be transformed to radiometric accuracy according to:

$$\rho_{0.5km} = 10 * \log_{10}(1 + \sqrt{ENL^{-1}}), \quad 4-2$$

where ENL refers to the equivalent number of looks. Given the ENL of the ASAR GM SSM product at 1 km estimated to be 9 the resulting accuracy is equal to 1.2 dB. Using equation 4-1 the accuracy of the 5 km product is estimated to decrease to 0.21 dB. This result explains the large drops in the RMSE maps over dry and rocky areas. Here, the magnitude of backscatter measurements during the year often does not exceed 4 dB and the low accuracy of the original product strongly impacted the soil moisture variations.

*Errors of  
AWRA-L*

Furthermore, changes in RMSE patterns are evident also over areas in northeastern Australia that receive sufficient amount of precipitation. The possible explanation is the increasing effect of AWRA-L errors acting at approximately 5 km (e.g. poor estimates over rock outcrops, salt lakes or clay soils) that becomes more pronounced in the RMSE computed at corresponding scale.

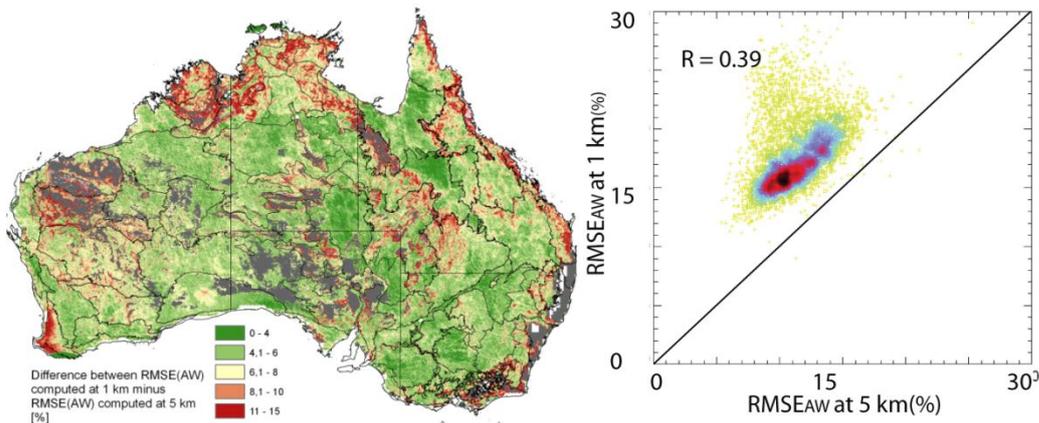


Figure 25. The difference between  $RMSE_{AW}$  computed at 1 km and 5 km spatial scales and the scatterplot demonstrating the relationship between  $RMSE$  computed at 1 and 5 km spatial scales. The units are percentage (%) of saturated soil moisture. The color of the right plot represents the density of points ranging from high (black) to very low (yellow).

Similar analyses were computed also on non-transformed datasets. Interestingly, the results *RMSE on non-transformed datasets* present identical patterns at both 1 and 5 km spatial scales (Figure 26). This may be explained by the fact that the bias exceeds the effect of random error and doesn't change with changing scale.

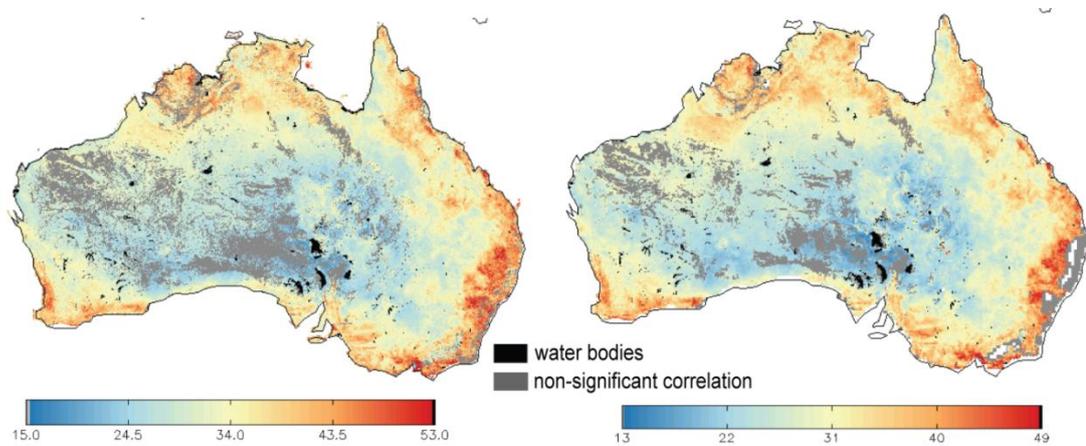


Figure 26. *RMSE computed between ASAR GM and AWRA-L SSM ( $RMSE_{AW}$ ) at 1 km (left) and 5 km (right) spatial scale. The grey areas had non-significant correlation values ( $p > 0.05$ ). The units are percentage (%) of saturated soil moisture.*

## 4.2.2 Relative evaluation measures

### 4.2.2.1 Pearson Correlation

An overall high agreement between the ASAR GM and four other SSM datasets is demonstrated in Figure 27. Significant correlations were found over 88% of the continent for AWRA-L, 95% for AMSR-E, 92% for GLDAS-NOAH, and 95% for the ERA-Interim model. High correlation values ( $R > 0.6$ ) dominate in southwest, southeast and northern Australia and coincides with areas with high RMSE values in Figure 17. The areas have in common high mean annual precipitation (Figure 6, right) and vegetation dominated by herbaceous plants (Figure 6, left).

The good correlation results from a combined effect of a) high seasonality increasing the variance of soil moisture and b) physics of radiation transfer. In particular, the sparse vegetation allows for a good penetration of C-band signals and increases the ASAR GM sensitivity to soil moisture. Likewise, the AMSR-E soil moisture sensitivity over these regions improves through the minor effect of the vegetation absorption and scattering. Australia's wetter regions generally also have a greater density of precipitation gauging stations, which will also enhance the quality of the rainfall forcing of the models and reduce error in soil moisture estimates. *Patterns due to medium scale processes*

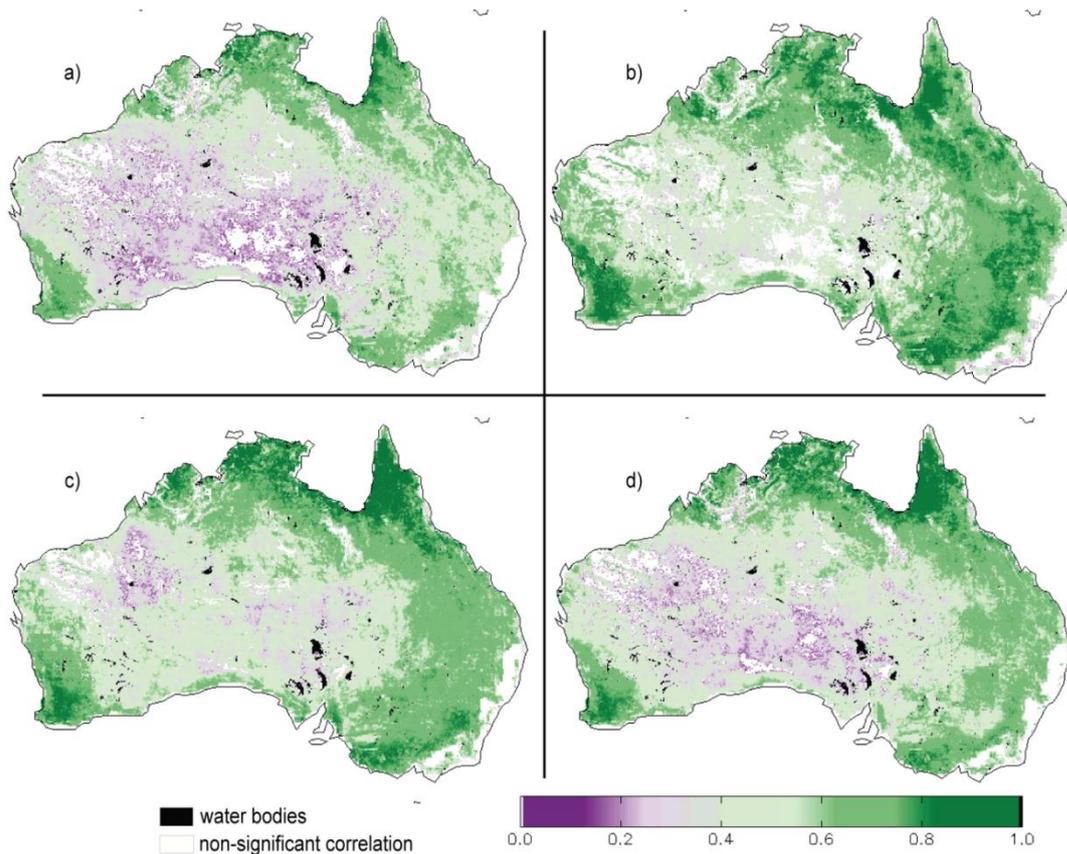


Figure 27. The Pearson's correlation coefficient between ASAR GM and a) AWRA-L SSM, b) AMSR-E SSM, c) ERA-Interim SSM and d) GLDAS-NOAH SSM. The white areas in the correlation map have non-significant correlation values ( $p > 0.05$ ). The analyses were performed at the 5 km scale.

The correlations are generally higher for AMSR-E and ERA-Interim and GLDAS models than for AWRA-L SSM. The maps appear linearly and inversely related to magnitudes of RMSE found in Figure 17. Furthermore, high correlations ( $R > 0.6$ ) dominate all of eastern Australia for the three coarse resolution products, including the Brigalow Belt of acacia woodland.

*Patterns due to large scale processes*

Insignificant correlations were found over portions of central arid, north-western and eastern coastal Australia and correspond to low values (< 10 %) of RMSE (c). The potential reasons for low agreement in dry areas are the low mean and variance of mean annual precipitation, causing low variance in soil moisture and the lower quality of the rainfall data and hence model estimates. The signal-to-noise ratio over arid regions is expected to be minimal considering the low mean and variance of soil moisture data and the poor radiometric resolution of the ASAR GM (~ 1.2 dB) data. Also, the density of in-situ precipitation inputs to models over these areas is minimal. The low correlation in eastern coastal areas may be explained by the limited ASAR GM and AMSR-E sensitivity to soil moisture due to dense vegetation, heterogeneous relief and widespread urban development.

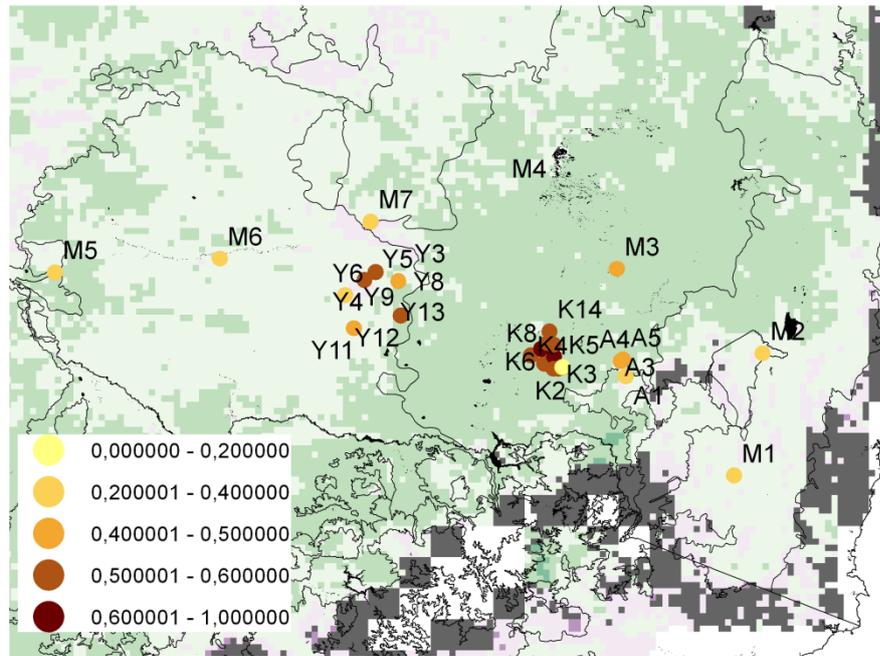


Figure 28. The Pearson's correlation coefficient between ASAR GM and OzNET in-situ stations; the background is Pearson's correlation coefficient between ASAR GM and AWRA-L SSM. The grey areas in the correlation map have non-significant correlation values ( $p > 0.05$ ) (scale as in Figure 27).

The lows and highs in the correlation coefficient for spatial data correspond to the lows and highs of correlation coefficients computed between ASAR GM and OzNET in-situ stations (Figure 28).

#### 4.2.2.2 Spearman correlation

Pearson's correlation coefficient computed on absolute values may be artificially enhanced by the effect of seasonality. This is given by the mathematical formulation of  $R$  that incorporates covariance and standard deviations, which increase with increasing magnitude of the absolute values. The Spearman correlation coefficient (Figure 29), given that it's computed on ranked datasets, doesn't relate to the actual values of soil moisture. As such, it reflects the quality of the data in an ordered sequence and mitigates (not eliminates!) the effect of seasonality. The results can thus be compared to the method introduced by (Dorigo et al., 2010; Scipal et al., 2008b) that calculates  $R$  on anomalies. Given the identical results achieved with all models (namely ERA-Interim, GLDAS-Noah and AWRA-L), the results are only demonstrated on an example of the AWRA-L model and AMSR-E passive microwave dataset.

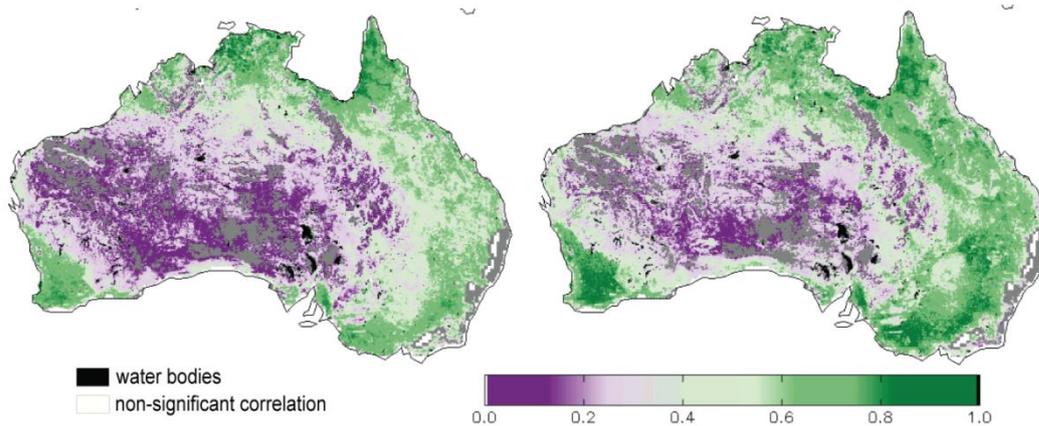


Figure 29. Spearman's correlation coefficient  $R_s$  between ASAR GM and a) AWRA-L SSM and, b) AMSR-E SSM. The grey areas in the correlation map have non-significant correlation values ( $p=0.05\%$ ). The analyses were performed at the 5 km scale.

For both maps the relative patterns correspond demonstrating higher values of  $R_s$  in forested and agricultural areas and higher values of  $R_s$  over central and western Australia. However, of special interest is the comparison of  $R_s$  and  $R$ .

## 4.3 Advanced evaluation methods

### 4.3.1 Error propagation

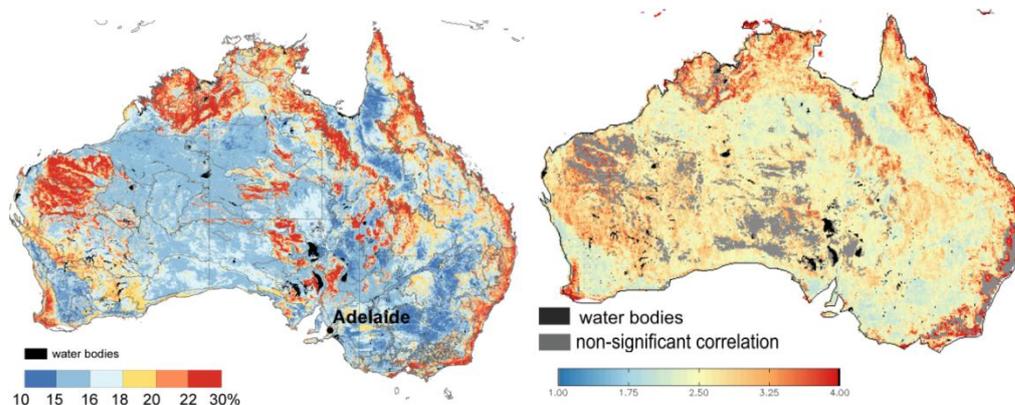


Figure 30. Maximum ASAR GM SSM retrieval error ( $\pi_{AS}$ ) for Australia calculated using the EP model (Pathe et al. 2009) at the 1 km (left) and at the 5 km (right) spatial scale. The units are percentage (%) of saturated soil moisture.

The maximum ASAR GM SSM propagated error ( $s_{AS}$ ) computed at the 1 and at the 5 km spatial scale (Figure 30) strongly coincides with the spatial patterns of a combination of vegetation types (Figure 6) and landscape geomorphology (Van Dijk & Warren, 2010). In particular, the error is less ( $< 18\%$ ) for herbaceous and shrub vegetation classes and greater for forested areas and areas covered with rock outcrops in western, northern, and eastern coastal Australia. The impact of vegetation and geomorphological patterns on  $s_{AS}$  is further emphasized by the correspondence of  $s_{AS}$  with the bioregions of the Interim Biogeographic Regionalization (IBRA; Thackway and

Creswell 1995) evident in Figure 5. The IBRA mapping combines attributes of climate, geomorphology, landform, and lithology.

### 4.3.2 Predicted RMSE

#### 4.3.2.1 RMSE model performance

$RMSE_{AW}$  was predicted according to the equation 2-26 at the 1 km spatial scale. The exact meaning and derivation of the predicted RMSE measure was in detail described in sections 2.2.3.1 and 3.2.2. It is here displayed along with observed  $RMSE_{AW}$  (Figure 31) to allow a direct method's evaluation. A high agreement of the spatial patterns of predicted  $RMSE_{AW}$  and  $RMSE_{AM}$  is evident. Also, the patterns coincide with the spatial distribution of  $s_{AS}$ . This was expected since  $s_{AS}$  comprise the major input into predicted  $RMSE_{AW}$ . The patterns seem to originate from processes at two spatial scales, medium (1-5 km) and a coarser (>25km) scale.

The areas with high values (> 30%) coincide in both maps and cover regions associated with steep slopes and rock outcrops (e.g. rock outcrops in northern and Western Australia). These values act at medium scale and can be attributed to ASAR GM observational errors that originate in foreshortening effects in steep slopes or in diffuse scattering from very rough areas. Such effects are not always corrected during geometric and radiometric correction due to the limitations of the DEM. *Comparison of observed and predicted RMSE*

Errors above 30% are also encountered along the eastern coast and might be associated with dense vegetation that lowers sensitivity of the C-band backscatter to soil moisture. Similar findings demonstrating the sensitivity of  $RMSE_{AW}$  to the topographical and geomorphological medium scale features were amply documented (Van Dijk & Warren, 2010).

The medium values (20 to 24%) coincide in both maps and often correspond to areas with high  $R$  (> 0.6) demonstrated in Figure 27. These are alluvial, topographically uniform areas, or areas covered with herbaceous growth (e.g. alluvial region in the Gulf Plains in northern Queensland or the Nullarbor bioregion in southern Australia that exhibit relatively high mean annual precipitation. Furthermore, medium values dominate in central arid regions, act at considerable coarser scale, and correspond to areas with only limited mean annual precipitation (Figure 5). These areas correspond to areas with low  $RMSE_{AW}$  values (< 10%) (Figure 17).

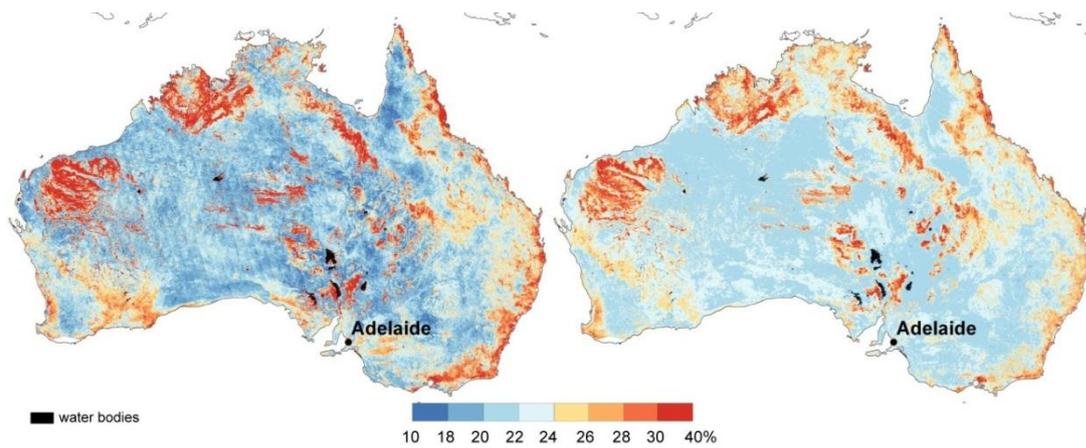


Figure 31. The maps represent  $RMSE_{AW}$  computed from the observations (left) and  $RMSE_{AW}$  predicted (right) at the 1 km scale. The units are percentage (%) of saturated soil moisture.

The match between both predicted  $RMSE_{AW}$  and observed  $RMSE_{AW}$  maps computed at the 1 km spatial resolution with growth forms is pronounced in southwestern and southeastern Australia. The crossover between herbaceous growing forms and shrubs is especially evident. An example is Menzies Line in southwest Australia dividing herbaceous vegetation on cleared land from native shrubland; there are sharp divisions between cropping and grazing land east of Adelaide, and the Cobar bioregion in eastern Australia that is dominated by small trees (Figure 7). These regions are easily detectable due to their specific land cover forms and also due to the specific soil and explicit bedrock type (Van Dijk and Warren 2010).

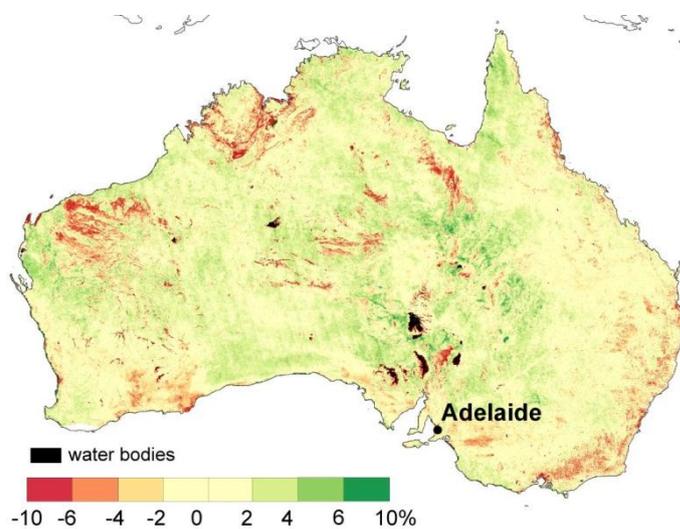


Figure 32. The difference between  $eRMSE_{AW}$  and  $RMSE_{AW}$  computed from the observations.

The predicted  $RMSE_{AW}$  and observed  $RMSE_{AW}$  maps correspond within  $\pm 4\%$  of saturated soil moisture over 89% of the land mass (Figure 32). The remaining 11% coincides mainly with rock outcrops, salt pans, and densely vegetated areas (Figure 7).

The predicted  $RMSE_{AW}$  underestimates the observed  $RMSE_{AW}$  over areas with steep slopes and rock outcrop areas in central, western, and northwestern Australia (red colors in Figure 32). This

underestimation may originate from ASAR GM (Figure 2) as well as from the AWRA-L error estimates. The AWRA-L soil moisture estimates are likely to be poor where the surface is dominated by hard rock outcrops or salt lakes, as the model parameterization does not explicitly consider these features. Nevertheless, the AWRA-L errors are expected to be mainly related to the errors in rainfall forcing (Van Dijk and Warren 2010), and thus correspond to relatively large scale patterns. The predicted  $RMSE_{AW}$  is also lower than observed RMSE in eastern coastal Australia. The reverse performance is found over large portions of central and western Australia (green colors in Figure 32). Given the limited mean annual precipitation (Figure 3) over these regions it is suggested that the error estimate of the AWRA-L model may be lower than the anticipated 15% ( $0.045 - 0.09 \text{ m}^3/\text{m}^3$ ).

The results demonstrate a high agreement between observed  $RMSE_{AW}$  and predicted  $RMSE_{AW}$ . Given the independence of the two methods their correspondence suggests good accuracy of the predicted RMSE and the observed ASAR GM error estimate  $s_{AS}$  at the 1 km spatial scale.

Importantly, the results demonstrated that the main source of spatial variation of observed  $RMSE_{AW}$  at the 1 km scale originates from the ASAR GM SSM data. This was expected given the effect of small scale (< 1km) patterns impacting the ASAR GM SSM product and the dominating effect of medium (> 5km) and large scale (> 25km) effects of atmospheric forcing impacting the AWRA-L dataset. As noted in section 4.2.1.6, the spatio-temporal behavior of absolute evaluation measures may change with the spatial scale. For instance, it can be expected that at the 5 km spatial scale, the medium resolution patterns impacting the ASAR GM SSM average out and the coarser resolution atmospheric processes become dominant. The change of predicted RMSE from the 1 to 5 km spatial scale is addressed in the following section.

#### *4.3.2.2 Performance of predicted RMSE at coarser spatial scales*

The results from section 4.3.2.1 were published as a separate article (Doubková et al., 2012) and suggested that the main source of predicted RMSE at the 1 km scale originates in the ASAR GM SSM dataset. Only later, the significant impact of the spatial scale on the  $RMSE_{AW}$  measure was revealed (section 4.2.1.6). This section investigates if the ASAR GM SSM error ( $s_{AS}$ ) remains the main contributor to predicted RMSE also at coarser spatial scale. In particular, it assesses the correspondence of the ASAR GM SSM error ( $s_{AS}$ ) to the observed  $RMSE_{AW}$  at 1 and at 5 km spatial scale (Figure 33).

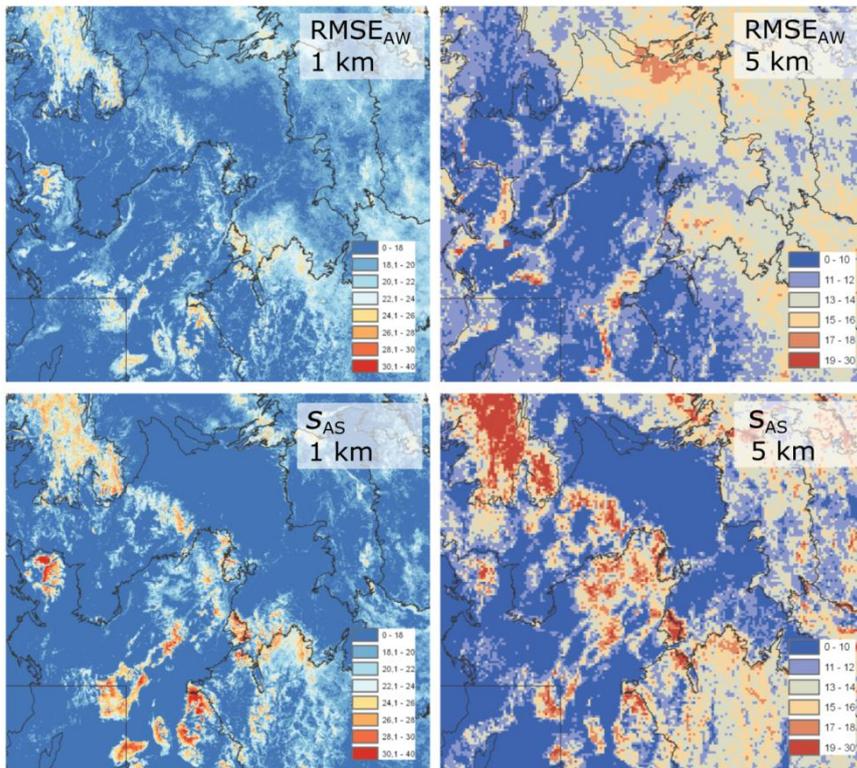


Figure 33. Maps of observed  $RMSE_{AW}$  (upper part) and EP error  $s_{AS}$  (below) computed at the 1 km (left) and 5 km (right) spatial scales.

The correspondence of  $s_{AS}$  and observed  $RMSE_{AW}$  is pronounced at the 1 km scale. The highs (> 24%) and lows (<18%) of both images correspond well. This corresponds to the findings of section 4.3.2.1. Nevertheless, the correspondence decreases at the 5 km scale (Figure 33 right panel). This is most probably due to the rapidly decreasing noise of the ASAR GM SSM dataset and the fact that  $s_{AS}$  loses its dominant effect to the AWRA-L error estimate in the computation of the predicted  $RMSE_{AW}$ . This is demonstrated by the higher observed  $RMSE_{AW}$  values in northeastern part of the image. This may be caused by the cracking clay soils (Liu et al., 2010). These were discussed in the literature as complicated to account in models and passive microwave data (Liu et al., 2010).

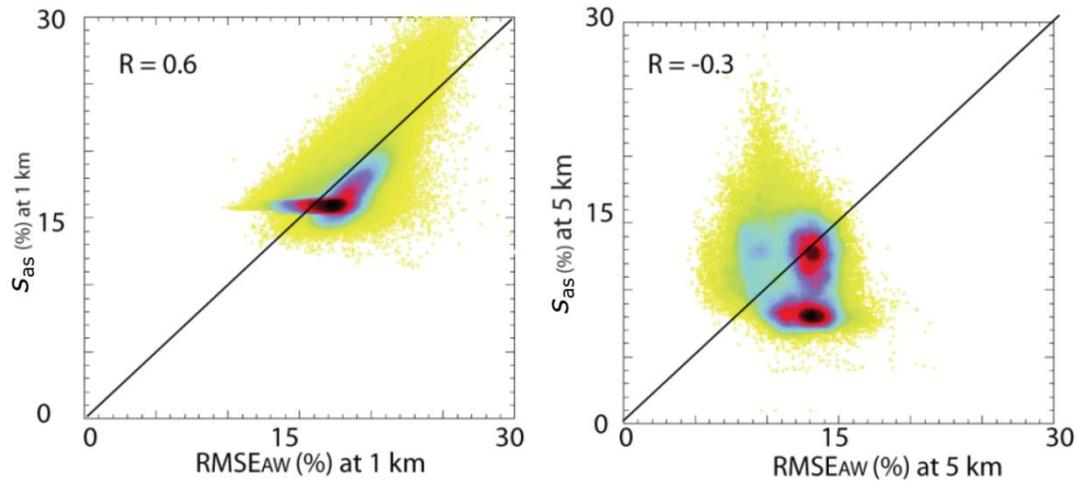


Figure 34. Scatterplots between observed  $RMSE_{AW}$  and EP error ( $S_{AS}$ ) computed at the 1 km (left) and at the 5 km (right) spatial scale computed over the area of interest from the Figure 33. The color represents the density of points ranging from high (black) to very low (yellow).

The above findings in Figure 34 are expressed as scatterplots. The scatterplot on the left side suggests that the equation 3-10 has probably a too high starting point. The scatterplot exhibiting the patterns at 5 km scale (Figure 34 right panel) reveals a dual behaviour of the image suggesting that by spatial averaging the low error values decrease more abruptly than the higher error values.

#### 4.3.3 Triple collocation (TC)

The TC method has been in detail described in section 2.2.2.2 and its usage for the ASAR GM was introduced in the section 3.2.2. The TC errors are marked with \* signaling that the data were prior transformed to dynamics of one of the datasets. The main objective was to assess  $\tau^*$  of ASAR GM SSM ( $\tau_{AS}^*$ ) and to investigate trends in  $\tau_{AS}^*$  related to a) the change in spatial resolution from 5 to 25 km, and b) the choice of a hydrological model.

The results and discussions are organized as follows. Section 4.3.3.1 assesses the spatial patterns of  $\tau_{AS}^*$  at 5 and 25 km scale. This is followed by the assessment of the TC errors of AMSR-E ( $\tau_{AM}^*$ ) and AWRA-L ( $\tau_{AW}^*$ ) in section 4.3.3.2. Finally, section 4.3.3.3 reveals trends in  $\tau_{AS}^*$  related to the choice of the third dataset. For this purpose the triple collocation is repeated using two different, globally available, model reanalysis datasets (ERA-Interim and GLDAS-NOAH). Section structure

##### 4.3.3.1 Assessment of the ASAR GM errors

The resulting ASAR errors ( $\tau_{AS}^*$ ) are displayed in Figure 35, at the 5 and 25 km spatial resolutions, respectively. The spatial relative patterns of  $\tau_{AS}^*$  are discussed on a continental scale and then assessed over a region in southeastern Australia with a high ( $> 0.7$ ) temporal stability (Cosh et al., 2004). A high temporal stability is expected to mitigate the effect of scaling errors. The temporal stability was computed as a correlation of the local (1 km) to the regional (25 km) ASAR GM soil moisture values.

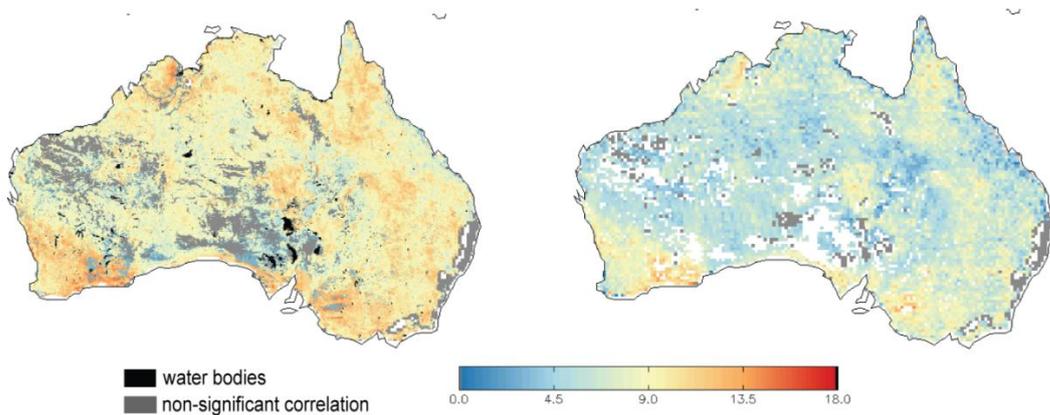


Figure 35. The spatial errors  $\tau_{AS}^*$  at 5 km and 25 km overlaid with the Interim Biogeographic Regionalization dataset for Australia (IBRA) (Thackway & Creswell, 1995). The units are percentage (%) of saturated soil moisture. The grey areas (and the white areas in the right image) had non-significant correlation values ( $p>0.05$ ).

Important to note is that the analyses were computed on absolute values according to Scipal et al. (Scipal et al., 2008b). The estimated errors provide the ability of the dataset to capture the absolute soil moisture levels and are hence comparable with a majority of the evaluation results presented earlier in this work.

The spatial patterns of  $\tau_{AS}^*$  act, similar to the standard evaluation measures (section 4.2.1), at large (>25 km) and medium (> 5 km) spatial scales. The medium patterns correspond to the different dense vegetation forms, soil roughness patterns or geomorphological features (e.g. mallee bushes in Mallee, or the hilly and rocky landscapes of the sparsely vegetated Pilbara region (Figure 36)).

**TC errors at 5 and 25 km** The absolute values of  $\tau_{AS}^*$  decrease when computed at the 25 km spatial resolution. This is explained by the improved radiometric accuracy achieved by averaging of the ASAR GM data. The original radiometric resolution of the ASAR GM mode is high (>1.25 dB) (equation 4-2) and further decreases according to equation 4-1 to about 0.21 dB for the 5 km resolution product. This is a reasonable estimate given that the estimated radiometric accuracy for the ERS scatterometer product was 0.2 dB (Attema & Lecomte, 1998).

The map computed at the 25 km spatial resolution provides significantly smoother patterns. Nevertheless, some of the features from the medium resolution map remain evident also at coarser resolution (i.e. the higher error estimates over eastern Mallee or the hilly landscapes of the sparsely vegetated Pilbara region). This is an interesting finding since the effect of these features was not encountered in the error assessments of the coarse resolution active microwave sensors (Dorigo et al., 2010). The pronounced effect of the roughness of soils and vegetation on the medium resolution SAR datasets is most probably caused by the coherent characteristic of the SAR signal that initiates interference over these areas and remains high even after averaging of the original SSM product to coarser scales.



Figure 36. Photographs downloaded from the Confluence Project website (“Degree confluence project,” 2008) representing (from upper left to lower right) portions of Pilbara, Mallee, Gulf Plains and Mount Isa Inlier IBRA Bioregions (Thackway & Creswell, 1995).

The low  $\tau_{AS}^*$  errors (< 10%) are found over central and south-central Australia and correspond to *The spatial patterns of the TC errors* regions with low annual precipitation and low soil moisture variation. Also the areas are dominated by relatively smooth surfaces with sand plains and sand dunes irregularly covered with shrubs or Eucalypts (e.g. Great Victoria Desert or Gibson Desert). Low values are also found in northern and northwestern Australia corresponding to areas with high soil moisture variation due to monsoonal systems. The low errors in the latter regions can be explained by the highly favorable conditions - high seasonality and penetrable vegetation - for detection of soil moisture with active microwave systems using change detection method.

Medium values (10 – 14%) dominate over southeastern Australia (i.e. NSW South Western Slopes or Riverina bioregion) and southwestern Australia (i.e. Avon Wheat belt); both are agricultural areas. The  $\tau_{AS}^*$  is higher over these regions than over northern Australia due to lower seasonality and attenuation by agricultural crops.

A high  $\tau_{AS}^*$  is found over densely vegetated areas and hilly and rugged hills. This is similar to results of the standard evaluation measures (chapter 4). In addition, high values are found over regions that flood periodically (i.e. Channel Country bioregion). This is expected, given that the ASAR GM SSM product is not masked for sporadically appearing water bodies. High errors were also found over central desert regions (i.e. Simpsons and Strzelecki deserts).

The actual soil moisture variation of the three compared products and their impact on  $\tau_{AS}^*$  are discussed in detail below for three selected points in southeastern Australia. The selection of this region was motivated by high temporal stability values (> 0.7) (Figure 37 right panel) that minimize error due to spatial scaling. The average error over southeastern Australia (region enlarged in Figure 37) is 14%.

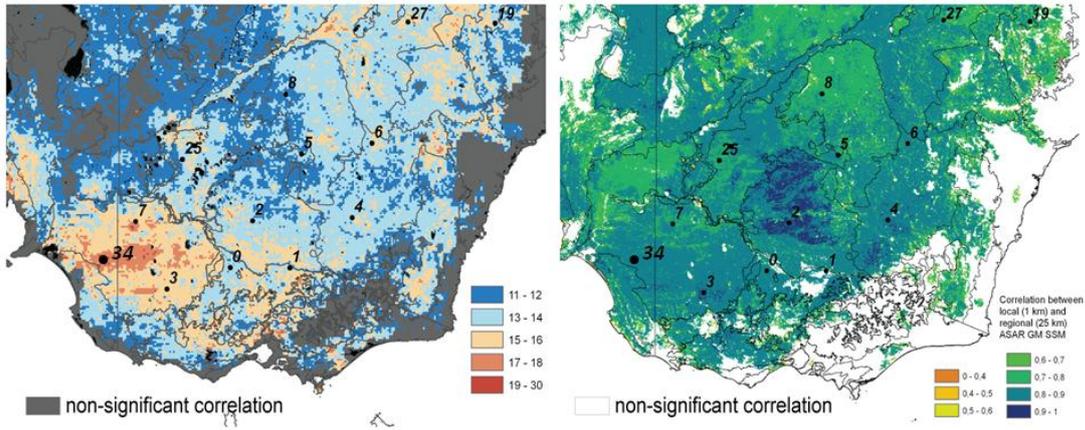


Figure 37. The spatial errors  $\tau_{ASAR}^*$  at 5km overlaid with the borders of the Interim Biogeographic Regionalization dataset for Australia (IBRA) (Thackway & Creswell, 1995) (left) and the correlation layer computed between the local (1 km) and the regional (25 km) soil moisture (right). OzNET station k4 is located in the vicinity of region 3. The units are percentage (%) of saturated soil moisture. Points 0 and 34 refer to regions 1 and 3, these were discussed later in this section. The grey and white areas had non-significant correlation values ( $p>0.05$ ).

The TC errors explained in time-series

The triple collocation method relies on a multiplication of differences computed between cross-calibrated datasets (see section 2.2.2.2). Therefore, by looking into time-series of absolute values and differences between the three datasets, ASAR GM, AWRA-L, and AMSR-E, over region 1, OzNET station K4, and region 2 (Figure 39 to Figure 41) one can better understand the origin of  $\tau_{AS}^*$ .

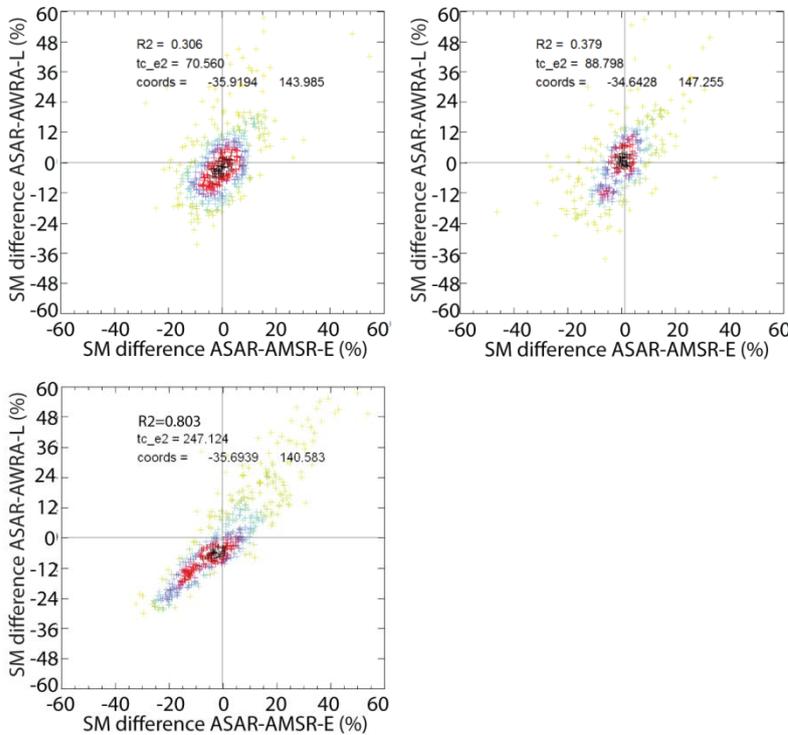


Figure 38. Scatterplot of differences between ASAR and AWRA-L and ASAR and AMSR-E SSM estimates over region 1, OzNET station K4, and region 3. The units are percentage (%) of

saturated soil moisture. The color represents the density of points ranging from high (black) to very low (yellow).

An interesting finding relates to the equation 2-22. While the left side of the equation can be assumed positive given the square root sign, the right side can have both signs as it is computed as an averaged multiplication of the differences between random soil moisture datasets. This is a rather interesting context that has not been addressed in the evaluation studies of soil moisture before. The results demonstrated that the vast majority of the retrieved  $\tau_{AS}^*$  remain positive (Figure 37) with only rare negative values if coarser resolution models were used instead of the AWRA-L model. Figure 38 and also Figure 39 to Figure 41 provide an explanation why  $\tau_{AS}^*$  remains positive. The separate multiplications of the differences are not random, rather jointly positive or negative. If ASAR GM SSM is higher than AWRA-L SSM, then it is also higher than AMSR-E SSM and vice versa. This signals large independency of the ASAR data caused probably by the combined effect of a) the coarser resolution or the forcing of the AMSR-E and AWRA-L data and b) the lower radiometric accuracy of the ASAR GM data.

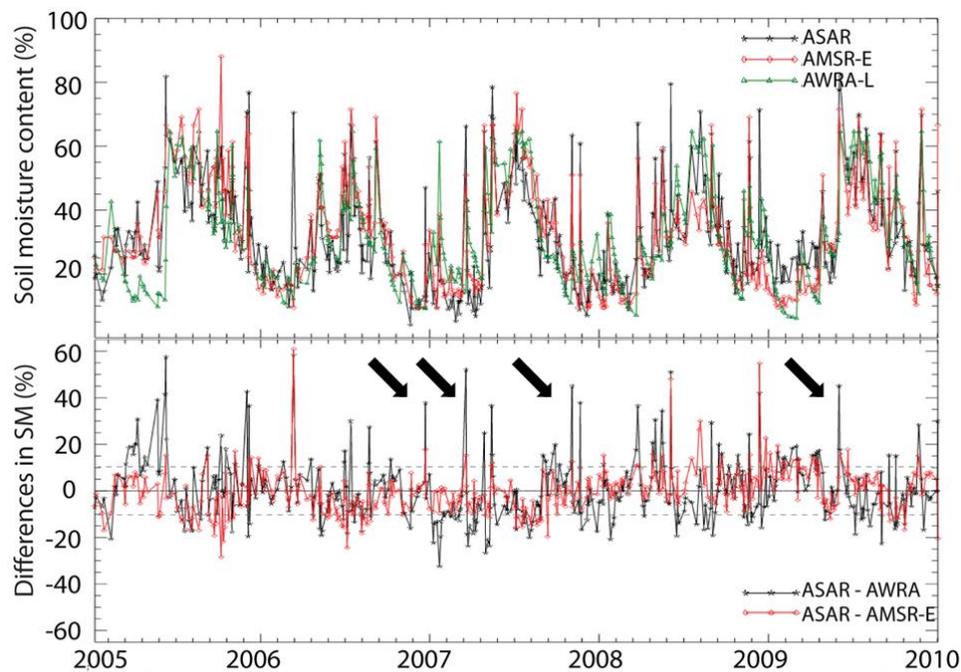


Figure 39. The time-series of the ASAR, AWRA-L, and AMSR-E SSM datasets (upper part) and the time-series of the differences between the ASAR and AWRA-L, and ASAR and AMSR-E SSM (lower part) for the region 1. The units are percentage (%) of saturated soil moisture.

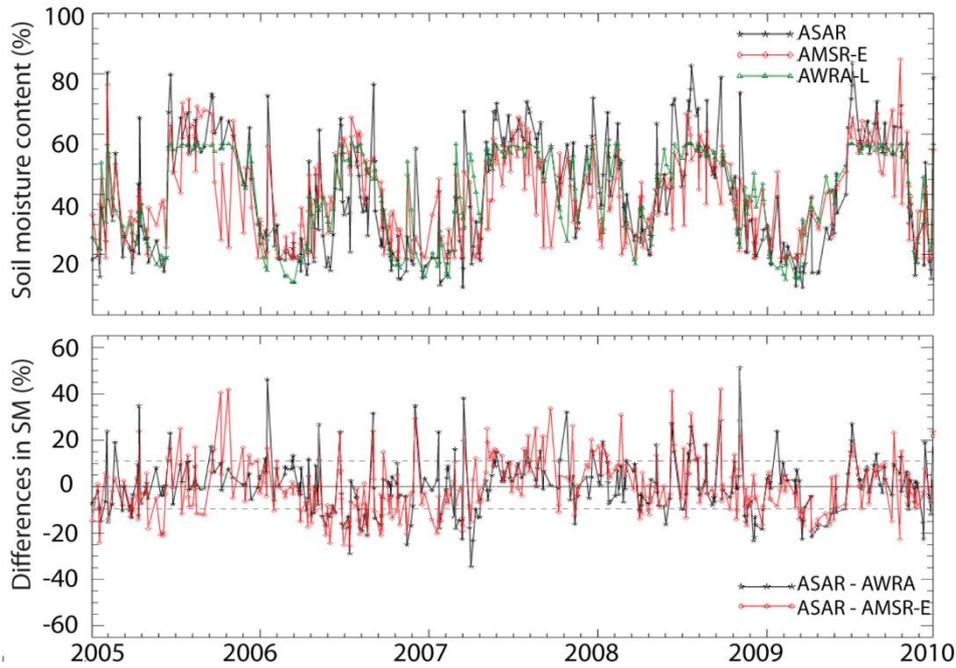


Figure 40. The time-series of the ASAR, AWRA-L, and AMSR-E SSM (upper) and the time-series of differences between ASAR and AWRA-L, and ASAR and AMSR-E SSM (lower) for OzNET station K4. The units are percentage (%) of saturated soil moisture.

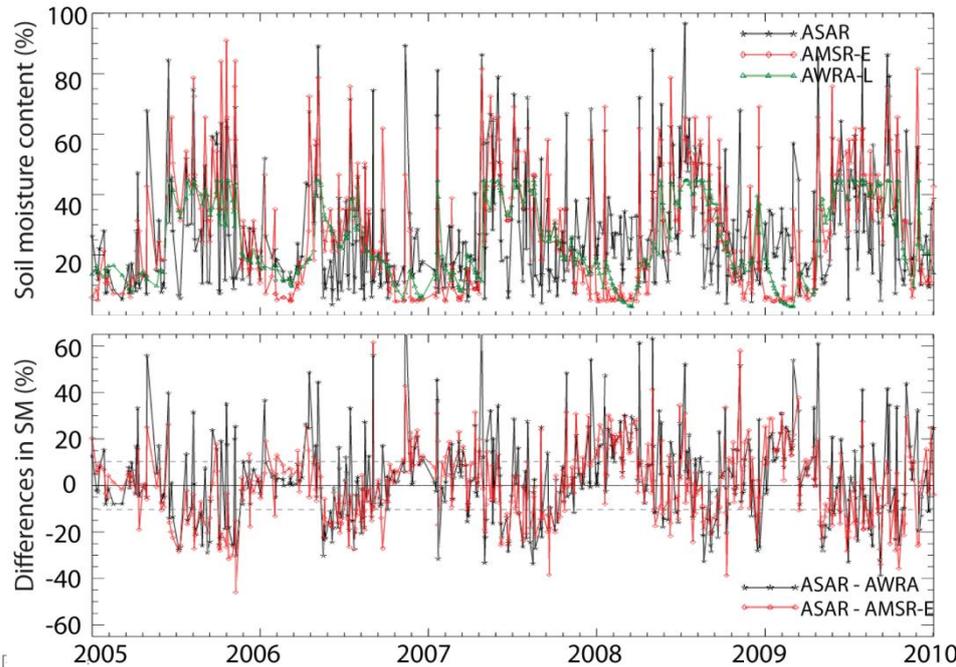


Figure 41. The time-series of the ASAR, AWRA-L, and AMSR-E SSM (upper) and the time-series of differences between the ASAR and AWRA-L, and ASAR and AMSR-E SSM (lower) for region 3. The units are percentage (%) of saturated soil moisture.

Assessing Figure 39 to Figure 41 in detail, the time-series appear impacted by two types of *Seasonality* fluctuations; one representing the seasonal effect and a second one, imposed on top of it, representing the effect of short-term events (several days). The ability to capture both

fluctuations is reflected in the final  $\tau_{AS}^*$  estimates. The ASAR GM data capture the absolute seasonality very well over region 1 and the OzNET station K4. This is evident in the absolute time-series and further supported by the limited seasonal trend in the SSM differences (Figure 39 and Figure 40, lower parts). On the contrary, Figure 41 demonstrates a low ability of the ASAR GM data to capture soil moisture fluctuations; this is mainly caused by the dense vegetation cover (open woodland and shrubland).

The ASAR GM and AMSR-E estimates document strong moisture peaks that are not captured by the AWRA-L model (depicted by the arrows in Figure 39). The latter is also exhibited in Figure 38 as a cloud of points with high values on y- and close to 0 values on x-axes. It can be assumed that the increase corresponds to precipitation events because of the sudden increase in both ASAR GM and AMSR-E time-series that have independent inputs. The possible reasons why AWRA-L does not capture these peaks might be a) the error of the input precipitation data in the AWRA-L model, b) the simplifying assumptions on infiltration and evaporation in the model, c) the temporal mismatch of several hours that was allowed between datasets, and d) the point character of the precipitation data, the main forcing of the AWRA-L model, that may be located out of the path of the storm.

To investigate the origin of the inconsistencies between ASAR GM, AWRA-L, and AMSR-E and to assess if these refer to the, by the AWRA-L dataset missed, precipitation event, the model was replaced with data from OzNET in-situ network and the soil moisture differences were re-plotted. Figure 42 demonstrates the differences from the Figure 40 (ASAR GM – AWRA-L and ASAR GM – AMSR-E) along with the differences between ASAR – OzNET and ASAR GM – AMSR-E. Only limited number of large differences between ASAR GM – AWRA-L disappeared or got mitigated when AWRA-L was exchanged with the OzNET data. Furthermore, additional peaks appeared. These can be explained by a) temporal mismatch of several hours that was allowed between datasets and b) point character of the precipitation data that may be located out of the path of the storm.

The above results demonstrated that exchanging the modelled data with an in-situ soil moisture data may lead to analogous problems of spatial scaling related to point characteristic of dataset forcing. Furthermore, the results showed the complexity in determining a reference true soil moisture dataset.

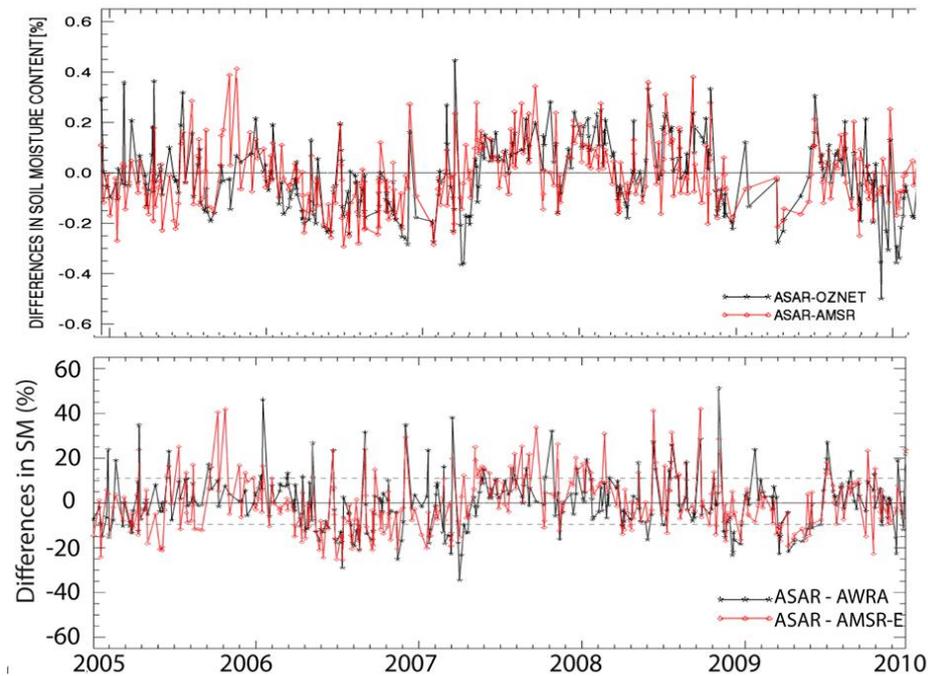


Figure 42. The time-series of soil moisture differences expressed as a percentage (%) of saturated soil moisture. The difference between ASAR and AWRA-L SSM, and ASAR and AMSR-E SSM (upper part), and the differences between the ASAR GM and OzNET SSM, and ASAR GM and AMSR-E SSM (lower part) over OzNET station K4.

A special attention should be given back to Figure 38. This demonstrates high correlations between distinct time-series differences. While the differences are by no means identical to dataset error, the demonstrated high correlations between separate differences may provide an inference that also the dataset errors may be correlated. This would violate the assumption on identity and zero cross-correlation (section 3.2.2).

#### 4.3.3.2 Assessment of the AMSR-E and AWRA-L errors

The triple collocation errors of the AMSR-E SSM ( $\tau_{AM}^*$ ) and AWRA-L SSM ( $\tau_{AW}^*$ ) datasets are displayed in Figure 43. The  $\tau_{AM}^*$  remains below 6 % over vast portion of the continent; this corresponds to the results of Dorigo et al. (Dorigo et al., 2010) and can be explained by the high radiometric accuracy as well as by the five times coarser spatial resolution of the AMSR-E sensor. An exception is the forested eastern and southeastern Australia and monsoonal, very wet and mangrove covered, northern Australian coast. Similar findings were found and in detail discussed for  $RMSE_{AM}$  (section 4.2.1.1).

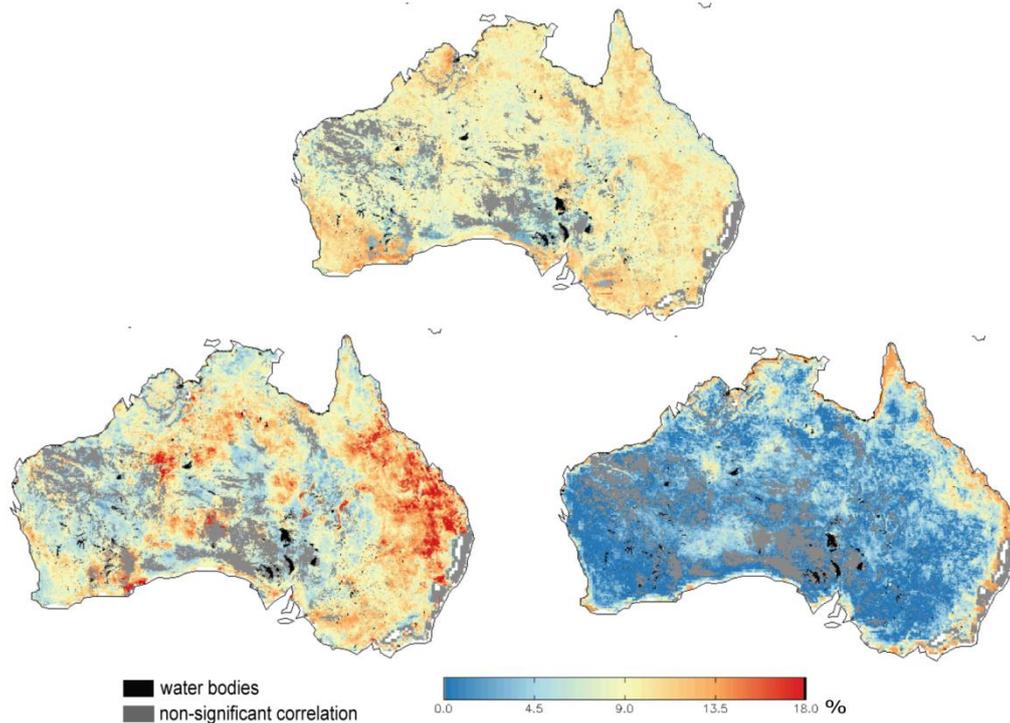


Figure 43. The triple collocation error for ASAR GM SSM ( $\tau_{AS}^*$ ) (above), AWRA-L SSM ( $\tau_{AW}^*$ ) (below left), and AMSR-E SSM ( $\tau_{AM}^*$ ) (below right) soil moisture datasets computed at the 5 km spatial resolution with ASAR GM as a reference. The units are percentage (%) of saturated soil moisture.

The  $\tau_{AW}^*$  demonstrates highs (>12%) over areas with medium annual precipitation (approximately 300-800 mm). Especially high values (>16%) are encountered over sparsely forested, clay soil dominated northeast regions. A possible explanation for high values over these areas is the simplifications in soil types in the AWRA-L model that mitigate differences between different soils and the fact that clay soils act differently in different time of the year. In particular, the soils are expected to fasten evapotranspiration, infiltration, and change porosity during part of the year. The lows (<8%) dominate central arid and northwestern Australia.

The  $\tau_{AW}^*$  is considerable larger than  $\tau_{AM}^*$  and  $\tau_{AS}^*$ . This can be explained by the compounded effect of the following: a) a minimizing effect of spatial errors if the major product's forcing has spatial instead of point character, b) a minimizing effect of spatial mismatch with coarsening of the spatial resolution, c) differences in data distribution of AWRA-L comparable to other datasets (AWRA-L can well demonstrate very low soil moisture and saturation stage; fails however when describing the mid-range soil moisture values), and lastly d) the fact that AWRA-L model could not depict several large-magnitude events detectable by both the ASAR GM and AMSR-E SSM (Figure 39 to Figure 41).

#### 4.3.3.3 TC computed with several models

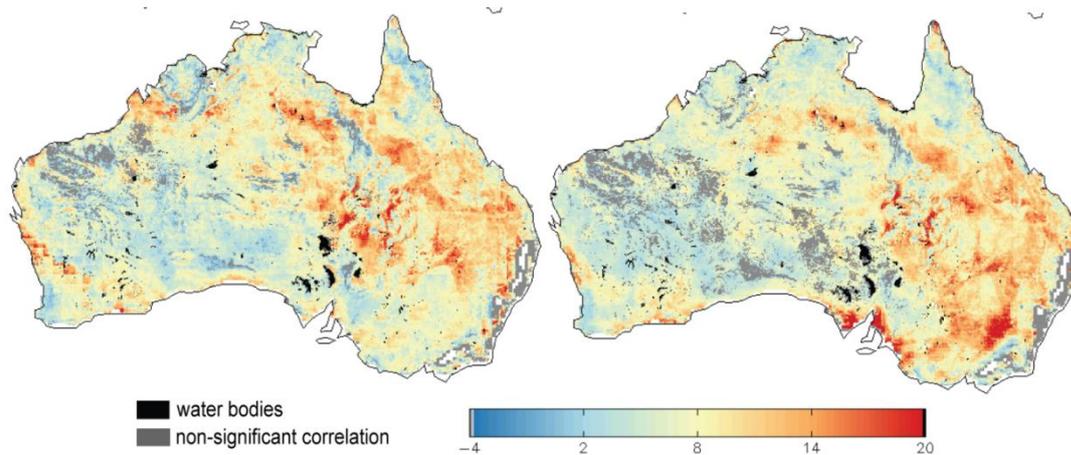


Figure 44. The triple collocation error of the a) ERA interim and b) GLDAS models computed at 5 km spatial resolution ASAR GM data as a reference. The units are percentage (%) of saturated soil moisture.

This section investigates the effect of exchanging a third dataset in the triple collocation with other models. In particular, the AWRA model was exchanged with GLDAS-NOAH and ERA-Interim models and the resulting  $\tau_{GL}^*$ ,  $\tau_{ER}^*$ ,  $\tau_{AS}^*$  and  $\tau_{AM}^*$  were assessed. The other two dataset, ASAR and AMSR-E, remain the same. The GLDAS-NOAH and ERA-Interim estimates were expressed in the ASAR dynamics. The selection of the third dataset should not influence the spatial patterns of  $\tau_{AS}^*$ , if the random errors of all three datasets are assumed to be uncorrelated (Dorigo et al., 2010).

#### Model error

The spatial patterns of the estimated error of the GLDAS-NOAH ( $\tau_{GL}^*$ ) is similar to  $\tau_{AW}^*$ ; lower error are demonstrated for the ERA Interim dataset ( $\tau_{ER}^*$ ). The quality of the rainfall forcing and the representative depth has a large impact on the quality of soil moisture. The better performance of  $\tau_{ER}^*$  may be explained by the fact that the ERA Interim assimilates several microwave radiances (rather than using rain rates) to improve sensitivity to atmospheric temperature, moisture; cloud water, and precipitation (cite Simmons ECMWF). Important to note is also that the ERA-Interim system runs at 80 km spatial resolution while GLDAS-NOAH provides soil moisture at the resolution corresponding to 0.25 degree.

#### Effect on the ASAR and AMSR-E error

Next, the impact of exchanging the modeled dataset on  $\tau_{AS}^*$  is investigated. The Figure 45 demonstrates  $\tau_{AS}^*$  computed with three different model inputs. The expectation was that exchanging a third dataset does not influence the spatial patterns of  $\tau_{AS}^*$  if all errors of the datasets are independent (Dorigo et al., 2010). The overall patterns of high ASAR GM SSM error seem to correspond well in all three maps. The largest errors are in southwestern, southeastern Australia, and in desert regions and act at coarse resolution as well as at medium resolution scales.

Nevertheless, the results also demonstrated alternations to  $\tau_{AS}^*$  that seemed to mirror the effect of a third dataset. For instance,  $\tau_{AS}^*$  appears as the highest when computed with the AWRA-L model and much lower when computed with the ERA-Interim model. Similar results were encountered also in the RMSE maps and were interpreted as effects of the second dataset (section 4.2.1.1). As a result, the alternations to  $\tau_{AS}^*$  might be explained by the error of AWRA-L

acting at medium resolution scale that is expected to be larger than error of the other models acting at coarser resolution scales (ERA-Interim and GLDAS-Noah) or, for instance, by the several microwave radiances assimilated by the ERA-Interim dataset.

It is suggested that introducing an additional datasets to the TC method may improve understanding of the ASAR GM SSM error. Other option may be to acquire the minimum  $\tau_{AS}^*$  from several TC runs that used diverse second and third dataset as an input.

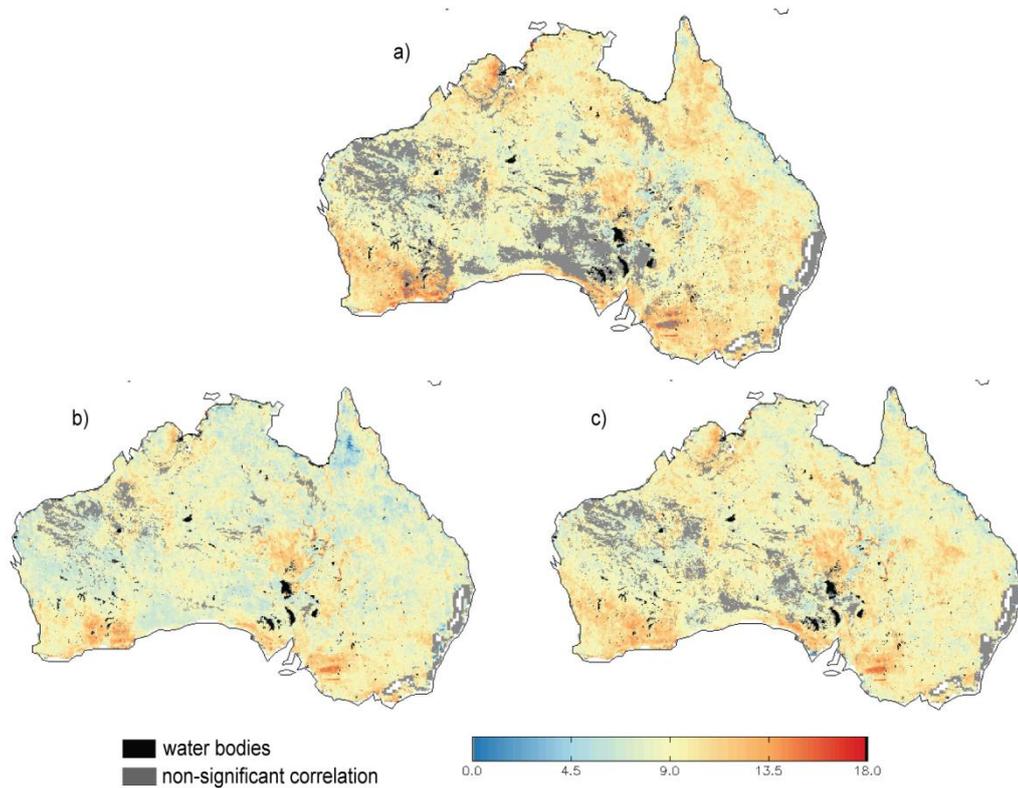


Figure 45. The triple collocation error of ASAR GM SSM ( $\tau_{AS}^*$ ) when computed with AMSR-E SSM as a second and a) AWRA-L, b) ERA-Interim, and c) GLDAS-NOAH as a third dataset, respectively. The units are percentage (%) of saturated soil moisture.

Similarly, the effect of different models on  $\tau_{AM}^*$  was investigated. The general patterns correspond and demonstrate low values over majority of the continent with exception of the coastal areas. Nevertheless, an evident decrease of  $\tau_{AM}^*$  is evident when ERA-Interim model was used.

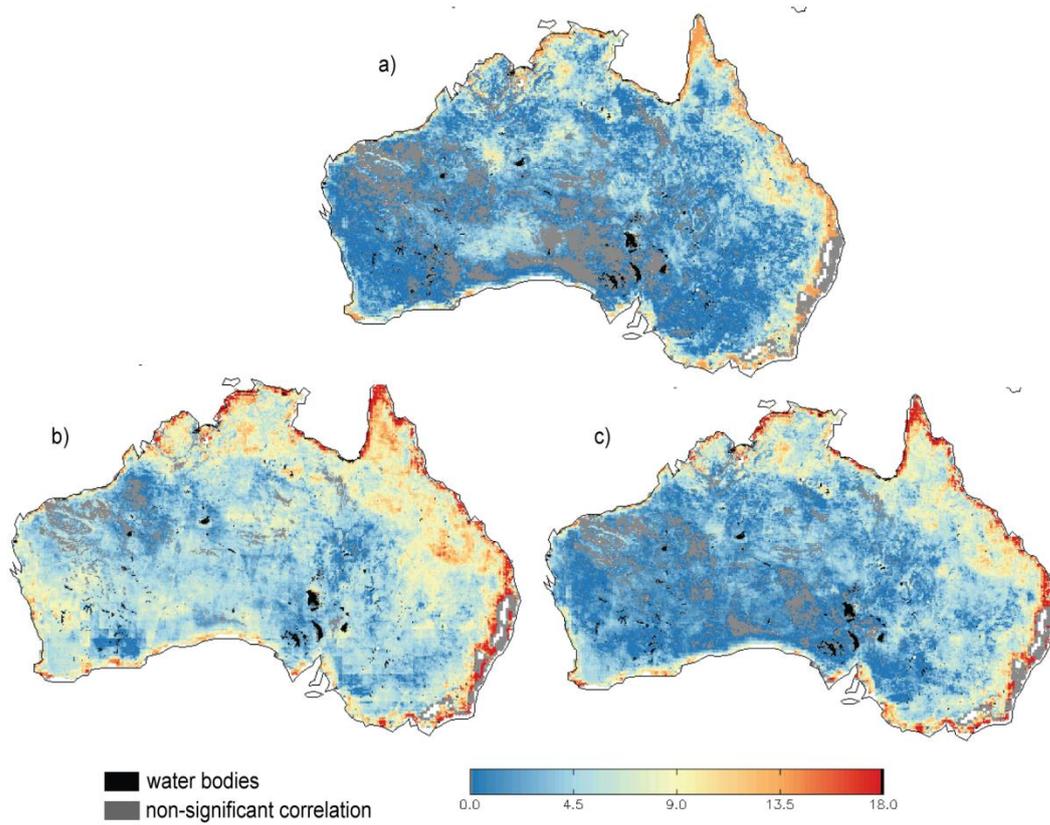


Figure 46. The triple collocation error of AMSR-E SSM ( $\tau_{AM}^*$ ) when computed with ASAR GM SSM as a second and a) AWRA-L, b) ERA-Interim, and c) GLDAS-NOAH as a third dataset, respectively. The units are percentage (%) of saturated soil moisture.

## 5. Discussion

This thesis was motivated with a view to evaluate the quality of the ASAR GM SSM medium resolution dataset and to provide guidance on appropriate evaluation methodology applicable to any SSM product. This chapter discusses the overall quality of the ASAR GM SSM and future Sentinel-1 SSM products and addresses common misconceptions and how these can be avoided in data evaluation studies. It is written in form of questions that were amply motivated in the introduction of this thesis (section 1.2) and addressed in chapters 2 and 4.

### 5.1 Can we apply the evaluation requirements of SMOS and SMAP to ASAR GM SSM?

The SMOS and SMAP missions SSM product requirement relies on a condition of  $RMSE < 0.04 \text{ m}^3/\text{m}^3$  (Kerr et al., 2010). Furthermore, the WMO recently published requirements on soil moisture quality that also rely on error defined as a single value per given spatial resolution (WMO, 2012). These requirements were driven by a) the results of in-situ campaigns that demonstrate  $0.04 \text{ m}^3/\text{m}^3$  as the typical spread of soil moisture observations, b) by an acceptable estimation of the evaporation and soil transfer when  $RMSE < 0.04 \text{ m}^3/\text{m}^3$ , and c) by the fact that RMSE is currently the most commonly used measure of precision. The two publications above motivated the scientific question: 'Can we apply the evaluation requirements of SMOS and SMAP to the medium resolution ASAR GM SSM product?' The question can be further split up between: "Shall we rely on an usage of a single RMSE criterion?" and "Can we setup an absolute threshold as an RMSE requirement?". *Justification of the question*

Before answering these questions, let us better understand the differences in the units. The relative units of the ASAR GM SSM product range between 0.0 and 100 % of saturation. The range of the volumetric units spans between 0.0 and  $1.0 \text{ m}^3/\text{m}^3$ . To transfer volumetric to relative units, the former needs to be divided by soil porosity. This is because the volumetric soil moisture expresses the amount of water in the entire soil mass (pores and solid form) while the relative soil moisture expresses only amount of water in soil pores. Given the soil porosity over Australia ranging between 0.3-0.6 (Rodell et al., 2004) the transferred requirement on relative soil moisture ranges between 13 % of saturated soil moisture over low-porosity soils to 6.6 % of saturation over high-porosity soils.

The findings from chapter 4 demonstrated that all measures describe slightly different data quality. Inversely, it can be expected that no single measure can describe all qualities of the ASAR GM SSM product. Furthermore, section 2.3 demonstrated that no application relies only on a single quality of SSM data. These findings let us conclude that assessing a single measure when evaluating ASAR GM SSM may not be sufficient and gives an evident answer to the first question.

Several arguments to the second question are provided below:

First, if an absolute agreement between different SSM time-series differences originating from different sensing depths, represented spatial extent, and different units need to be removed. The selection of an appropriate transformation needs to be performed with a careful consideration of data distribution and assumptions of transformation techniques (section 4.1). *Bias removal*

*Spatially variant RMSE* Second, the section 4.2.1.1 showed that RMSE is a spatially variant measure that strongly reflects variances of datasets. As a result, a threshold of  $0.04 \text{ m}^3/\text{m}^3$  can be easier met in desert regions than over wet areas with high soil moisture variance. Furthermore, the Australian soils in wet areas have higher porosity than soils in dry regions (Rodell et al., 2004), can hold more water, and the soil moisture variance is therefore higher. Consequently, the threshold of  $0.04 \text{ m}^3/\text{m}^3$  becomes even more challenging to fulfill. It is evident that the general requirement on RMSE should be spatially varying and reflect PDF of soil moisture. Another possibility might be using the nRMSE measure as a threshold that is independent of data distribution (section 4.2.1.2).

*No absolute true SSM* Third, RMSE is computed between two datasets where one is often but wrongly assumed to be more accurate than the other. Nevertheless, there is no true estimation and assumptions about a true estimation have been demonstrated to generate a bias (Entekhabi et al., 2010). This is because every dataset, even a reference, possess random and systematic errors that affect the final value of RMSE. The systematic errors reflect the differences at spatial scales, sensed depths, and the dynamic of the models. These significantly influence RMSE (Figure 23). Removing systematic bias prior to RMSE computation is therefore essential (section 4.1 and 4.2.1.5).

*Effect of other datasets in the evaluation study* Fourth, RMSE estimates differed even after the removal of systematic bias (Figure 17). Such differences reflect random error combined with a systematic time variant error (e.g. due to missing parameter in the model) of the second dataset. These cannot be removed during dataset transformation (4.1.2). This finding suggests that independent methods need to be used to specify an absolute RMSE quality threshold (e.g. TC).

*Effect of spatial scale* Sixth, RMSE computed at 1 and 5 km spatial scale significantly differed (Figure 24) due to a) the noise that decreased with spatial resolution, b) the minimizing effect of spatial mismatch, and c) the increasing probability that the products react on identical atmospheric forcing (the AWRA-L SSM acts at 5 km scale). The requirements on accuracy of SMOS and SMAP missions were dimensioned to sensors with 25 km spatial resolution. If higher resolutions are evaluated the thresholds are expected to increase and contra versa.

The above findings demonstrated that the requirement on  $\text{RMSE} < 0.04 \text{ m}^3/\text{m}^3$  is not directly applicable to the medium resolution ASAR GM SSM dataset. The major reasons are summarized below:

- A single evaluation measure, such as RMSE, can only assess one dataset quality. Nevertheless, different applications were demonstrated to require a variety of qualities of the ASAR GM SSM product.
- Defining a single RMSE value as a threshold doesn't judge quality of the ASAR GM SSM over very wet and over very dry areas correspondingly.
- A portion of the  $\text{RMSE}_{\text{AW}}$  originated in the error of the reference dataset. The selection of diverse reference datasets with varying spatial resolution resulted in varying RMSE results.

Several recommendations are listed below that should enable an evaluation of the ASAR GM SSM and possibly of all other SSM products:

- Do not rely on a single evaluation measure. Use a combination of evaluation measures to estimate dataset quality (e.g. RMSE combined with assessment of evolution in time (e.g.  $R$  or  $R_s$ )).
- If assessing RMSE, set up spatially variant thresholds for RMSE reflecting PDFs or rely on nRMSE
- Transform acquired measures to a common dynamics (bias corrected measures).
- The level of acceptance of the evaluation measures should be derived using independent methods that are not effected by random or systematic error of the reference dataset (e.g. TC method).
- The RMSE threshold must refer to a specific spatial resolution.

## 5.2 How does the selection of spatial resolution influence error estimates?

Several methods were introduced in section 4.2 that estimate random and/or systematic errors of soil moisture datasets (TC, RMSE, MAE, and EP). It was demonstrated that the absolute values decrease and the spatial patterns of RMSE change with decreasing spatial resolution (Figure 24). It is anticipated that change in spatial scale influences also other absolute error estimates such as MAE, TC, or EP. This anticipation is supported by the findings of Martinez-Fernandez and Ceballos (Martinez-Fernandez & Ceballos, 2005) who demonstrated that scaling errors can be larger than the retrieval error of a single dataset. Similarly, Cosh et al. (Cosh et al., 2008) demonstrated that spatial representativeness of the ground observations largely influence the absolute error measure.

These findings are of a great importance because the usage of datasets with varying spatial resolution is very common in evaluation studies and will increase with the upcoming launch of new coarse (e.g. SMAP) and medium (e.g. Sentinels) resolution sensors. This justified the question: ‘How does the selection of spatial resolution influence error estimates?’ While numerous studies addressed the role of scaling error for evaluations in-situ – remotely sensed data (e.g. Michael H Cosh et al., 2008; Miralles et al., 2010) there is only limited literature addressing the scaling problem in evaluations studies performed between spatial data with different resolutions. *Justification of the question*

The absolute values of  $RMSE_{AW}$  decreased with decreasing spatial resolution. Important to note is that also the spatial patterns changed; especially so over northeastern Australia and desert regions (Figure 24). The observed RMSE was impacted by several factors. First, it is influenced by random and time-variant systematic errors of the two datasets (see section 2.2.3.1); for instance  $RMSE_{AW}$  is impacted by errors of ASAR GM and AWRA-L SSM. The latter errors act at different spatial scales. In particular, the AWRA-L errors are expected to be related mainly to rainfall forcing acting at large scales ( $> 25$  km), and to rock outcrops, salt lakes and soil types acting at medium scales ( $\geq 5$  km). The ASAR GM errors are affected by medium scale geomorphological and land cover patterns acting at medium scales ( $\geq 1$  km). Second, if errors of the datasets are not fully independent their error covariances may also impact RMSE (Zwieback et al., 2012) and this is not easy to estimate. Third, RMSE decreases with decreasing resolution as a result of improved radiometric accuracy of the dataset (section 4.2.1.6). *Effect of scale on RMSE*

The above findings suggest that, at 1 km resolution, RMSE is impacted mainly by the errors of the ASAR GM SSM product whereas errors of both datasets impact the patterns at 5 km scale. This results in differences in spatial patterns as well as in absolute values seen in the two RMSE maps. The theory was supported in section 4.3.2 by a close correspondence between predicted RMSE and  $s_{AS}$  at 1 km but an insignificant correlation at 5 km scale (Figure 34).

The question is if one can generalize this finding to say that the absolute errors are always dominated by the errors of the finer dataset. This, however, does not appear so evident. For instance,  $RMSE_{AG}$ ,  $RMSE_{AE}$ , and  $RMSE_{AM}$  seemed to be impacted by random errors of both datasets when assessed at 5 km scale (section 4.2.1.1) even though the ASAR GM SSM was retrieved at finer, 5 km, and the other datasets at coarser, > 25 km, spatial resolution. The assumption is that the ASAR GM errors largely decreased when averaged to 5 km (section 4.2.1.6) what transformed their errors to a comparable range with the errors of the coarser resolution datasets. Importantly, the actual proportion of the separate errors remains unknown.

#### *Effect of scale on TC*

Similar scaling effects influenced the triple collocation method. The interpretation is more complex given that a third dataset is required by the method. Figure 45 demonstrates results of  $\tau_{AS}^*$  computed using AMSR-E, and three models acting at 5 km and at 25 km spatial scale, respectively. The triple collocation method relies on the availability of three datasets; their errors are expected to mitigate each other given their independency. The AMSR-E errors are expected to be of smaller magnitude than errors of the medium resolution datasets (ASAR GM and AWRA-L), partly also due to the high noise of the ASAR GM data. As a result, the independent error differences cannot fully mitigate themselves (Figure 39 – Figure 41) because the individual covariances are not equal to 0. As a result,  $\tau_{AS}^*$  may reflect some portion of the random and systematic error of AWRA-L SSM.

Figure 38 demonstrates that different error magnitudes may result in a dependency of the separate differences. This may signify dependency of the errors themselves what would violate the TC presumptions. Such violation could be a reason why the selection of the third dataset, despite the differing expectations (Dorigo et al., 2010), influenced the final estimates of  $\tau_{AS}^*$ .

Despite the above hesitations the study was performed using data with different spatial resolutions simply because no other than AWRA-L dataset existed at the time of this research at a scale corresponding to the ASAR GM scale.

### **5.3 Is there a best combination of measures to describe the quality of ASAR GM SSM dataset?**

#### *Justification of the question*

This work demonstrated that a) every application has slightly different requirements on dataset quantities (section 2.3), and that b) there is no single measure that can describe all dataset qualities (chapter 4). The appropriate evaluation method seems to consist of combination of several evaluation measures based on the application (some were suggested in Table 3). Often, the application of the data is not known at the stage of the algorithm development. In such cases, several measures need to be assessed that each represents different dataset quality and so that possibly all qualities are described. To easier assess the needs for such method the evaluation measures were divided in to three groups according to the quality these describe: a) random and

systematic error assessment (RMSE, MAE, EP, TC), b) assessment of the evolution in time ( $R$ ,  $R_S$ ), and c) bias.

The discussion below summarizes and confronts evaluation methods for these three groups. The comparison is based on the results achieved in chapter 4 for the ASAR GM data.

### 5.3.1 Absolute evaluation measures

This chapter summarizes and confronts following absolute evaluation measures – RMSE, MAE, TC, and EP. The discussions assume a good quality of the ASAR GM SSM model. While RMSE, MAE, TC, and EP measures are similar in the sense that they refer to dataset absolute error they differ due to the different weight they assign to time-variant systematic and random errors (section 2.2.3.2).  $\tau_{AS}^*$  was computed using ASAR GM, AMSR-E, and AWRA-L SSM dataset.

Throughout this sections MAE was used instead of RMSE. Following few paragraphs justify why MAE was prioritized. *RMSE versus MAE*

RMSE is one of the most common evaluation measures used to estimate dataset error. An important feature is that it quadratically penalizes residuals between parameters to address problem of large residuals. By doing so the measure accounts for the dataset variability and magnitude field but loses its functional relationship with absolute error (Willmott & Matsuura, 2005). MAE, on the contrary, has a functional relationship with the mean absolute error.

Section 4.2.1.3 demonstrated that the relative patterns of RMSE and MAE are extremely similar and the absolute values only slightly differed. While MAE doesn't warn about existence of large variability within datasets, it can be easily interpreted as a mean absolute error between the ASAR GM and AMSR-E SSM datasets.

The discussion above and the demonstrated results in 4.2.1.3 suggest that MAE has a better potential as an estimate of the dataset mean absolute compared to RMSE. These findings are supported by findings of Willmott and Matsuura (Willmott & Matsuura, 2005) who first suggested that teasing variance out of evaluations of absolute error may be inappropriate.

Importantly, Figure 21 demonstrated that the difference between error (MAE) and standard error (RMSE) is minimal and thus MAE can be used to compare against other absolute standard errors such as EP or TC, using MAE as a measure that is very easy to interpret.

The theoretical differences and similarities between RMSE and  $\tau_{AS}^*$  were discussed in section 2.2.3.2. Given the similarities between RMSE and MAE the findings can be easily transformed also to MAE. Assuming a good quality of the ASAR GM SSM model a good correspondence between  $\tau_{AS}^*$  and MAE would indicate a) a minimum error of the reference datasets in MAE, and b) a complete fulfilment of the assumptions on  $\tau_{AS}^*$ ; especially about the independency of the residual errors (section 2.2.3.2). *MAE and TC*

The results are displayed in Figure 47 and demonstrate a significant correlation between  $\tau_{AS}^*$  and MAEs. This finding that would, according to the theory from section 2.2.3.2, suggest minimum error of the reference dataset in MAE is surprising mainly because RMSE and MAE was found strongly impacted by the error of the second dataset in section 4.2.1. As more realistic therefore appears that both of the assumptions a) and b) were violated and that both MAE and TC methods

address a combination of random and a systematic time-variant error of both (all three in case of the TC method) datasets.

The highest correspondence was found between  $\tau_{AS}^*$  and  $MAE_{AM}$ . This is probably influenced by the selection of the reference dataset (AMSR-E) in both the TC and MAE method. The impact of the third dataset on  $\tau_{AS}^*$  when computed at 5 km scale has been demonstrated in Figure 45.

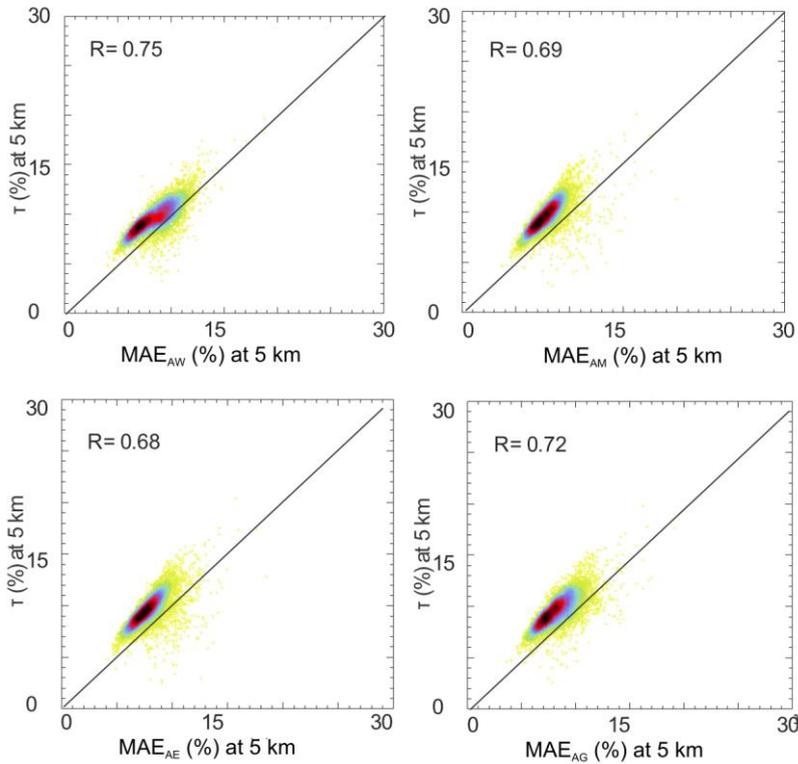


Figure 47. Scatterplots displaying relationship between MAE and  $\tau_{AS}^*$ . The MAE was computed using ASAR GM SSM and a) AWRA-L dataset ( $MAE_{AW}$ ), b) AMSR-E dataset ( $MAE_{AM}$ ), c) ERA-Interim dataset ( $MAE_{AE}$ ), and d) GLDAS-Noah ( $MAE_{AG}$ ). The analyses were performed at the 5 km scale. The units are percentage (%) of saturated soil moisture. The color represents the density of points ranging from high (black) to very low (yellow).

**Spatial patterns** The spatial patterns between  $MAE_{AM}$  and  $\tau_{AS}^*$  are displayed in Figure 47. Given the above listed results the comparison was performed using  $MAE_{AM}$  that demonstrated the highest correspondence with  $\tau_{AS}^*$ . The relative patterns of highs and lows in MAE and  $\tau_{AS}^*$  maps correspond very well with highs in southwestern, eastern, northern Australia, and over desert regions. As already discussed in section 4.2.4, MAE is higher than  $\tau_{AS}^*$  because it weights the systematic time-variant error differently.

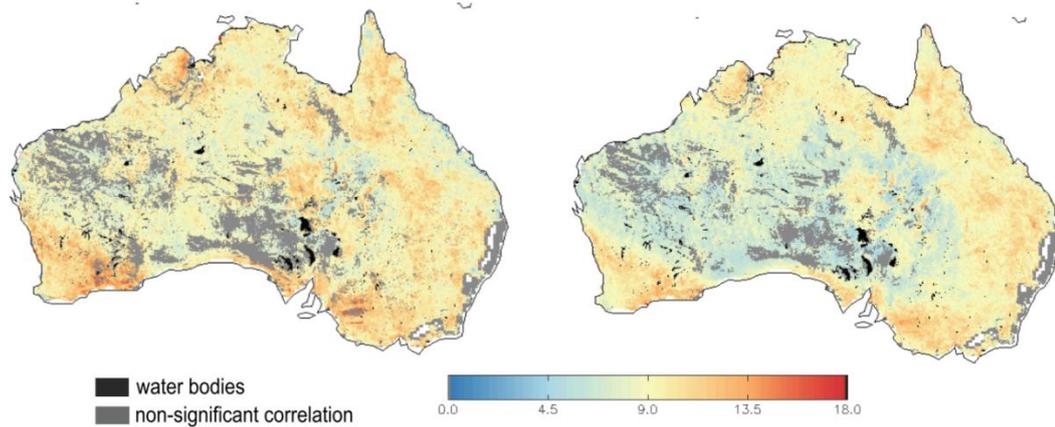


Figure 48.  $\tau_{AS}^*$  computed using AMSR-E and AWRA-L (left) and  $MAE_{AS}$  computed between ASAR GM and AMSR-E. The grey areas display the non-significant correlation values ( $p > 0.05$ ). The analyses were performed at the 5 km scale. The units are percentage (%) of saturated soil moisture.

The differences and similarities between RMSE and  $s_{as}$  were amply discussed in section 2.2.3.2. A *MAE and EP* high correspondence of EP and MAE would indicate a) a minimal error of the reference dataset, and b) a good understanding of the propagated errors in the EP method. Vice versa, low correspondance would refer to the error of the reference dataset, or to a wrong understanding of the propagated errors in the EP method.

The results comparing MAE and  $s_{AS}$  are displayed in Figure 50 and demonstrate non-significant correlation for all combinations of  $s_{AS}$  and MAEs. The low correlation is expected to be impacted by a time-variant systematic error introduced by the second dataset. This doesn't impact  $s_{AS}$  but does MAE. The importance of the time-variant systematic error has been addressed in 4.2.1.1. This is a new result suggesting that previous findings (Doubková et al., 2012) demonstrating the ASAR GM SSM error as the main source in RMSE hold only at the resolution of the ASAR GM dataset (1 km).

Other explanations of the low correlations explore the propagated errors in EP. First, the errors propagated in EP were assumed to be constant in time. This may have been violated especially over areas with high seasonality (northern Australia). Second, given that EP, as derived in this thesis, doesn't address the variance of SSM data. Last reason, though probably of minor importance, is the fact that the separate propagated errors are not necessary independent as demonstrated on the relationship of slope ( $\beta$ ) and sensitivity ( $S$ ) over Oklahoma (Pathe et al., 2009b).

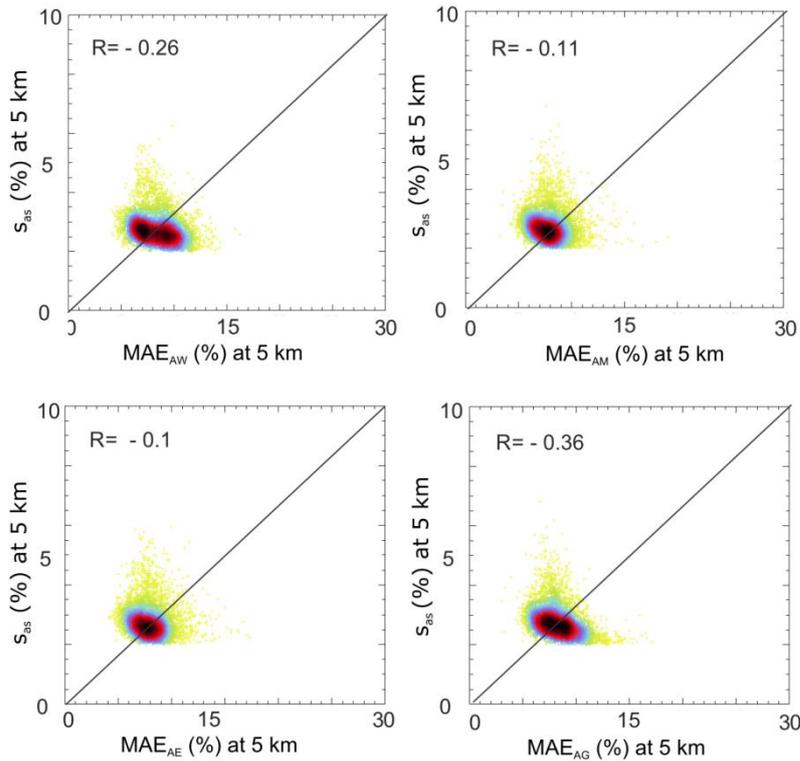


Figure 49. Scatterplots displaying the relationship between MAE and  $s_{AS}$ . The MAE was computed using ASAR GM SSM and a) AWRA-L dataset ( $MAE_{AW}$ ), b) AMSR-E dataset ( $MAE_{AM}$ ), c) ERA-Interim dataset ( $MAE_{AE}$ ), and d) GLDAS-Noah ( $MAE_{AG}$ ). The analyses were performed at the 5 km scale. The units are percentage (%) of saturated soil moisture. The color represents the density of points ranging from high (black) to very low (yellow).

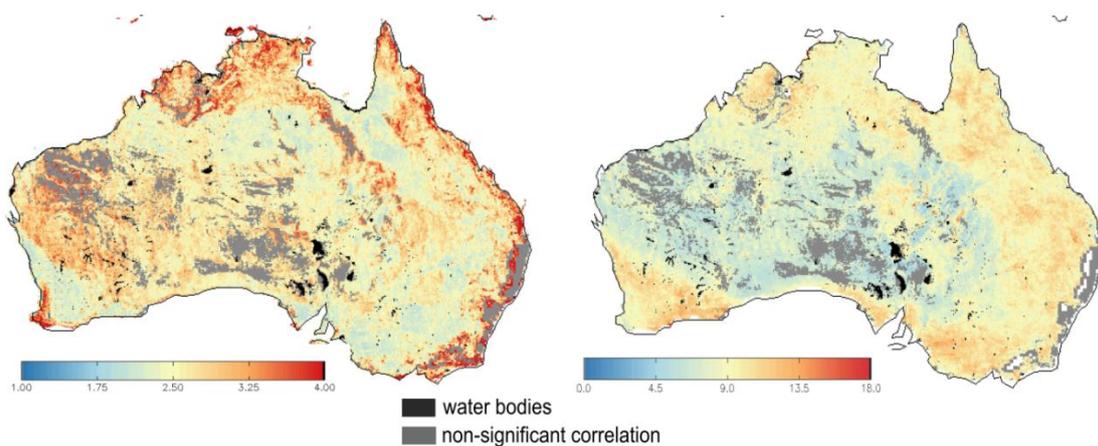


Figure 50.  $s_{AS}$  (left), and  $MAE_{AW}$  computed between ASAR GM and AWRA-L. The grey areas display the non-significant correlation values ( $p > 0.05$ ). The analyses were performed at the 5 km scale. The units are percentage (%) of saturated soil moisture.

**Spatial patterns** Figure 49 provides a spatial understanding to the relationship between  $MAE_{AW}$  and  $s_{AS}$ . Severe differences between MAE and  $s_{AS}$  were evident over most of the continent. In fact, the relationship seems inversed and supports the findings in Figure 49.

The absolute values differ substantially with the 0 – 4 % saturation ranges for  $s_{AS}$  and 0 – 18 % saturation ranges for  $MAE_{AW}$ . Differences in absolute values were expected (section 4.2.4) as  $MAE_{AW}$  estimates both random and systematic error, while  $s_{AS}$  only estimates the random error. The following discussion only compares relative values evident in Figure 50 as highs (red) and lows (blue).

While relative MAE is higher than  $s_{AS}$  in the desert regions, over coastal regions, and in southeastern Australia, the opposite trend is shown in eastern and northern Australia. The possible reason for the high  $MAE_{AW}$  over desert regions may be the high, but not accounted in the EP, variability of the ASAR GM sensitivity ( $S$ ) (section 3.1.1) in the central deserts caused by the severe but exceptional rains. The high sensitivity propagates the EP through the model and decreases the final  $s_{AS}$  without considering its high seasonal variance. Also, the parameter  $S$  is derived from the reference probabilities of the ERS scatterometer (Pathe et al., 2009b). Some inaccuracies in the references should therefore be expected given the differences in the sensitivity of the SAR and scatterometer to roughness effects of soils and vegetation. Similarly, the higher relative values of  $s_{AS}$  in northern and northeastern Australia may be explained by the potential underestimation of  $S$ .

The differences and similarities between  $s_{AS}$  and  $\tau_{AS}^*$  were amply discussed in section 2.2.3.2. A *EP and TC* good correspondence of the EP and TC error maps would indicate a) a complete fulfilment of the TC assumptions and b) a good understanding of the propagated errors in the EP.

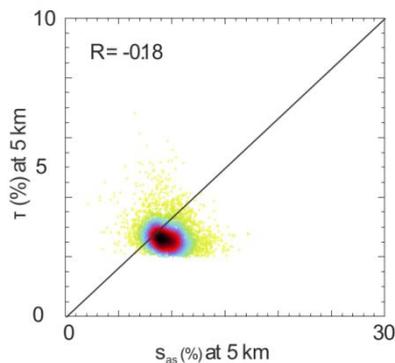


Figure 51. Scatterplot displaying the relationship between  $s_{AS}$  and  $\tau_{AS}^*$  errors. The analyses were performed at the 5 km scale. The units are percentage (%) of saturated soil moisture. The color represents the density of points ranging from high (black) to very low (yellow).

The results are displayed in Figure 51 and Figure 52 and demonstrate a non-significant *Spatial patterns* correlation between the two measures. The possible explanations coincides with those explaining the differences between  $s_{AS}$  and  $MAE_{AW}$  and relate to the fact that a) the  $s_{AS}$  error represents only random error while other measures, included MAE and  $\tau_{AS}^*$ , may be effected by the time-variant systematic error and b) some assumption of the EP method may be violated (e.g. independency of input parameters, or the assumption that these are stable in time).

The first point deserves further attention. The parameter  $\tau_{AS}^*$  was assumed not to be effected by random and time-invariant errors of the second and the third dataset (Dorigo et al., 2010) as the independent dataset errors mitigate each other. Nevertheless, this assumption has been shown violated in section 4.3 when data of different spatial resolutions and noise levels were used. As a

result the spatial patterns of  $\tau_{AS}^*$  derived with other models somewhat differed (Figure 45) and reflected errors of the models used in the TC. Despite the latter, the TC method demonstrated a good ability to detect errors of AMSR-E dataset (Dorigo et al., 2010; Parinussa et al., 2011).

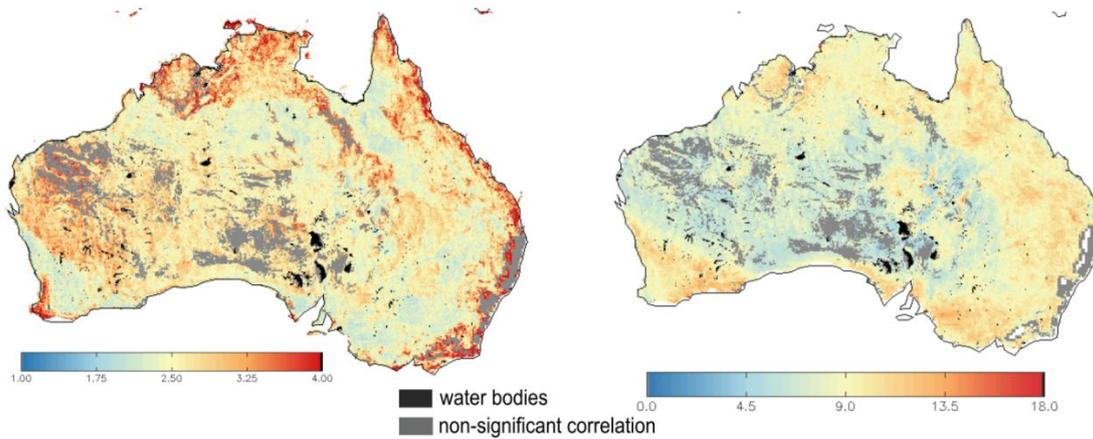


Figure 52.  $s_{AS}$  (left) and  $\tau_{AS}^*$  computed using AMSR-E and AWRA-L (right). The grey areas display the non-significant correlation values ( $p > 0.05$ ). The analyses were performed at the 5 km scale. The units are percentage (%) of saturated soil moisture.

### 5.3.2 Relative evaluation measures

Two evaluation measures were introduced in chapter 4.2.2 – namely  $R$  and  $R_5$ . It was expected that  $R_5$  mitigates the effect of seasonality that is so strongly imprinted in  $R$ . Following section investigates the differences in qualities between these two measures.

As demonstrated in Figure 53 the difference between the two coefficients for the ASAR and AMSR estimates ranges between -0.4 to 0.3.  $R_5$  is generally lower over central and western Australia. Here, a minimal soil moisture variation, intercepted with rare but severe rainfall events, can be expected. While  $R$  values may be increased by the magnitude of these rains, the effect on  $R_5$  is expected to be minimal as  $R_5$  is computed on ranked dataset and doesn't so strongly relate to rain event's magnitudes.

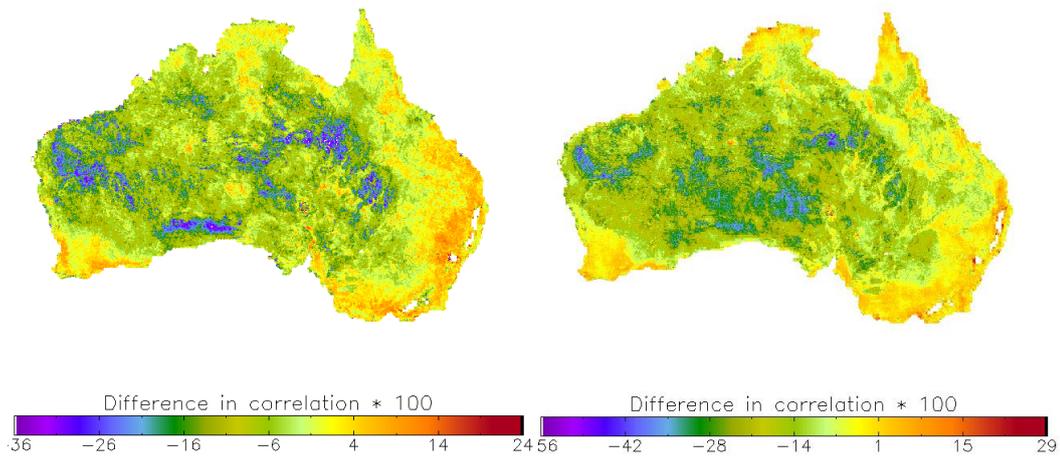


Figure 53. The difference between  $R_s$  and  $R$  (computed as  $R_s - R$ ) between ASAR GM and AWRA-L SSM (left) and ASAR GM and AMSR-E SSM (right).

An opposite trend is evident over southwestern and southeastern Australia and, for the case of AWRA-L data, also over eastern Australia. The potential reason for the improvement in  $R_s$  over vegetated areas may be the non-linear behavior of AWRA-L dataset that impacts to a lesser extent  $R_s$  than  $R$ . Another explanation may be connected to the large biomass. Above a certain biomass level the backscatter signal is less sensitive to soil moisture, causing a saturation effect and thus non-linear behavior. While this may deteriorate  $R$ , it has a negligible effect on  $R_s$ .

### 5.3.3 Resume

Important to note is that the goal of the above section was not to decide why one measure is better than the other, rather to understand their discrepancies and correspondences.

The assumptions of the TC method have been summarized in section 2.2.2.2. Not fulfilling some of these (e.g. independency of errors) may influence the final TC assessment what was amply addressed by (Zwieback et al., 2012). In this thesis, the TC method appeared not to be able to eliminate the random and time-variant systematic errors of the second and the third dataset. The reasons are not evident but some suggestions are provided below. *TC method*

First, it is suggested that the TC method could not fully function with datasets acquired at different spatial resolutions, and so, with datasets demonstrating different error magnitudes. The second possibly reason may be that the errors were not fully independent as initially assumed (section 3.2.2). This is not easy to proof, but an indication of the latter may be the dependency of the separate differences demonstrated in Figure 38.

The effect of errors of ancillary dataset was assessed by exchanging the third dataset and studying the impact on  $\tau_{AS}^*$  (Figure 45). Similar results can be expected if exchanging the second dataset. Despite the latter, the TC method demonstrated a good ability to detect errors of the three models and AMSR-E dataset, when compared to independent studies (Dorigo et al., 2010; Parinussa et al., 2011). The general spatial relative patterns in all three  $\tau_{AS}^*$  acquired with three different models also corresponded (Figure 45). Nevertheless, there seemed to be a secondary impact of the errors of other datasets used in the evaluation. This fact was further supported by the high correspondence between  $\tau_{AS}^*$  and MAE (Figure 47).

*MAE* The interpretation of MAE is the most evident from all error estimates – it claims to reflect both the random and the systematic time-variant error of both datasets. Nevertheless, the errors of the second dataset are not desired when assessing error characterization.

*EP* An important finding was presented in the preceding chapter expanding results of recent publication (Doubková et al., 2012). In particular, the random error of the ASAR GM SSM has been addressed by the latter publication as the main contributor to the spatial variability of RMSE computed between ASAR GM SSM and an independent model. This thesis study demonstrated that this no longer holds after averaging both datasets to 5 km scale. At this scale, the errors of both datasets contribute to the final RMSE estimate.

Second, the EP error is not affected by the systematic errors of other datasets and importantly, it doesn't account for the variance of SSM. A correspondence with MAE or TC (both reflecting SSM variance and errors of other datasets) should have not be expected at first place. Instead, it is suggested that the difference between MAE and  $\tau_{AW}^*$  may provide a good estimate about the random error of the ASAR GM SSM dataset comparable to  $s_{as}$ . This assumption is only valid under the condition that all errors in the EP are addressed and correctly propagated.

*Bias* Bias between soil moisture datasets has only minor importance for the majority of soil moisture applications as it is often removed prior to the applied studies. It was added at the end of this chapter and analyzed in section 4.2.1.4 as it a) plays an important role in preliminary dataset evaluations that are commonly performed by a simple visual comparison of absolute values of soil moisture maps and b) may bring further understanding to the separate algorithms.

*Relative eval. measures* Interesting finding arose from the comparison of  $R$  and  $R_s$  measurement. Their differences were able to depict effects of high seasonality of saturation of SSM data.

## 5.4 What is the quality and what are the limitations of ASAR GM SSM data?

The evaluation results from chapter 4 were summarized in the preceding chapter. The goal was to identify and understand the reasons for the discrepancies and correspondence between separate evolution measures. This chapter summarizes the same results from the user point of view.

*Spatial resolution* The largest advantage of the ASAR GM SSM dataset is undoubtedly its spatial resolution. Until recently, only coarse resolution soil moisture datasets were available that often discouraged the hydrological community operating at local (meters or few km) scales. The ASAR GM data are provided at 1 km spatial resolution and allow detection of spatial patterns that cannot be detected in coarse resolution sensors. This is an advantage especially over areas with precipitation and landcover heterogeneity (Meier et al., 2010; Pathe et al., 2009b).

*Temporal resolution* Furthermore, the ASAR GM SSM onboard ENVISAT allows for SAR acquisitions every 2-3 days over Australia and approaches so the requirement of WMO (WMO, 2012) of at least daily coverage of soil moisture for medium resolution products. As such the ASAR GM SSM product is a first sensor with quasi-operational abilities (Doubkova et al., 2009).

High spatial resolution decreases radiometric accuracy. Analyses demonstrated that at 1 km resolution the ASAR GM SSM is influenced by large noise (> 1.2 dB) that rapidly decreases to 0.21 dB when averaged to 5 km spatial resolution (section 4.2.1.6). It is expected that providing ASAR GM SSM dataset at > 5 km resolution may increase the number of data users. Similar findings were suggested by (Thoma et al. 2008) who showed a need of spatial averaging and filtering to a regional scale (i.e. 0.5 or 1 km) for high resolution radar imagery.

*Radiometric resolution*

Table 5. The mean and standard deviation of EP, TC, MAE, R, and Rs measures computed as an average over the entire continent. The analyses were performed at 5 km and 1 km (only for the purpose of comparison) spatial resolution. Over 250 000 pixels were used.

	Median (%) (5km)	Stdev (%) (5 km)	Median (%) (1 km)	Stdev (%) (1 km)
TC <sub>AW</sub>	9.44	2.12		
TC <sub>AE</sub>	8.45	1.88		
TC <sub>AG</sub>	9.20	1.80		
EP	2.65	0.54		
MAE <sub>AW</sub>	8.42	1.75	17.8	2.90
MAE <sub>AM</sub>	7.81	1.49		
MAE <sub>AE</sub>	7.81	1.49		
MAE <sub>AG</sub>	8.19	1.66		
R <sub>AW</sub>	0.44	0.19	0.31	0.17
R <sub>AM</sub>	0.62	0.17		
R <sub>AE</sub>	0.62	0.17		
R <sub>AG</sub>	0.57	0.17		
RS <sub>AW</sub>	0.39	0.21		
RS <sub>AM</sub>	0.46	0.22		

The estimated errors (Table 5) are several times higher than the average error of the ERS SSM (0.028 m<sup>3</sup>/m<sup>3</sup>) (Scipal et al., 2008b) and, as expected, also than the ASCAT SSM product over Australia. The only exceptions are the results of the EP that remain below 3 %. Given the very low estimated radiometric noise of the data at 5 km (0.21 dB) this number appears to be impacted mainly by the errors of the dry and wet references, respectively. These were approximated to 2% in equation 3-12. Given the assessment with all other independent error estimates this number appears unrealistically low and requires further investigation.

*Absolute and relative error estimate*

If approximated from the Table 5, the actual median error of the ASAR GM SSM lies between 2.65–9.44 % with standard deviation between 0.54–2.12 % of saturated soil moisture. Importantly, these are median values; the actual spatial distributions of the errors provided in chapter 4 should be given an equal level of importance.

While the continental averages of the correlation coefficients stayed rather low (Table 5), several regions demonstrated a very high correlation ( $R_s > 0.8$ ) and suggest an ability of the ASAR GM SSM to detect anomalies that are essential for drought and flood monitoring. These were namely southeastern, southwestern, and northern Australia

The high error estimates of the ASAR GM SSM in the original resolution (right columns in Table 5) may be a reason why, until now, the data were only applied after averaging of the original product to coarser spatial resolutions. The ASAR GM SSM data were applied for crude qualitative (Water Research Commission, 2012) as well as quantitative (Van Dijk & Warren, 2010) evaluations of soil moisture datasets. Other applications implemented the data to detect a bias in the coarse resolution precipitation dataset (Milzow et al., 2010). Given the average error between 2.65–9.44 % at 5 km resolution further applications are anticipated.

Input of soil moisture data into assimilation systems as well as direct input studies demonstrated a diversity of results (section 2.3). No evident indication was provided explaining the conditions under which remotely-sensed dataset improve a model.

To conclude, several reasons were addressed explaining the current lack of applications of the ASAR GM SSM data. These include: a) a missing understanding of the required quality and quantity of remotely-sensed data in data assimilation systems, b) a low radiometric quality, and c) a low revisit period when compared to the coarse resolution dataset. Nevertheless, as demonstrated in this work, the error of the ASAR GM SSM improves rapidly with averaging already to 5 km scale. Also, order of magnitude improved radiometric quality is expected from the upcoming Sentinel-1 sensor. Second, the problem of temporal gaps could be solved using double collocation technique (Jin & Henderson, 2011) and third, new assimilation studies are on the way that are expected to better characterize the requirements of remotely sensed data in data assimilation schemes.

## 5.5 Learning from ASAR GM SSM errors for Sentinel-1

The evaluation approach demonstrated in this work is applicable to any remotely sensed soil moisture dataset and can therefore be used to assess the quality of Sentinel-1. It is expected that the influence of surface features, such as vegetation and roughness, will be more pronounced in the Sentinel-1 scale than it was in the ASAR GM. As a result, additional parameters will need to be added to the algorithm that will account for such effects in the change detection algorithm.

In addition, the demonstrated modelling difficulties of soil moisture at fine scales (Thoma et al. 2008) suggest averaging and filtering of the raw Sentinel-1 data to a regional scale (i.e. 0.5 or 1 km). The question arises how the effect of surface features will emerge at this resolution. A detailed discussion on the Sentinel-1 algorithm and error model is beyond the scope of this work, but anticipated modifications are likely to include:

- The improved revisit period might improve the estimation of the individual model parameters.
- Due to a characteristic local incidence angle (Hornacek et al., 2012) the final error of Sentinel-1 product is expected to be smaller due to the missing effect of slope error
- The final error is expected to improve by an order of magnitudes due to a) the improved radiometric resolution of the Sentinel-1 backscatter measurements (0.128 dB) (Snoeij et al. 2010) compared to ASAR GM (1.2 dB) and b) the averaging and filtering of the raw data to regional scale (i.e. 0.5 or 1 km)

Currently, data assimilation and direct input of the ASAR GM SSM to models may be restricted by its poor radiometric resolution and revisit period. The proposed soil moisture product from

Sentinel-1 has a foreseen coverage every six days globally, nearly daily over Europe and Canada (depending on latitude) (Hornacek et al., 2012) and by a greatly improved radiometric accuracy. As such, it has the potential to be of great benefit for data assimilation, anomaly and threshold detection, as well as, direct input into models operating at medium resolution scales.

## 6. Conclusion

This work comprised three major objectives. First of all, it provided state of the art methods for evaluation measures of SSM supplemented by a concise overview of existing applications and their suggested evaluation measures. Second, it comprehensively evaluated the ASAR GM SSM medium resolution product, and third, it provided general guidance on an appropriate evaluation methodology applicable to any SSM product. The results were discussed in the form of answers to several questions identified in the introductory part of the thesis, which include following:

1. Can we apply the evaluation requirement of SMOS and SMAP to ASAR GM SSM?
2. How does the selection of spatial resolution influence the error estimates?
3. Is there a best combination of measures to describe the quality of the ASAR GM SSM dataset?
4. What is the quality and what are the limitations of the ASAR GM SSM data?
5. Learning from ASAR GM SSM errors for Sentinel-1.

As an answer to the first question, the results demonstrated that the SMOS and SMAP evaluation requirements on soil moisture datasets relying solely on RMSE evaluation measure are not applicable to other SSM dataset due to the fact that a) no single evaluation measure can assess all the qualities of a dataset, b) RMSE is a spatially variant. Therefore, it doesn't judge the corresponding quality of the data over wet and over dry areas, and c) a portion of the  $RMSE_{AW}$  originates in the error of the reference dataset and thus the selection of diverse reference datasets with varying spatial resolution results in varying RMSE. It was recommended:

- not to use a single evaluation measure, rather to rely on a combination of evaluation measures to estimate dataset quality (e.g. RMSE combined with the assessment of evolution in time (e.g.  $R$  or  $R_s$ ));  
if assessing RMSE then to:
- setup a spatially variant level of acceptance for RMSE reflecting its PDFs (or to rely on nRMSE);
- derive the level of acceptance of the evaluation measures using independent methods that are not affected by random or systematic error of the reference dataset (e.g. TC method);
- transform the acquired measures to a common dynamics (bias corrected measures);
- derive the thresholds with a consideration of a specific spatial resolution.

As answer to the second, the major importance of the effect of scaling on the evaluation results was highlighted. In particular, it was demonstrated that the absolute evaluation measures (e.g. RMSE, MAE, TC) between two or more datasets are impacted unequally by errors of the datasets and that the individual error weights are linked to the dataset spatial resolutions. For instance, at 1 km resolution, the spatial patterns of RMSE were dominated by the errors of the ASAR GM SSM product (Doubková et al., 2012), whereas errors of both the datasets impacted the patterns at 5 km scale. These findings are of a major importance due to the frequent usage of datasets with varying spatial resolution in evaluation studies, which is expected to increase with the upcoming launch of new sensors operating at coarse (e.g. SMAP) and medium (e.g. Sentinels) resolutions.

It was also interesting to note that the expected error of SSM dataset increases as the spatial resolution decreases; most probably due to various factors such as soil and vegetation roughness. The requirement of WMO with inverse requirement trend appears inadequate (WMO, 2012).

The third question was addressed by assessing the quality of the ASAR GM SSM product using a number of standard (RMSE, MAE, or bias) and advanced (EP, TC) evaluation measures, as well as a combination of remotely sensed (AMSR-E), modeled (AWRA-L, ERA-Interim, GLDAS-NOAH), and in-situ (OzNET) SSM estimates. The measures were divided into three groups in accordance to the attributes: a) random and systematic error assessment (RMSE, MAE, EP, TC), b) assessment of the evolution in time ( $R$ ,  $R_s$ ), and c) bias.

The results demonstrated that no single measure could depict all qualities of the dataset. Even within the same group, the measures demonstrated different data qualities, which prevented finding one best combination of measures – absolute and relative – to use in evaluation studies. Yet, following important aspects were found:

- The impact of errors of the supplementary datasets in the TC method could not be mitigated probably due to the different spatial resolutions used in the experiment.
- As such, TC and MAE closely corresponded yielding a good estimate of the combined random and systematic errors of two (and three in the case of TC) datasets.
- Despite the fact that TC, MAE, and EP are representative of standard error, their meaning was in reality not the same. First, TC and MAE were impacted by random as well as systematic time-variant errors whereas EP only reflected random error. Second, under some conditions the TC method could be affected by errors of the second and the third dataset (as demonstrated in section 4.3.3). This was not the case for the EP method. Third, the EP method doesn't address the variance of data while MAE and TC do.
- Other methods are needed to evaluate the quality of the EP method at coarser (> 5 km) resolution. It is suggested to follow up on the method suggested by Doubkova et al. while improving the understanding of the error of the modeled dataset. For the latter the TC method can be applied. Needless to say, there is a need of better understanding of the impact of errors of the second and the third datasets.
- An interesting finding arose from the comparison of  $R$  and  $R_s$  measurements. Their differences were able to depict effects of high seasonality or saturation of SSM data.
- Though less important for the majority of soil moisture applications, the demonstrated large bias highlighted the fact that this measure could play an important role in the preliminary dataset evaluations that are commonly performed by a simple visual comparison of absolute values of soil moisture maps. In addition, it might bring further understanding in different algorithms.

The above discussion further motivated why relying on a single evaluation measure is not sufficient to describe overall data quality. Different situation arose if application of the data was known. For such cases a method called *Evaluation based on Application* (EbA) was introduced, which suggests a selection of the appropriate evaluation methods according to the application. This was summarized in section 2.3.

Fourth and fifth, the median values of error estimates for ASAR GM SSM over the Australian continent ranged between 2.65–9.44% with the standard deviation between 0.54–2.12% of

saturated soil moisture. The median of  $R$  over the entire continent varied between 0.44–0.62. The original error estimates were estimated to approximately 17% of saturated soil moisture. The ASAR GM SSM data were provided to users in the original form of 1 km data resolution (Doubkova et al., 2009) and applied for crude qualitative (Water Research Commission, 2012) and quantitative (Van Dijk & Warren, 2010) evaluations of soil moisture datasets. Other applications implemented the ASAR GM SSM data to detect a bias in the coarse resolution precipitation dataset (Milzow et al., 2010).

As of now, the application of the ASAR GM SSM data is still rather limited. The reasons for this limited applications were found to be a) a missing understanding of the required quality and quantity of remotely-sensed data in data assimilation systems, b) a low radiometric quality, and c) a low revisit period of the ASAR GM SSM when compared to the coarse resolution dataset. Nevertheless, these problems can be partly avoided. For instance, as demonstrated in this work, the error of the ASAR GM SSM improves rapidly with averaging already to 5 km scale. In addition, the problem of low radiometric accuracy should be eliminated in the upcoming Sentinel-1 sensor due to its by an order of magnitudes improved radiometric quality. The problem of temporal gaps could be solved using double collocation technique (Jin & Henderson, 2011) and third, new assimilation studies are on the way that are expected to better characterize the requirements of remotely sensed data in data assimilation schemes.

While individual evaluation methods have been introduced in earlier studies, this work is an innovative study as it provided a concise summary of evaluation measures combined with a demonstration of their shared use. Furthermore, the innovativeness of the theses lies in the transformation of the triple collocation evaluation method, applied until now only to evaluate coarse resolution (~25 km) datasets (i.e. Dorigo et al., 2010; Scipal et al., 2008), to the ASAR GM medium resolution SSM product.

Finally the last aspect of this study is that the findings and suggestions originating from the discussion are transferable to other satellite-derived soil moisture data. Of special interest is its transfer to data from the planned Sentinel-1 SAR sensor that shares similar technical characteristics but has an improved retrieval error comparable to the ASAR GM sensor. The operationally available medium resolution soil moisture from Sentinel-1 with a well-characterized error is likely to yield benefits for modelling and monitoring of land surface-atmosphere fluxes, crop growth and water balance applications.

## 7. Reference

- Albergel, C., Calvet, J. C., De Rosnay, P., Balsamo, G., Wagner, W., Hasenauer, S., Naeimi, V., et al. (2010). Cross-evaluation of modelled and remotely sensed surface soil moisture with in situ data in Southwestern France. *Hydrology and Earth System Sciences*, 14(11), 2177–2191.
- Attema, E. P. W., & Lecomte, P. The ERS-1 and ERS-2 wind scatterometers, system performance and data products. , 4 IGARSS 98 Sensing and Managing the Environment 1998 IEEE International Geoscience and Remote Sensing Symposium Proceedings Cat No98CH36174 2–4 (1998). doi:10.1109/IGARSS.1998.703710
- Aubert, D., Loumagne, C., & Oudin, L. (2003). Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall - Runoff model. *Journal of Hydrology*, 280(1-4), 145–161.
- Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., Van Den Hurk, B., Hirschi, M., & Betts, A. K. (2009). A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System. *Journal of Hydrometeorology*, 10(3), 623–643. doi:10.1175/2008JHM1068.1
- Bartalis, Z., Wagner, W., Anderson, C., Bonekamp, H., Naeimi, V., & Hasenauer, S. (2008). Validation of coarse resolution microwave soil moisture products. *International Geoscience and Remote Sensing Symposium (IGARSS)* (Vol. 2, pp. II173–II176). Boston, MA, July, 7-11, 2008.
- Bartsch, A., Wagner, W., Pathe, C., Scipal, K., Sabel, D., & Wolski, P. (2009). Global monitoring of wetlands – the value of ENVISAT ASAR global mode. *Journal of Environmental Management*, 90(7), 2226–2233.
- Beck, H. E., De Jeu, R. A. M., Schellekens, J., Van Dijk, A. I. J. M., & Bruijnzeel, L. A. (2009). Improving curve number based storm runoff estimates using soil moisture proxies. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2(4), 250–259.
- Bellocchi, G., Rivington, M., Donatelli, M., & Matthews, K. (2011). Validation of Biophysical Models : Issues and Methodologies. *Agriculture*, 2(2010), 109–130. doi:10.1051/agro
- Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., et al. (2011). Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. *Remote Sensing of Environment*, 115(12), 3390–3408. doi:10.1016/j.rse.2011.08.003
- Brocca, L., Melone, F., Moramarco, T., & Morbidelli, R. (2009). Antecedent wetness conditions based on ERS scatterometer data. *Journal of Hydrology*, 364(1-2), 73–87.

- Brocca, L., Melone, F., Moramarco, T., & Morbidelli, R. (2010a). Spatial-temporal variability of soil moisture and its estimation across scales. *Water Resources Research*, 46(2), art. no. W02516.
- Brocca, L., Melone, F., Moramarco, T., Wagner, W., & Hasenauer, S. (2010b). ASCAT soil wetness index validation through in situ and modeled soil moisture data in central Italy. *Remote Sensing of Environment*, 114(11), 2745–2755. doi:10.1016/j.rse.2010.06.009
- Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., & Hasenauer, S. (2010c). Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrology and Earth System Sciences Discussions*, 14(10), 1881–1893. doi:10.5194/hess-14-1881-2010
- Brocca, L., Moramarco, T., Melone, F., Wagner, W., Hasenauer, S., & Hahn, S. (2012). Assimilation of Surface- and Root-Zone ASCAT Soil Moisture Products Into Rainfall-Runoff Modelling. *IEEE Transactions on Geoscience and Remote Sensing*, 50(7), 2542 – 2555.
- Calvet, J.-C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., & Pigué, B. (2007). In situ soil moisture observations for the CAL/VAL of SMOS: the SMOSMANIA network. *international Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 1196–1199). Barcelona, Spain, July, 23-28, 2007: IEEE. doi:10.1109/IGARSS.2007.4423019
- Ceballos, A., Scipal, K., Wagner, W., & Martínez-Fernández, J. (2005). Validation of ERS scatterometer-derived soil moisture data in the central part of the Duero Basin, Spain. *Hydrological Processes*, 19(8), 1549–1566.
- Champagne, C., Berg, A. A., McNairn, H., Drewitt, G., & Huffman, T. (2012). Evaluation of soil moisture extremes for agricultural productivity in the Canadian prairies. *Agricultural and Forest Meteorology*, 165, 1–11.
- Cheng, R. T., Burau, J. R., & Gartner, J. W. (1991). Interfacing data analysis and numerical modelling for tidal hydrodynamic phenomena. In B. B. Parker (Ed.), *Tidal hydrodynamics* (pp. 201–219). New York: John Wiley & Sons.
- Corner, B. R., Narayanan, R. M., & Reichenbach, S. E. (2003). Noise estimation in remote sensing imagery using data masking. *International Journal of Remote Sensing*, 24(4), 689–702. doi:10.1080/01431160210164271
- Cosh, M. H., Jackson, T. J., Moran, S., & Bindlish, R. (2008). Temporal persistence and stability of surface soil moisture in a semi-arid watershed. *Remote Sensing of Environment*, 112(2), 304–313. doi:10.1016/j.rse.2007.07.001
- Cosh, M., Jackson, T. J., Bindlish, R., & Prueger, J. H. (2004). Watershed scale temporal and spatial stability of soil moisture and its role in validating satellite estimates. *Remote Sensing of Environment*, 92(4), 427–435.

- Crawford, T. M., Stensrud, D. J., Carlson, T. N., & Capehart, W. J. (2000). Using a Soil Hydrology Model to obtain regionally averaged soil moisture values. *Journal of Hydrometeorology*, 1(4), 353–363.
- Crow, W. T. (2007). A novel method for quantifying value in spaceborne soil moisture retrievals. *Journal of Hydrometeorology*, 8(1), 56–67.
- Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., De Rosnay, P., et al. (2012). Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Reviews of Geophysics*, 50(2).
- Crow, W. T., Huffman, G. J., Bindlish, R., & Jackson, T. J. (2009). Improving satellite-based rainfall accumulation estimates using spaceborne surface soil moisture retrievals. *Journal of Hydrometeorology*, 10(1), 199–212.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597.
- Degree confluence project. (2008). Retrieved September 6, 2012, from <http://confluence.org/>
- Dharssi, I., Bovis, K. J., Macpherson, B., & Jones, C. P. (2011). Operational assimilation of ASCAT surface soil wetness at the Met Office. *Hydrology and Earth System Sciences*, 15(8), 2729–2746.
- Dinku, T., Chidzambwa, S., Ceccato, P., Connor, S. J., & Ropelewski, C. F. (2008). Validation of high-resolution satellite rainfall products over complex terrain. *International Journal of Remote Sensing*, 29(14), 4097–4110.
- Dorigo, W., Scipal, K., Parinussa, R. M., Liu, Y. Y., Wagner, W., de Jeu, R. A. M., & Naeimi, V. (2010). Error characterisation of global active and passive microwave soil moisture datasets. *Hydrology and Earth System Sciences*, 14(12), 2605–2616. doi:10.5194/hess-14-2605-2010
- Dorigo, W., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., et al. (2011). The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences*, 15(5), 1675–1698. doi:10.5194/hess-15-1675-2011
- Doubkova, M., Bartsch, A., Pathe, C., Sabel, D., & Wagner, W. (2009). The medium resolution soil moisture dataset: Overview of the share esa due tiger project. *Proceedings of IEEE International Symposium on Geoscience and Remote Sensing* (Vol. 1, pp. 1116–1119). Cape Town, South Africa, July 2009.
- Doubkova, M., Bartsch, A., & Wagner, W. (2012). SHARE: Soil Moisture for Hydrometeorological Applications. *Earthzine*, (4), 4.

- Doubková, M., Van Dijk, A. I. J. M., Sabel, D., Wagner, W., & Blöschl, G. (2012). Evaluation of the predicted error of the soil moisture retrieval from C-band SAR by comparison against modelled soil moisture estimates over Australia. *Remote Sensing of Environment*, *120*, 188–196.
- Draper, C., Mahfouf, J.-F., Calvet, J.-C., Martin, E., & Wagner, W. (2011). Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France. *Hydrology and Earth System Sciences*, *15*(12), 3829–3841.
- Draper, C. S., Reichle, R. H., De Lannoy, G. J. M., & Liu, Q. (2012). Assimilation of passive and active microwave soil moisture retrievals. *Geophysical Research Letters*, *39*(4), art. No. L04401.
- Draper, C. S., Walker, J. P., Steinle, P. J., de Jeu, R. A. M., & Holmes, T. R. H. (2009). An evaluation of AMSR-E derived soil moisture over Australia. *Remote Sensing of Environment*, *113*(4), 703–710.
- Drusch, M. (2007). Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set. *Journal of Geophysical Research*, *112*(D3), 1–14. doi:10.1029/2006JD007478
- Drusch, M., Wood, E. F., & Gao, H. (2005). Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture. *Geophysical Research Letters*, *32*(15), art. No. L15403.
- Drusch, M., Wood, E., Gao, H., & Thiele, A. (2004). Soil moisture retrieval during the Southern Great Plains Hydrologic Experiment 1999: a comparison between experimental remote sensing data and operational products. *Water Resources Research*, *40*(W02504), doi:10.1029/2003WR002441.
- Entekhabi, D., Reichle, H. R., Koster, D. R., & Crow, T. W. (2010). Performance metrics for soil moisture retrievals and application requirements. *Journal of Hydrometeorology*, *11*(3), 832–840.
- Entin, J. K., Robock, A., Vinnikov, K. Y., Hollinger, S. E., Liu, S., & Namkhai, A. (2000). Temporal and spatial scales of observed soil moisture variations in the extratropics. *Journal of Geophysical Research*, *105*(D9), 11865–11877.
- Famiglietti, J. S., Devereaux, J. A., Laymon, C. A., Tsegaye, T., Houser, P. R., Jackson, T. J., Graham, S. T., et al. (1999). Ground-based investigation of soil moisture variability within remote sensing footprints during the Southern Great Plains 1997 (SGP97) Hydrology Experiment. *Water Resources Research*, *35*(6), 1839–1851. doi:10.1029/1999wr900047
- Famiglietti, J. S., Rudnicki, J., & Rodell, M. (1998). Variability in surface moisture content along a hillslope transect: Rattlesnake Hill, Texas. *Journal of Hydrology*, *210*, 259–281.

- Francois, C., Quesney, A., & Ottlé, C. (2003). Sequential assimilation of ERS-1 SAR data into a coupled land surface-hydrological model using an extended Kalman filter. *Journal of Hydrometeorology*, 4(2), 473–487.
- Gouveia, C., Trigo, R. M., & DaCamara, C. C. (2009). Drought and vegetation stress monitoring in Portugal using satellite data. *Natural Hazards and Earth System Science*, 9(1), 185–195.
- Gruhier, C., De Rosnay, P., Hasenauer, S., Holmes, T., De Jeu, R., Kerr, Y., Mougín, E., et al. (2010). Soil moisture active and passive microwave products: Intercomparison and evaluation over a Sahelian site. *Hydrology and Earth System Sciences*, 14(1), 141–156.
- Hornacek, M., Wagner, W., Sabel, D., Truong, H., Snoeij, P., Hahmann, T., Diedrich, E., et al. (2012). Potential for High Resolution Systematic Global Surface Soil Moisture Retrieval via Change Detection Using Sentinel-1. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 5(4), 1303–1311.
- Isbell, R. F. (1996). *The Australian Soil Classification (1st ed.)*. Colingwood: CSIRO Publishing.
- Iso. (1994). Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions. *Measurement*. International Standards Organization.
- Jackson, T. J., Cosh, M. H., Bindlish, R., Starks, P. J., Bosch, D. D., Seyfried, M., Goodrich, D. C., et al. (2010). Validation of Advanced Microwave Scanning Radiometer Soil Moisture Products. *IEEE Transactions on Geoscience and Remote Sensing*, 48(12), 4256–4272.
- Jackson, T., Member, S., Vine, D. M. L., Hsu, A. Y., Oldak, A., Starks, P. J., Swift, C. T., et al. (1999). Soil Moisture Mapping at Regional Scales Using Microwave Radiometry: The Southern Great Plains Hydrology Experiment, 37(5), 2136–2151.
- Jacobs, J. (2004). SMEX02: Field scale variability, time stability and similarity of soil moisture. *Remote Sensing of Environment*, 92(4), 436–446. doi:10.1016/j.rse.2004.02.017
- Jakeman, A., Letcher, R., & Norton, J. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling Software*, 21(5), 602–614. doi:10.1016/j.envsoft.2006.01.004
- Jeu, R. A. M., Wagner, W., Holmes, T. R. H., Dolman, A. J., Giesen, N. C., & Friesen, J. (2008). Global soil moisture patterns observed by space borne microwave radiometers and scatterometers. *Surveys in Geophysics*, 29(4-5), 399–420.
- Jin, H., & Henderson, B. (2011). Towards a daily soil moisture product based on incomplete time series observations of two satellites. *19th International Congress on Modelling and Simulation* (p. 7). Perth, Australia, December, 12-16, 2011.
- Kerr, Y. H., Waldteufel, P., Wigneron, J. P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M. J., et al. (2010). The SMOS L: New tool for monitoring key elements of the global water cycle. *Proceedings of the IEEE*, 98(5), 666–687.

- Kim, Y., & Wang, G. (2007). Impact of initial soil moisture anomalies on subsequent precipitation over North America in the coupled land-atmosphere model CAM3-CLM3. *Journal of Hydrometeorology*, 8(3), 513–533.
- Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., et al. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, 305(5687), 1138–1140.
- Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K., & Puma, M. J. (2009). On the Nature of Soil Moisture in Land Surface Models. *Journal of Climate*, 22(16), 4322–4335. doi:10.1175/2009JCLI2832.1
- Kuenzer, C., Zhao, D., Scipal, K., Sabel, D., Naeimi, V., Bartalis, Z., Hasenauer, S., et al. (2009). El Niño southern oscillation influences represented in ERS scatterometer-derived soil moisture data. *Applied Geography*, 29(4), 463–477.
- Kumar, S. V., Reichle, R. H., Koster, R. D., Crow, W. T., & Peters-Lidard, C. D. (2009). Role of subsurface physics in the assimilation of surface soil moisture observations. *Journal of Hydrometeorology*, 10(6), 1534–1547.
- Laguardia, G., & Niemeier, S. (2008). On the comparison between the LISFLOOD modelled and the ERS/SCAT derived soil moisture estimates. *Hydrology and Earth System Sciences Discussions*, 5(3), 1227–1265.
- Liu, Y. Y., Evans, J. P., McCabe, M. F., De Jeu, R. A. M., Van Dijk, A. I. J. M., & Su, H. (2010). Influence of cracking clays on satellite estimated and model simulated soil moisture. *Hydrology and Earth System Sciences*, 14(6), 979–990.
- Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., Van Dijk, A. I. J., McCabe, M. F., et al. (2011). Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrology and Earth System Sciences*, 15(2), 425–436.
- Liu, Y. Y., Van Dijk, A. I. J. M., De Jeu, R. A. M., & Holmes, T. R. H. (2009). An analysis of spatiotemporal variations of soil and vegetation moisture from a 29-year satellite-derived data set over mainland Australia. *Water Resources Research*, 45(7), art. no. W07405.
- Mahfouf, J.-F. (2010). Assimilation of satellite-derived soil moisture from ASCAT in a limited-area NWP model. *Quarterly Journal of the Royal Meteorological Society*, 136(648), 784–798.
- Martinez-Fernandez, J., & Ceballos, A. (2003). Temporal Stability of Soil Moisture in a Large-Field Experiment in Spain. *Soil Sci. Soc. Am. J.*, 67, 1647–1656.
- Martinez-Fernandez, J., & Ceballos, A. (2005). Mean soil moisture estimation using temporal stability analysis. *Journal of Hydrology*, 312, 28–38.
- Matgen, P., Heitz, S., Hasenauer, S., Hissler, C., Brocca, L., Hoffmann, L., Wagner, W., et al. (2011). On the potential of MetOp ASCAT-derived soil wetness indices as a new aperture for

hydrological monitoring and prediction : a field evaluation over Luxembourg. *Hydrological Processes*. doi:10.1002/hyp

- McCabe, M. F., Wood, E. F., Wójcik, R., Pan, M., Sheffield, J., Gao, H., & Su, H. (2008). Hydrological consistency using multi-sensor remote sensing data for water and energy cycle studies. *Remote Sensing of Environment*, *112*(2), 430–444.
- McCabe, M. F., Wood, E., & Gao, H. (2005). Initial soil moisture retrievals from AMSR-E: large scale comparisons with SMEX02 field observations over Iowa. *Geophysical Research Letters*, *32*, art. No. L06403. doi:10.1029/2005GL021222
- Meier, P., Frömel, A., & Kinzelbach, W. (2010). Hydrological real-time modelling using remote sensing data. *Hydrology and Earth System Sciences Discussions*, *7*(6), 8809–8835.
- Meier, P., Frömel, A., & Kinzelbach, W. (2011). Hydrological real-time modelling in the Zambezi river basin using satellite-based soil moisture and rainfall data. *Hydrology and Earth System Sciences*, *15*(3), 999–1008. doi:10.5194/hess-15-999-2011
- Metropolis, N., & Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association*, *44*(247), 335–341. doi:10.1137/0106028
- Mielke, P. W., & Berry, K. J. (2007). *Permutation methods: a distance function approach*. Book. New York: Springer.
- Milzow, C., Krogh, P. E., & Bauer-Gottwein, P. (2010). Combining satellite radar altimetry, SAR surface soil moisture and GRACE total storage changes for model calibration and validation in a large ungauged catchment. *Hydrology and Earth System Sciences Discussions*, *7*(6), 9123–9154.
- Miralles, D. G., Crow, W. T., & Cosh, M. H. (2010). Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations. *Journal of Hydrometeorology*, *11*(6), 1423–1429. doi:10.1175/2010JHM1285.1
- Mladenova, I., Lakshmi, V., Walker, J. P., Panciera, R., Wagner, W., & Doubkova, M. (2010). Validation of the ASAR global monitoring mode soil moisture product using the NAFE'05 data set. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(6), 2498–2508.
- Mölders, N., Jankov, M., & Kramm, G. (2005). Application of Gaussian error propagation principles for theoretical assessment of model uncertainty in simulated soil processes caused by thermal and hydraulic parameters. *Journal of Hydrometeorology*, *6*(6), 1045–1062.
- Naeimi, V., Bartalis, Z., & Wagner, W. (2009a). ASCAT soil moisture: An assessment of the data quality and consistency with the ERS scatterometer heritage. *Journal of Hydrometeorology*, *10*(2), 555–563.

- Naeimi, V., Scipal, K., Bartalis, Z., Hasenauer, S., & Wagner, W. (2009b). An improved soil moisture retrieval algorithm for ERS and Metop Scatterometer observations. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7), 1999–2013. doi:art. no. 4814564
- Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, T. K., & Nghiem, S. V. (2003). Soil moisture retrieval from AMSR-E. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2), 215–229.
- Oreskes, N. (1998). Evaluation (not validation) of quantitative models. *Environmental Health Perspectives*, 106(6), 1453–1460.
- Owe, M., De Jeu, R. A. M., & Holmes, T. (2008). Multi-sensor historical climatology of satellite-derived global land surface moisture. *Journal of Geophysical Research*, 113, art. No. F01002. doi:10.1029/2007jf000769
- Panciera, R., Walker, J. P., Kalma, J. D., Kim, E. J., Hacker, J. M., Merlin, O., Berger, M., et al. (2008). The NAFE ' 05 / CoSMOS Data Set : Toward SMOS Soil Moisture Retrieval , Downscaling , and Assimilation. *Cosmos*, 46(3), 736–745. doi:10.1109/TGRS.2007.915403
- Parajka, J., Naeimi, V., Blöschl, G., & Komma, J. (2009). Matching ERS scatterometer based soil moisture patterns with simulations of a conceptual dual layer hydrologic model over Austria. *Hydrology and Earth System Sciences*, 13(2), 259–271.
- Parajka, J., Naeimi, V., Blöschl, G., Wagner, W., Merz, R., & Scipal, K. (2006). Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale. *Hydrology and Earth System Sciences*, 10(3), 353–368. doi:10.5194/hess-10-353-2006
- Parinussa, R. M., Meesters, A. G. C. A., Liu, Y. Y., Dorigo, W., Wagner, W., & De Jeu, R. A. M. (2011). Error estimates for near-real-time satellite soil moisture as derived from the land parameter retrieval model. *IEEE Geoscience and Remote Sensing Letters*, 8(4), 779–783.
- Pathe, C., Wagner, W., Sabel, D., Bartalis, Z., Doubkova, M., & Naeimi, V. (2009a). Scatterometer and ScanSAR soil moisture observations of the contiguous United States. *Proceedings of the IEEE National Radar Conference, IEEE National Radar Conference*. Pasadena, California, USA, April 13-18, 2009.
- Pathe, C., Wagner, W., Sabel, D., Doubkova, M., & Basara, J. (2009b). Using ENVISAT ASAR Global Mode Data for Surface Soil Moisture Retrieval Over Oklahoma, USA. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2), 468–480.
- Pauwels, V. R. N., Hoeben, R., Verhoest, N. E. C., & De Troch, F. P. (2001). The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation. *Journal of Hydrology*, 251(1-2), 88–102.
- Pauwels, V. R. N., Hoeben, R., Verhoest, N. E. C., De Troch, F. P., & Troch, P. A. (2002). Improvement of TOPLATS-based discharge predictions through assimilation of ERS-based remotely sensed soil moisture values. *Hydrological Processes*, 16(5), 995–1013.

- Prisley, S. P., & Mortimer, M. J. (2004). A synthesis of literature on evaluation of models for policy applications, with implications for forest carbon accounting. *Forest Ecology and Management*, *198*(1-3), 89–103. doi:10.1016/j.foreco.2004.03.038
- Rebel, K. T., De Jeu, R. A. M., Ciais, P., Viovy, N., Piao, S. L., Kiely, G., & Dolman, A. J. (2012). A global analysis of soil moisture derived from satellite observations and a land surface model. *Hydrology and Earth System Sciences*, *16*(3), 833–847.
- Reichle, R. H., Crow, W. T., & Keppenne, C. L. (2008a). An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resources Research*, *44*(3), art. No. W03423.
- Reichle, R. H., Crow, W. T., Koster, R. D., Sharif, H. O., & Mahanama, S. P. P. (2008b). Contribution of soil moisture retrievals to land data assimilation products. *Geophysical Research Letters*, *35*(1), art. No. L01404.
- Reichle, R. H., & Koster, R. D. (2004). Bias reduction in short records of satellite soil moisture. *Geophysical Research Letters*, *31*(19), art. No. 19501 1–4.
- Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., et al. (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, *85*(3), 381–394.
- Rüdiger, C., Calvet, J.-C., Gruhier, C., Holmes, T. R. H., De Jeu, R. A. M., & Wagner, W. (2009). An Intercomparison of ERS-Scat and AMSR-E Soil Moisture Observations with Model Simulations over France. *Journal of Hydrometeorology*, *10*(2), 431–447. doi:10.1175/2008JHM997.1
- Sabel, D., Bartalis, Z., Wagner, W., Doubkova, M., & Klein, J.-P. (2012). Development of a Global Backscatter Model in support to the Sentinel-1 mission design. *Remote Sensing of Environment*, *120*, 102–112.
- Sass, G. Z., Wheatley, M., Aldred, D. A., Gould, A. J., & Creed, I. F. (2012). Defining protected area boundaries based on vascular-plant species richness using hydrological information derived from archived satellite imagery. *Biological Conservation*, *147*(1), 143–152.
- Scipal, K., Drusch, M., & Wagner, W. (2008a). Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system. *Advances in Water Resources*, *31*(8), 1101–1112.
- Scipal, K., Holmes, T., De Jeu, R., Naeimi, V., & Wagner, W. (2008b). A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophysical Research Letters*, *35*(24), Art. No.: L24403.
- Settin, T., Botter, G., Rodriguez-Iturbe, I., & Rinaldo, A. (2007). Numerical studies on soil moisture distributions in heterogeneous catchments. *Water Resources Research*, *43*(5), 1–13. doi:10.1029/2006WR005737

- Simmons, A., Uppala, S., Dee, D., & Kobayashi, S. (2007). ERA-Interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF Newsletter*, 110, 25–35.
- Smart, R., Knapp, S., Glover, J., Randall, L., & Barry, S. (2006). *Regional scale land use mapping of Australia: 1992/93, 1993/94, 1996/97, 1998/99, 2000/01 and 2001/02 maps, Version 3*. Canberra, Australia.
- Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modelling and calibration using triple collocation. *Journal of Geophysical Research C: Oceans*, 103(3334), 7755–7766.
- Taylor, J. R. (1997). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. Sausalito: University Science Books.
- Thackway, R., & Creswell, R. I. (1995). An interim biogeographic regionalisation for Australia : a framework for setting priorities in the National Reserves System Cooperative Program Version 4.0 Canberra. *Australian Nature Conservation Agency, Reserve Systems Unit*.
- Vachaud, G., Passerat de Silans, A., Balabanis, P., & Vauclin, M. (1985). Temporal Stability of Spatially Measured Soil Water Probability Density Function. *Soil Sci. Soc. Am. J.*, 49, 822–828.
- Van Dijk, A. I. J. M. (2010). AWRA Technical Report 3. Landscape Model (version 0.5) Technical Description. *WIRADA/CSIRO Water for a Healthy Country Flagship, Canberra*.
- Van Dijk, A. I. J. M., & Marvanek, S. (2010). Derivation of a simplified soil drainage model. AWRA background paper 2010/1. *WIRADA/CSIRO Water for a Healthy Country Flagship, Canberra*.
- Van Dijk, A. I. J. M., & Renzullo, L. J. (2011). Water resource monitoring systems and the role of satellite observations. *Hydrology and Earth System Sciences*, 15(1), 39–55. doi:10.5194/hess-15-39-2011
- Van Dijk, A. I. J. M., & Warren, G. A. (2010). AWRA Technical Report 4. Evaluation Against Observations. *WIRADA/CSIRO Water for a Healthy Country Flagship, Canberra*.
- Van Doninck, J., Peters, J., Lievens, H., De Baets, B., & Verhoest, N. E. C. (2012). Accounting for seasonality in a soil moisture change detection algorithm for ASAR Wide Swath time series. *Hydrology and Earth System Sciences*, 16(3), 773–786.
- Van der Velde, M., Tubiello, F. N., Vrieling, A., & Bouraoui, F. (2011). Impacts of extreme weather on wheat and maize in France: evaluating regional crop simulations against observed data. *Climatic Change*, 113, 751–765.
- Verhoest, N. E. C., Lievens, H., Wagner, W., Álvarez-Mozos, J., Moran, M. S., & Mattia, F. (2008). On the soil roughness parameterization problem in soil moisture retrieval of bare surfaces from synthetic aperture radar. *Sensors*, 8(7), 4213–4248.

- Verstraeten, W. W., Veroustraete, F., Wagner, W., van Roey, T., Heyns, W., Verbeiren, S., & Feyen, J. (2010). Remotely sensed soil moisture integration in an ecosystem carbon flux model. The spatial implication. *Climatic Change*, *103*(1-2), 117–136.
- Vinnikov, K. Y., Robock, A., Qiu, S. A., Entin, J. K., Owe, M., Choudhury, B. J., Hollinger, S. E., et al. (1999). Satellite remote sensing of soil moisture in Illinois, United States. *Journal of Geophysical Research - Atmospheres*, *104*(D4), 4145–4168. doi:10.1029/1998JD200054
- Viterbo, P., & Beljaars, A. C. M. (1995). An improved land surface parameterization scheme in the ECMWF model and its validation. *Journal of Climate*, *8*(11), 2716–2748.
- WMO. (2012). O.S.C.A.R. - Observing Systems Capability Analysis and Review Tool. Retrieved September 6, 1BC, from <http://www.wmo-sat.info/oscar/variables/view/149>
- Wagner, W., Lemoine, G., Borgeaud, M., & Rott, H. (1999a). A Study of Vegetation Cover Effects on ERS Scatterometer Data. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(2 (II)), 938–948.
- Wagner, W., Lemoine, G., & Rott, H. (1999b). A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data. *Remote Sensing of Environment*, *70*(2), 191–207.
- Wagner, W., Naeimi, V., Scipal, K., de Jeu, R., & Martinez-Fernandez, J. (2007). Soil moisture from operational meteorological satellites. *Hydrogeology Journal*, *15*(1), 121–131.
- Wagner, W., Noll, J., Borgeaud, M., & Rott, H. (1999c). Monitoring soil Moisture over the Canadian Prairies with the ERS Scatterometer. *IEEE Trans. Geosci. Rem. Sens.*, *37*(1), 206–216.
- Wagner, W., Sabel, D., Doubkova, M., Bartsch, A., & Pathe, C. (2009). The potential of Sentinel-1 for monitoring soil moisture with a high spatial resolution at global scale. *Earth Observation and Water Cycle Science Symposium* (Vol. ESA Specia, p. 8). Frascati, Italy, 18 - 20 November 2009.
- Wang, J. M., Shiue, J., Engman, E., Lawless, P., Schmutge, T. J., Gould, W., Fuchs, J., et al. (1980). *Remote Measurements of soil moisture by microwave radiometers at BARC test site*. (NASA GFSC, Ed.) (TM-80720th ed., p. 113).
- Water Research Commission. (2012). *South African Water Research Commission. Remote sensing evapotranspiration (SEBS) evaluation using water balance*.
- Western, A. W., & Grayson, R. B. (1998). The Tarrawarra data set: Soil moisture patterns, soil characteristics, and hydrological flux measurements. *Water Resources Research*, *34*(10), 2765–2768. doi:10.1029/98WR01833
- Western, A. W., Grayson, R., & Blöschl, G. (2002). Scaling of soil moisture: a hydrologic perspective. *Ann. Rev. Earth Planet*, *30*, 149–180.

- Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79–82. doi:10.3354/cr030079
- Wong, F. S. (1985). First-order, second-moment methods. *Computers and Structures*, 20(4), 779–791.
- Wooldridge, S. A., Kalma, J. D., & Walker, J. P. (2003). Importance of soil moisture measurements for inferring parameters in hydrologic models of low-yielding ephemeral catchments. *Environmental Modelling and Software*, 18(1), 35–48.
- Young, R., Walker, J., Yeoh, N., Smith, A., Ellett, K., Merlin, O., & and Western, A. (2008). *Soil moisture and meteorological observations from the Murrumbidgee catchment*. (T. U. of M. Department of Civil and Environmental Engineering, Ed.).
- Zribi, M., Paris Anguela, T., Duchemin, B., Lili, Z., Wagner, W., Hasenauer, S., & Chehbouni, A. (2010). Relationship between soil moisture and vegetation in the Kairouan plain region of Tunisia using low spatial resolution satellite data. *Water Resources Research*, 46(6), art. No. W06508.
- Zwieback, S., Scipal, K., Dorigo, W., & Wagner, W. (2012). Structural and statistical properties of the collocation technique for error characterization. *Nonlinear Processes in Geophysics*, 19(1), 69–80.
- de Wit, A. M., & van Diepen, C. A. (2007). Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, 146(1-2), 38–56.