Improvements and Applications of Satellite-Derived Soil Moisture Data for Flood Forecasting

Seokhyeon Kim

A thesis in fulfilment of the requirements for the degree of

Doctor of Philosophy

UNSW
AUSTRALIA

School of Civil and Environmental Engineering

Faculty of Engineering

October 2017

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: **Kim**

First name: **Seokhyeon** Other name/s:

Abbreviation for degree as given in the University calendar: **PhD**

School: **School of Civil and Environmental Engineering** Faculty: **Faculty of Engineering**

Title: Improvements and applications of satellite-derived soil moisture data for flood forecasting

Abstract 350 words maximum: (PLEASE TYPE)

Accurate knowledge of the spatiotemporal behaviour of soil moisture can greatly improve hydrological forecasting capability. While ground-based soil moisture measurements are ideal, they tend to be sparse in space and only available for limited periods. To overcome this, a viable alternative is space-borne microwave remote sensing because of the observational capability it offers for retrieving soil moisture in near real-time at the global scale. However, its direct applications have been limited due to the uncertainty associated, and the coarse spatial resolution these are available at.

Therefore, this thesis aims to use satellite soil moisture products for assessing flood risk by redressing their drawbacks in terms of accuracy and spatial resolution. The research consists of three inter-dependent focal areas; evaluation, improvement and application of the soil moisture products.

For the first objective, this thesis compared two alternate soil moisture products using spatiotemporally identical passive microwave observations but different retrieval algorithms. Complementarity in the performance of the products was identified and accordingly provided the basis for the improvement in soil moisture.

For the second objective, based on the identified complementarity, different formulations of weighted linear combination were proposed as a means of reducing the structural uncertainty associated with each retrieval algorithm. To address the limitation of resulting retrievals existing over coarse grid resolutions, an approach was presented to spatially disaggregate coarse soil moisture by only using a remotely sensed vegetation index product. The method provides a continuous timeseries of disaggregated soil moisture with a persistence structure closer to what is observed.

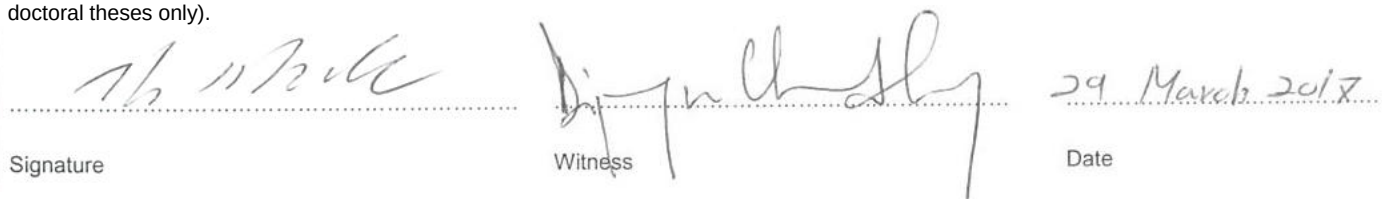
Lastly, for the third objective, a fully remote sensing based flood warning method using readily available soil moisture and rainfall data, open-access topographic and soil data, was developed. This method was applied over a number of anthropogenically unaffected river basins and was shown to have promise for flood warning in ungauged watersheds.

Ongoing and future research will form an integrated pathway for producing an improved soil moisture product available at finer spatial resolution, which can be used for various regional applications, along with using this to provide real-time flood warnings using freely available information to rural and remote communities worldwide.

Declaration relating to disposition of project thesis/dissertation

I hereby grant to the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all property rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstracts International (this is applicable to doctoral theses only).



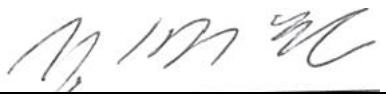
Signature Witness Date

The University recognises that there may be exceptional circumstances requiring restrictions on copying or conditions on use. Requests for restriction for a period of up to 2 years must be made in writing. Requests for a longer period of restriction may be considered in exceptional circumstances and require the approval of the Dean of Graduate Research.

FOR OFFICE USE ONLY Date of completion of requirements for Award:

Originality Statement

'I hereby declare that this thesis is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in this thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged. '

Signed:  (Seokhyeon Kim)

Date: 28 March 2017

Copyright Statement

Signed:  (Seokhyeon Kim)

Date: 28 March 2017

Authenticity Statement

Signed:  (Seokhyeon Kim)

Date: 28 March 2017

Abstract

Accurate knowledge of the spatiotemporal behavior of soil moisture can greatly improve hydrological forecasting capability. While ground-based soil moisture measurements are ideal, they tend to be sparse in space and only available for limited periods. To overcome this, a viable alternative is space-borne microwave remote sensing because of the observational capability it offers for retrieving soil moisture in near real-time at the global scale. However, its direct applications have been limited due to the uncertainty associated, and the coarse spatial resolution these are available at.

Therefore, this thesis aims to use satellite soil moisture products for assessing flood risk by redressing their drawbacks in terms of accuracy and spatial resolution. The research consists of three inter-dependent focal areas; evaluation, improvement and application of the soil moisture products.

For the first objective, this thesis compared two alternate soil moisture products using spatiotemporally identical passive microwave observations but different retrieval algorithms. Complementarity in the performance of the products was identified and accordingly provided the basis for the improvement in soil moisture.

For the second objective, based on the identified complementarity, different formulations of weighted linear combination were proposed as a means of reducing the structural uncertainty associated with each retrieval algorithm. To address the limitation of resulting retrievals existing over coarse grid resolutions, an approach was presented to spatially disaggregate coarse soil moisture by only using a remotely sensed vegetation index product. The method provides a continuous timeseries of disaggregated soil moisture with a persistence structure closer to what is observed.

Lastly, for the third objective, a fully remote sensing based flood warning method using readily available soil moisture and rainfall data, open-access topographic and soil data, was developed. This method was applied over a number of anthropogenically unaffected river basins and was shown to have promise for flood warning in ungauged watersheds.

Ongoing and future research will form an integrated pathway for producing an improved soil moisture product available at finer spatial resolution, which can be used for various regional applications, along with using this to provide real-time flood warnings using freely available information to rural and remote communities worldwide.

Acknowledgements

I am putting a period to this thesis at this moment, the outcome is by no means my own. Four years ago, I came to University of New South Wales (UNSW) as an empty-handed researcher. However, innumerable helping hands have driven me to soundly devote myself to the work without major difficulties and finally reach this finish line. I have so many people to be thankful for.

Firstly, I would like to mention that this work would not be achieved without supports from my academic supervisors. All of their guidance, suggestions, critiques, and scientific opinions have been foundation of this work and improved this study greatly. Specifically, I would like to acknowledge my grateful appreciation to Professor Ashish Sharma who has always provided accurate and timely supports for my academic needs as well as personal difficulties. For my joint supervisor, Dr. Fiona Johnson, her sharp critiques pointing out key points have frequently saved me from falling into confirmation bias on my works. Dr. Yi Y. Liu and Dr. Robert Parinussa, my co-supervisors, have been kind guides to the world of remote sensing which was quite inexperienced for me. Thank you very much once again.

Also, I would like to acknowledge the guidance and advice from the panel members for my four research progress reviews during my Ph.D. program. The appreciation is for Professor Richard Stuetz, Dr. Lucy Marshall, Dr. Hoori Ajami, A/Professor Sivakumar Bellie and Dr. Raj Mehrotra. Particularly, I would like to appreciate A/Professor Sivakumar Bellie who have allocated his time for discussing some hydrology topics related with 'network science' that I wish to work on at some future day.

I would also like to thank administrative supports given by staff in the School of Civil and Environmental Engineering. The gratitude goes to Patricia McLaughlin, Robert Steel, Patricia Karwan, Patrick Vuong and Hugh McMullen (retired). Thanks to their help, my working environment has been maintained at the best. Special appreciation to Robert Steel who was willing to polish the introduction and conclusion of the thesis from third-person point of view.

Last four years in Sydney were great time for me to become a better researcher and to have new friends and colleagues who have shared feelings and knowledge together. Their attentiveness, wit, considerateness and clear-sightedness have been consolation, rest, joy and knowledge to me. I send them my sincere gratitude and hope they will have a bright and successful future. I would like to thank Ahmad Hasan Nury, Ademir Abdala, Anna Yeung, Clare Stephens, Conrad Wasko, Dipayan Choudhury, Ditiro Moalafhi, Eytan Rocheta, Guido Ortega, Ha Nguyen, Mohammad Mahadi Hasan, Ruth Fisher, Sahani Pathiraja, Sajjad Eghdamirad, Suresh Hettiarachchi, Syed Shoaib, Taeho Kang, Thanh Hung Pham, Tomas Beuzen, Jannatun Nahar, Juan Castilla, Jingwan Li, Yating Tang, Yingying Sun and Zaved Khan. Especially, I would like to appreciate Dr. Conrad Wasko for his proofreading the introduction and conclusion sections of the thesis.

Also, I would like to thank Professors Joong Hoon Kim, Chulsang Yoo and Kyungrock Paik in my alma mater, Korea University. I have been always encouraged by their advices and tips whenever I visited. I would like to specially appreciate Professor Kyungrock Paik who firstly introduced me to Professor Ashish Sharma and this academically plentiful grounds. Furthermore, he has always provided me with a place to stay and allowed me to use their resources during my frequent visits to Korea. With this, he also gave me valuable chances to have academic exchanges with students and scholars through a few seminars and a special workshop, Synthesis Workshop on Dynamics of Structure and Functions of Complex Networks at Korea University in 2015.

I also wish to thank many Korean friends I have met in UNSW. Special thanks to Jongho Baek, Dohyung Kim, Ekyu Han, Saerom Jo, Sean Young, Shinyoung Noh and Byungkwan Park who have/had sincerely worked as community members of Korean Postgraduate Student Association at UNSW while I had been served as the president for 2014-2016. And I also really appreciate Lavina Lee, Seungho Lee, Kiho Kwon, Jongphil Lee and Sangho Lee who had greatly helped me to settle in the beginning of life in UNSW.

Financial support for this research was from UNSW Tuition Fee Scholarship linking to a Discovery Project funded by the Australian Research Council. I am very grateful for this support that helped me get this achievement and thank the School of Civil and Environmental Engineering for providing me the opportunity to carry out this research.

My heartfelt thanks to my parents, Cheolsoo Kim and Miyeon Park, for transferring me legacy of sincerity with sound body by which I can come here, and for raising me with love and devotion. I am also appreciative to my sister, Seokji Kim, her husband, Moosang Baek and her children Seojin and Woojin for their scrupulous care for making my research successful.

Most importantly, I want to send my sincere gratitude to my wife, Kyungmin Lee. I am truly thankful for her strong stand beside me during this study. Without her devotion, love, and cares, I could not finalize this work successfully. Also, I would like to give my special thanks to Kyungmin's parents: Takejin Lee and Insook Cho; my sister-in-law: Hyemin Lee. This success is accredited to their wishes and prayers.

Seokhyeon Kim

List of Publications

Journal Papers

Kim, *S.*; Paik, K.; Johnson, F.M.; Sharma, A., Flood Warning in A Watershed Using Readily Available Soil Moisture and Rainfall Data. Submitted

Kim, *S.*; Balakrishnan, B.; Liu, Y.; Johnson, F.M.; Sharma, A. (2017). Spatial Disaggregation of Coarse Soil Moisture Data by Using High Resolution Remotely Sensed Vegetation Products. *IEEE Geoscience and Remote Sensing Letters*

Kim, *S.*, Parinussa, R., Liu, Y., Johnson, F., & Sharma, A. (2016). Merging Alternate Remotely-Sensed Soil Moisture Retrievals Using a Non-Static Model Combination Approach. *Remote Sensing, 8*(6), 518.

Kim, *S.*, Parinussa, R. M., Liu, Y., Johnson, F. M., & Sharma, A. (2015). A framework for combining multiple soil moisture retrievals based on maximizing temporal correlation. *Geophysical Research Letters, 42*(16), 2015GL064981.

Kim, *S.*, Liu, Y., Johnson, F. M., Parinussa, R. M., & Sharma, A. (2015). A global comparison of alternate AMSR2 soil moisture products: Why do they differ? *Remote Sensing of Environment, 161*(0), 43-62.

Conference Presentation

Kim, *S.*; Liu, Y.; Johnson, F. M. & Sharma, A. (2016). A temporal correlation based approach for spatial disaggregation of remotely sensed soil moisture. American Geophysical Union (AGU), San Francisco, USA. (Poster presentation)

Sharma, A. & **Kim, *S.*** (2016). Merging alternate remotely sensed soil moisture by a dynamic linear combination. 7th Statistical Hydrology Workshop, Quebec, Canada (Oral presentation by Ashish Sharma)

Parinussa, R. M.; **Kim, *S.***, Liu, Y.; Johnson, F. M. & Sharma, A. (2015). Comparing and Combining Surface Soil Moisture Products from AMSR2. American Geophysical Union, San Francisco (AGU), CA, USA. (Poster presentation by Robert Parinussa)

Kim, *S.*; Liu, Y.; Johnson, F. M.; Parinussa, R. M. & Sharma, A. (2014). Reducing Structural Uncertainty in AMSR2 Soil Moisture Using a Model Combination Approach. American Geophysical Union (AGU), San Francisco, CA, USA. (Poster presentation)

Kim, *S.*; Liu, Y.; Johnson, F. M.; Parinussa, R. M. & Sharma, A. (2014). Improvement of Soil Moisture Dataset Combining AMSR2 Soil Moisture Products. The Australian Energy and Water Exchange Initiative (OzEWEX), Canberra, ACT, Australia (Poster presentation)

Research-related Activities

Soil Moisture Active Passive Experiment - the 4th campaign (SMAPEx-4)
May 2015 / Yanco, NSW, Australia

Synthesis Workshop on Dynamics of Structure and Functions of Complex Networks
June-July 2015 / Korea University, Seoul, South Korea

Table of Contents

List of Figures

List of Tables

Acronyms and Abbreviations

AMC	Antecedent Moisture Condition
AMSR2	Advanced Microwave Scanning Radiometer 2
AMSR-E	Advanced Microwave Scanning Radiometer for Earth Observing System
API	Antecedent Precipitation Index
AWAP	Australian Water Availability Project
CCI	Climate Change Initiative
CDF	Cumulative Distribution Function
CHIRPS	Climate Hazards group Infrared Precipitation with Stations
CN	Curve Number
CONUS	Contiguous United States
COSMOS	COsmic-ray Soil Moisture Observing System
DCT-PLS	Penalized Least Square regression based on Discrete Cosine Transform
DEM	Digital Elevation Model
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA	ECMWF re-analysis
ESA	European Space Agency
EVI	Enhanced Vegetation Index
FOH	Frequency of Hits
GCOM-W	Global Change Observation Mission
GLOBE DEM	Global Land One-kilometre Base Elevation Digital Elevation Model
HRS	Hydrologic Reference Stations
HSG	Hydrologic Soil Groups
HWSD	Harmonized World Soil Database
ISMN	International Soil Moisture Network
ISW	Index of Soil Wetness
JAXA	Japan Aerospace Exploration Agency
LAI	Leaf Area Index
LC	Land Cover
LPRM	Land Parameter Retrieval Model
LST	Land Surface Temperature
MDB	Murray-Darling basin
MERRA-Land	Modern-Era Retrospective Analysis for Research and Applications Land
MODIS	Moderate Resolution Imaging Spectroradiometer
MPDI	Microwave Polarization Difference Index
MSE	Mean Square Error
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index

NRCS-CN	National Resources Conservation Service
PC	Proportion Correct
PI	Polarization Index
POD	Probability of Detection
QC	Quality Control
RFI	Radio Frequency Interference
RMSE	Root Mean Square Error
RS	Remote Sensing or Remotely Sensed
RZSM	Root Zone Soil Moisture
SM	Soil Moisture
SMA	Soil Moisture Accounting
SMAP	Soil Moisture Active Passive
SMC	Soil Moisture Content
SMMR	Scanning Multichannel Microwave Radiometer
SMOS	Soil Moisture and Ocean Salinity
SRTM	Shuttle Radar Topographic Mission
SZSM	Surface Zone Soil Moisture
TC	Triple Collocation
TRMM	Tropical Rainfall Measuring Mission
TS	Threat Score
USCRN	U.S. Climate Reference Network
USDA	United States Department of Agriculture
UTC	Universal Time Coordinated
VI	Vegetation Index
VOD	Vegetation Optical Depth
VTCI	Vegetation Temperature Condition Index

List of Symbols

e_r	Rough surface emissivity
f_c	Fractional vegetation cover
h	Roughness parameter
I_a	Initial abstraction (mm)
k	Dielectric constant
R	Pearson correlation coefficient
R^2	Coefficient of determination
S	Potential maximum retention (mm)
T_b	Brightness temperature (K)
T_c	Vegetation canopy temperature (K)
T_s	Soil temperature (K)
u	Incidence angle (°)
w	Optimal weight for linear combination
W_c	Vegetation water content (kg/m^2)
Γ	Transmissivity
θ	Soil moisture (m^3/m^3)
λ	Microwave wavelength (m)
ξ	Euclidean distance
ρ	Spearman correlation coefficient
σ	Standard deviation
τ	Vegetation optical depth
ω	Single scattering albedo

Chapter 1 Introduction

Soil moisture retrieval from passive microwave observations has room for improvement and a possibility to be used for flood forecasting in watersheds. This thesis aims to make use of satellite soil moisture for assessing flood risk by redressing their drawbacks in terms of accuracy and spatial resolution. As a starting point, this chapter provides a general introduction to the research field of this thesis, and then problems to be addressed are stated through a number of questions. Finally, objectives and structure of thesis are presented.

1.1. Background

1.1.1. Soil Moisture in Water Cycle

Water, the most plentiful material on Earth, is the principal component of all living things as well as key to human existence and progress (Chow et al., 1988). The water cycle (Figure 1.1), representing the continuous circulation of different phases of water without beginning or end, is one of the important processes in the Earth system. Through the cycle, water vapor is evaporated from land/sea surfaces and condensed into clouds in the atmosphere. The water returns to the surface as precipitation and flows as various types of runoff, infiltrates into the ground, or remains on the surface as water or snow. It is critically important to understand how water circulates in space and time. It is also important to forecast significant anomalies in the water cycle, to ensure sufficient time is available to protect societies against possible adverse impacts. In this regard, hydrological forecasting, especially for extreme events such as floods in a watershed, forms an essential means for preventing socioeconomic damage (Carsell et al., 2004). Such forecasting is based on implementation and calibration of hydrologic models to forecast flow rates and water levels, by using real-time data such as rainfall (*i.e.* external forcing) and soil moisture (*i.e.* antecedent condition).

Of the variables in the hydrologic cycle, this thesis focuses on surface soil moisture, which is defined here as the volumetric amount of water in the uppermost soil layer. It is well known that the surface soil moisture is a key variable in understanding the water balance, influencing

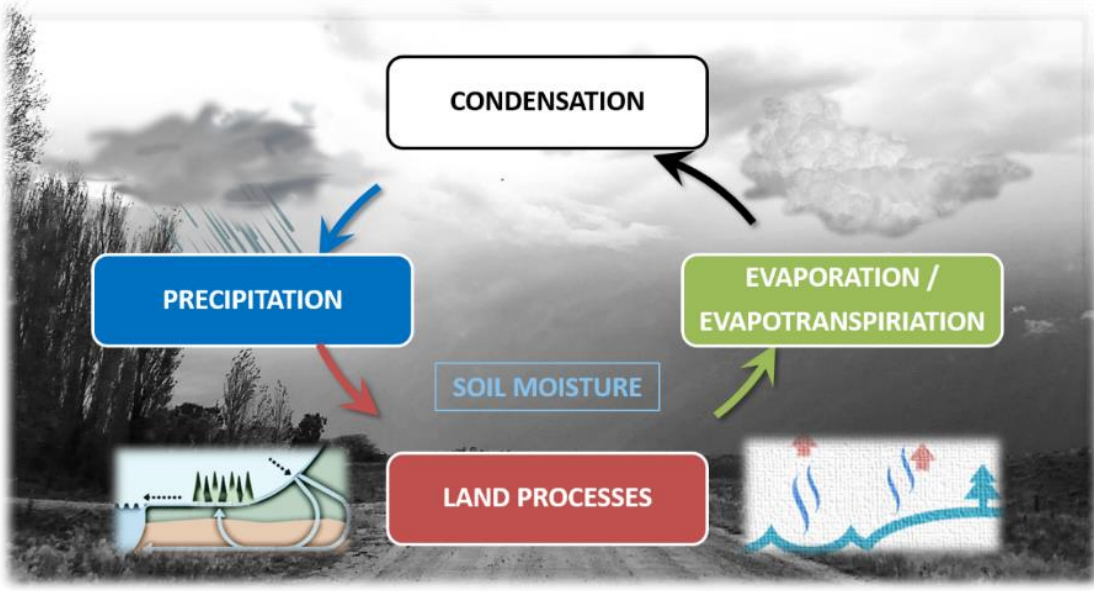


Figure 1.1. Schematic diagram of the water cycle

hydrological and meteorological processes. Soil moisture has persistence, also called “memory” (Koster & Suarez, 2001), which directly interacts with precipitation and evaporation through infiltration, evaporation and runoff (Dong et al., 2007).

Therefore, accurate knowledge of the spatiotemporal behavior of soil moisture can improve the skill of hydrological forecasting (Alvarez-Garreton et al., 2015; Brocca et al., 2010; Jin, 1999; Komma et al., 2008). Accordingly, availability of soil moisture at appropriate spatial and temporal resolutions is important in formulating effective forecasting systems (D'Odorico et al., 2007; Rautiainen et al., 2014; Walker & Houser, 2004).

1.1.2. Measurements of Soil Moisture

While ground-based measurements have been a common source of soil moisture information these tend to be sparsely located in space and only available for limited temporal periods. Therefore, many parts of the world remain ungauged, with even those that are gauged generally poor in spatial distribution (Figure 1.2). To overcome this, a viable alternative to ground measurements is microwave remote sensing using space-borne passive observations (De Jeu & Owe, 2003; Kerr et al., 2001), active microwave observations (Wagner et al., 2003) or a combination (Entekhabi et al., 2010a).

Figure 1.2. Distribution of ground stations for measuring soil moisture, taken from the International Soil Moisture Network (http://ismn.geo.tuwien.ac.at/ismn/) (Dorigo et al., 2011)

Measuring microwave signals is useful in estimating soil moisture in the upper few centimeters of the soil due to a large contrast that exists among dielectric constants of liquid water (~ 80), soil-water mixture (4 ~ 40) and dry soil (~ 4). In turn the emission and scattering of microwave radiation is related to the variation in dielectric constant (Schmugge et al., 1986). Added to this, microwave-based observations are unaffected by cloud, do not require the sun, are less sensitive to roughness conditions of the surface (Wigneron et al., 1998), and can provide information on the water content of the top soil layer even under vegetation coverage (Njoku & Entekhabi, 1996). Accordingly, these observations have provided a unique capability for retrieving soil moisture in near real-time at the global scale (De Jeu et al., 2008; Petropoulos et al., 2015). Microwave radiometry at X- (8-12 GHz), C- (4-8 GHz) and L-band (1-2 GHz) are usually used for the soil moisture estimation (Mohanty et al., 2017). Among them, it has been proven that L-band is physically more suitable than others because it is a protected band and has better penetrating capacity through vegetation canopy, and available from deeper soil layer (Kerr et al., 2001; Walker et al., 2013; Wang & Choudhury, 1981).

The first passive microwave observations resulted from the Scanning Multichannel Microwave Radiometer (SMMR) onboard the Nimbus-7 satellite in 1978, observing brightness

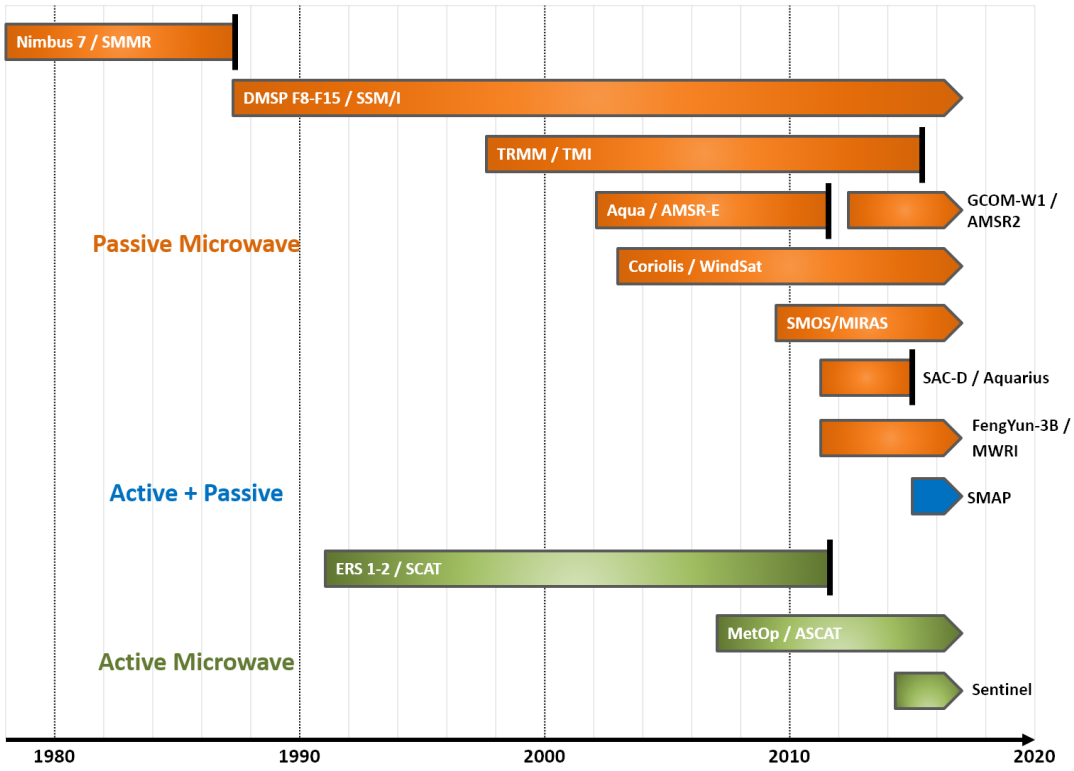


Figure 1.3. Timeline of space-borne coarse resolution radiometers (passive microwave), scatterometers (active microwave) and a combination that can be used for retrieving surface soil moisture on global scale (updated fromDorigo et al. (2010))

temperatures through multiple channels including C-band (6.6 GHz) used for soil moisture retrieval. Data has now been available using several active/passive sensors for the past four decades (Figure 1.3). L-band has been used for Soil Moisture and Ocean Salinity (SMOS), SAC-D (Spanish for Satellite for Scientific Applications-D) and Soil Moisture Active Passive (SMAP), and X- and C-band for others such as Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and Advanced Microwave Scanning Radiometer 2 (AMSR2).

The active and passive microwave-derived products are systemically different in terms of retrieval approaches (Ulaby et al., 1982) as well as their performances by spatial resolution and sensitivity to surface conditions (Petropoulos et al., 2015; Walker et al., 2013). Passive microwave remote sensing is different from active type by the source of electromagnetic energy used. Radiometers (passive) observe the microwave emission from the ground, but active sensors detect distance and backscatter by transmitting and receiving pulses of microwave

energy. Therefore, retrieval of soil moisture from passive observations involves different approaches than those in active sensors, and has distinctive characteristics in terms of spatial resolution and sensitivity to surface conditions. The passive microwave-derived soil moisture generally has a coarser spatial resolution due to large antenna sizes for capturing the typically weak signal from thermal emission but they are less sensitive to land surface conditions such as vegetation and surface roughness. On the contrary, the active microwave-derived products tend to have finer spatial resolutions but are difficult to be interpreted due to their high sensitivity to the surface conditions. In consideration of the complementarity, the SMAP mission combines passive and active microwave observations to retrieve soil moisture at a moderate spatial resolution (Entekhabi et al., 2010a).

1.2. Problem Statement

Despite this observational capability of the microwave-derived soil moisture products, direct applications have been limited due to uncertainty and coarse spatial resolution (>100 km^2). This thesis attempts to address these limitations by focusing on the following research questions with relation to chapters:

- How to use and design model validation algorithms to evaluate satellite-derived soil moisture products? - Chapter 3
- How to improve the soil moisture products based on an improved characterization of the processes that cause uncertainty in measurements? - Chapter 4 and Chapter 5
- How to reduce structural uncertainty in soil moisture products through effective combinational algorithms that tap into the positive aspects of the alternate retrieval algorithms being combined? - Chapter 4 and Chapter 5
- How to disaggregate coarse soil moisture to a finer spatial resolution for regional applications? - Chapter 6

Finally, this thesis presents a new basis for using remotely sensed soil moisture products for assessing flood risk and issuing flood warnings in a watershed where ground observations are not available (Chapter 7).

1.3. Objectives and Structure of Thesis

This thesis aims to make use of satellite-based soil moisture estimates for assessing flood risk by redressing their drawbacks in terms of accuracy and spatial resolution. As outlined in Figure 1.4, this thesis is structured with the three primary research objectives, evaluation, improvement and application of the satellite soil moisture, followed by Chapter 2 including a literature review about the previous studies related with the research objectives.

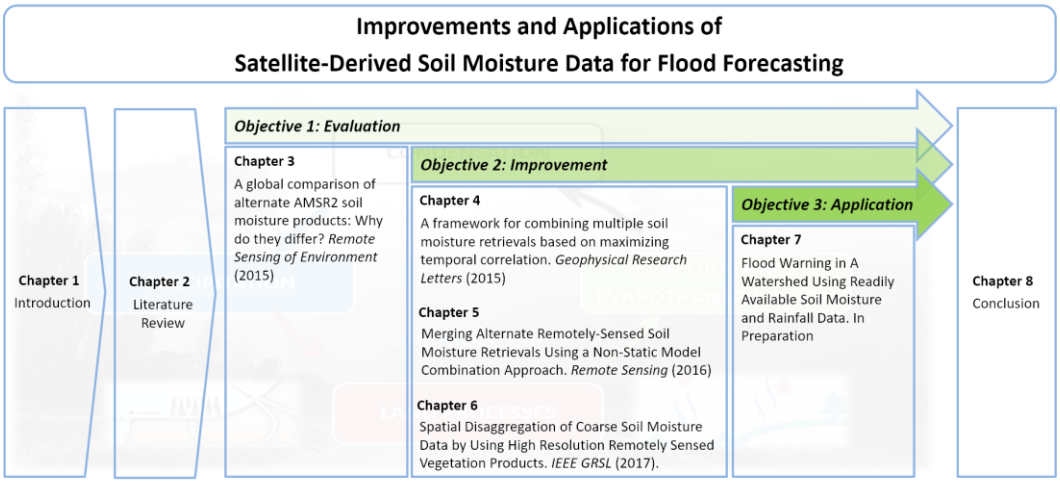


Figure 1.4. Structure of thesis

First, with relation to the first objective for evaluating characteristics of soil moisture products, two alternate soil moisture products from a space-borne radiometer are compared in Chapter 3. Details of their retrieval algorithms have been subsequently used throughout this thesis. The key findings from the comparisons are used as a basis of the second objective for improving the soil moisture product in terms of accuracy using two alternative approaches for reducing model structural uncertainty implicit in each retrieval algorithm (Chapter 4 and Chapter 5). In addition to this, a simple approach is presented in Chapter 6 for spatially disaggregating coarse soil moisture without gaps in disaggregated data. For the third objective, Chapter 7 presents a

methodology for using soil moisture and rainfall data as antecedent conditions and an external forcing respectively to assess flood risk. Finally, Chapter 8 presents the main conclusions and outlines future works.

1.4. Scope and Limitations

In consideration of the systematic differences between the active and passive microwave-derived products, the work scope of this thesis is limited to soil moisture products derived from passive microwave observations alone because of its long-term availability and diversity compared to the active products (Dorigo et al., 2010). Accordingly, all soil moisture products in this thesis are from passive microwave observations hereafter unless otherwise stated.

The main body of this thesis (Chapter 3 to 7) was written on the basis of journal manuscripts published or in preparation for publication. They contain some additional information which was not included in the original publications due to editorial limits. Data, study area and period of each study were selected depending on the best possible availability of key data and/or validation at the time of each study as summarized in Table 1.1.

Table 1.1. Summary of key data, study period and area used for each study.

Objective	Ch.	Key data	Study period	Study area	General reason
Evaluation	3	AMSR2 -JAXA/ LPRM surface soil moisture	1 Aug. 2012 to 31 July 2013	Global	AMSR2 products available from July 2012
Improvement	4	AMSR2 -JAXA/ LPRM surface soil moisture	1 Aug. 2012 to 31 July 2014	Global	-
	5	AMSR2 -JAXA/ LPRM surface soil moisture	1 Jan. 2013 to 31 Dec. 2014	Global	-
	6	ESA CCI surface soil moisture	1 Jan. 2010 to 31 Dec. 2011	CONUS, Spain	For a direct comparison with an existing study
Application	7	AMSR2 -LPRM surface soil moisture / SMAP root zone soil moisture	1 Apr. 2015 to 31 Mar. 2016	Murray-Darling basin, Australia	SMAP products available from Mar. 2015

Chapter 2 Literature Review

This chapter provides a literature review about the previous studies related with the three research objectives - evaluation, improvement and application of surface soil moisture products derived from satellite passive microwave observations.

2.1. Evaluation of Soil Moisture Products

Satellite microwave soil moisture products arise from different instruments and algorithms (Entekhabi et al., 2010a; Fujii et al., 2009; Jackson, 1993; Kerr et al., 2001; Njoku & Kong, 1977; Owe et al., 2008), resulting in differences in their performances (Liu et al., 2012). In other words, uncertainties exist in the products due to many complex factors affecting the soil moisture retrievals, including imperfect retrieval algorithms and measurement errors (Parinussa et al., 2011c). The retrieval algorithms link the remotely sensed microwave observations to soil moisture values and generally introduce an uncertainty due to methods and model parameterization used in the retrievals algorithms. The measurement error is defined as the difference between a measured value of soil moisture and its true value, and consists of random and systematic errors. The random error is caused by any random factors such as accuracy limits of the measuring instrument. The systematic error results from any systematic factors like incorrect calibration of the measurement instrument.

Previous studies have reported necessity of accurately estimated soil moisture products for various applications including weather prediction (Drusch, 2007) and flood forecasting (Brocca et al., 2010). Therefore, it is essential when using or improving soil moisture products to identify their error characteristics in time and space through an evaluation or validation procedure. Traditionally, the quality of a soil moisture product is ascertained through comparison with ground based observations which serve as the assumed truth, generally using metrics such as bias, root mean square error (RMSE), standard error, and correlation coefficient (Draper et al., 2009; Entekhabi et al., 2010b; Yilmaz & Crow, 2013). However, there are systemic differences between the assumed truth and the satellite soil moisture products. The differences come from

various sources including the spatial representation of the ground based (point-scale) observations (Crow et al., 2012), differences in measurement depth of the *in situ* sensor and the microwave emission, and uncertainties in the (limited) parameterization of the land surface roughness and vegetation required for the retrieval of soil moisture values (Parinussa et al., 2011c). Preprocessing of remotely sensed soil moisture is commonly applied for the majority of applications. This is a prerequisite to minimize systematic differences between that would otherwise result in (Reichle & Koster, 2004). Preprocessing steps include restricting the maximum measurement depth of the ground measurements being used, applying a quality control procedure, and/or to checking the area representativeness of stations by considering independent data (Dorigo et al., 2014; Yee et al., 2016). Such preprocessing is often achieved by removing the climatology and/or scaling to match the unique model soil moisture climatology (Reichle & Koster, 2004).

Alternate validation techniques are sometimes needed to make up for the limitations of the verification using ground observations due to the point scale and sparse distribution. A simple method for validation is to compare soil moisture products with other independent products for which the spatial resolution is identical (*e.g.* re-analysis products). Other large scale verification techniques are the triple collocation (TC) and R-value methods (Parinussa et al., 2011a). The former, TC, is a statistical tool for estimating the variance of the random error term from three collocated data sets (e.g. soil moisture) without a high-quality reference data set. It is based on an assumption that error structures of the three data sets are independent from each other if their measurement techniques and retrieval approaches are different. Since (Stoffelen, 1998) used TC for evaluating error characteristics for wind vector data, TC has been frequently used for evaluating soil moisture data sets from various sources such as passive, active microwave and models (Dorigo et al., 2014; Gruber et al., 2016; Scipal et al., 2010; Scipal et al., 2008). Crow (2007) first introduced the R-value method for evaluating remotely sensed soil moisture products without using ground-based soil moisture observations. The evaluation is implemented by calculating the correlation coefficient, called the R-value, between antecedent

rainfall errors and analysis increments by the Kalman filter-based assimilation of remotely sensed soil moisture data into an Antecedent Precipitation Index (API) based model. The R-values are typically ranged from 0 to 0.7, and a higher value indicates a higher-quality soil moisture data and increased efficiency in filtering errors in the API predictions which are from random errors in rainfall data for generating API.

2.2. Improvements of Soil Moisture Products

2.2.1. Reduction of Uncertainty

Many studies have investigated reducing uncertainties of soil moisture products through mathematical methods, and fusion or combination of existing products. For example, Du (2012) applied Fourier analysis to a time series of soil moisture for detecting high-frequency components that reflect soil moisture changes, and then generated long-term datasets with a better accuracy. Su et al. (2013a) reduced noise in soil moisture time series using a semi-empirical model of the observed power spectral density. Kornelsen and Coulibaly (2015) tried to reduce multiplicative bias of satellite soil moisture by a few methods such as cumulative distribution function (CDF) matching, linear rescaling and copulas. For the latter, Liu et al. (2011b) developed improved soil moisture data by blending passive and active microwave based products through CDF matching. Recently,Tomer et al. (2016) proposed an algorithm to merge strengths of active (high spatial resolution) and passive (high temporal resolution) soil moisture products. The algorithm sequentially converts a temporal differential of the passive soil moisture to a soil moisture value at the finer spatial (active) and temporal (passive) resolutions by using information on spatiotemporal changes of the active soil moisture and spatial heterogeneity.

Many of the previous studies have focused on evaluating the performance of soil moisture products with or without efforts to improve the soil moisture products. However, this thesis explores reasons for differences and similarities in the performance of soil moisture products from the same instrument but, using different retrieval algorithms, uses complementarity in their performances as a basis for improvement. Every soil moisture product has relative

strengths and weaknesses compared to others, and therefore a product can be supported by others and can be made more reliable using this approach. For this, Chapter 3 first investigates key differences in soil moisture retrieval algorithms as a means to understand why different retrievals from the same satellite sensor can lead to major differences in the derived values. These differences are then used to develop new alternatives for combining alternate soil moisture fields representing the same domain, as a means of deriving a better product that is stable and consistent. Two such combination algorithms are developed, as detailed in Chapters 4 and 5.

2.2.2. Spatial Disaggregation

The coarse resolution of soil moisture products is another key limitation for local applications such as hydrologic modelling and agricultural monitoring. To address this limitation, several attempts have been made to disaggregate the soil moisture fields to finer spatial resolutions using statistical methods such as kriging or fractal interpolation (Kim & Barros, 2002; Loew & Mauser, 2008), along with the use of complementary information such as topographic/soil properties (Pellenq et al., 2003), land surface temperature, vegetation and soil evaporative efficiency derived from high spatial resolution optical/thermal infrared sensors, land surface model and radar (Fang & Lakshmi, 2014; Merlin et al., 2008; Narayan et al., 2004; Peng et al.,

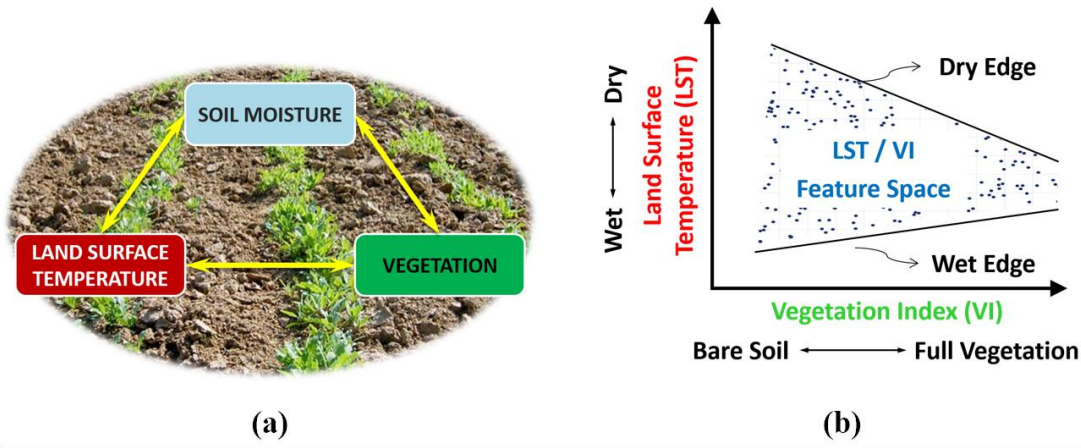


Figure 2.1. Schematic diagrams representing (a) relationship among soil moisture, land surface temperature and vegetation, and (b) feature space constructed by land surface temperature and vegetation index

2015; Piles et al., 2011). These methods have developed relationships among microwave-derived soil moisture at a course spatial resolution, and land surface temperature (LST) and vegetation index (VI) at finer spatial resolutions (Figure 2.1a). For example, Fang et al. (2013) suggested a method which uses look-up tables to relate the datasets at the coarse spatial scale to the daily temperature difference at a finer spatial scale to give the daily average soil moisture at the finer scale. Peng et al. (2015) proposed a soil moisture disaggregation method by using the vegetation temperature condition index (VTCI) as a soil moisture proxy. When plotting paired LST and VI over an area for a period, the data generally forms a triangular or trapezoidal shape which is called as the LST/VI feature space (Figure 2.1b) where a degree of surface wetness is stipulated by a pair of LST and VI at a location within the area, called VTCI (Wan et al., 2004).

For the latter set of approaches that rely on complementary surface information, one of key issues to overcome has been the cloud masks on optical/thermal infrared sensor derived data. Such masks tend to cause large gaps in the disaggregated soil moisture and therefore decrease the usability of the soil moisture dataset. To overcome this limitation, this thesis presents a new disaggregation approach for soil moisture that offers an alternative to the gap-contained data using the algorithms described above. More details on this approach are presented in Chapter 6.

2.3. Soil Moisture for Flood Warning

Soil moisture and rainfall play key roles for runoff generation during flooding. Soil moisture serves as the antecedent conditions to a flood event modulating rainfall, the main external forcing causing flood in any watershed. Therefore, the combined effect of the spatiotemporal dynamics of soil moisture and rainfall can explain most of variability in runoff prediction (Durán-Barroso et al., 2016).

In past decades, passive microwave based soil moisture products have been widely used for flood forecasting or monitoring. For example, Jin (1999) proposed a simple flooding index using passive microwave observations and suggested regional thresholds of the index for flood monitoring. Lacava et al. (2005) presented a passive microwave based method for improved

monitoring of soil wetness variations in relation to a flood in Hungary in 2000. Rao and Sharma (2006) used thresholds of a soil moisture product for estimating the flood affected area in watersheds in India. Along with these, many approaches have assimilated soil moisture products into distributed rainfall-runoff models for improving flood forecasting (Alvarez-Garreton et al., 2014; Brocca et al., 2010; Chen et al., 2011; Houser et al., 1998; Komma et al., 2008).

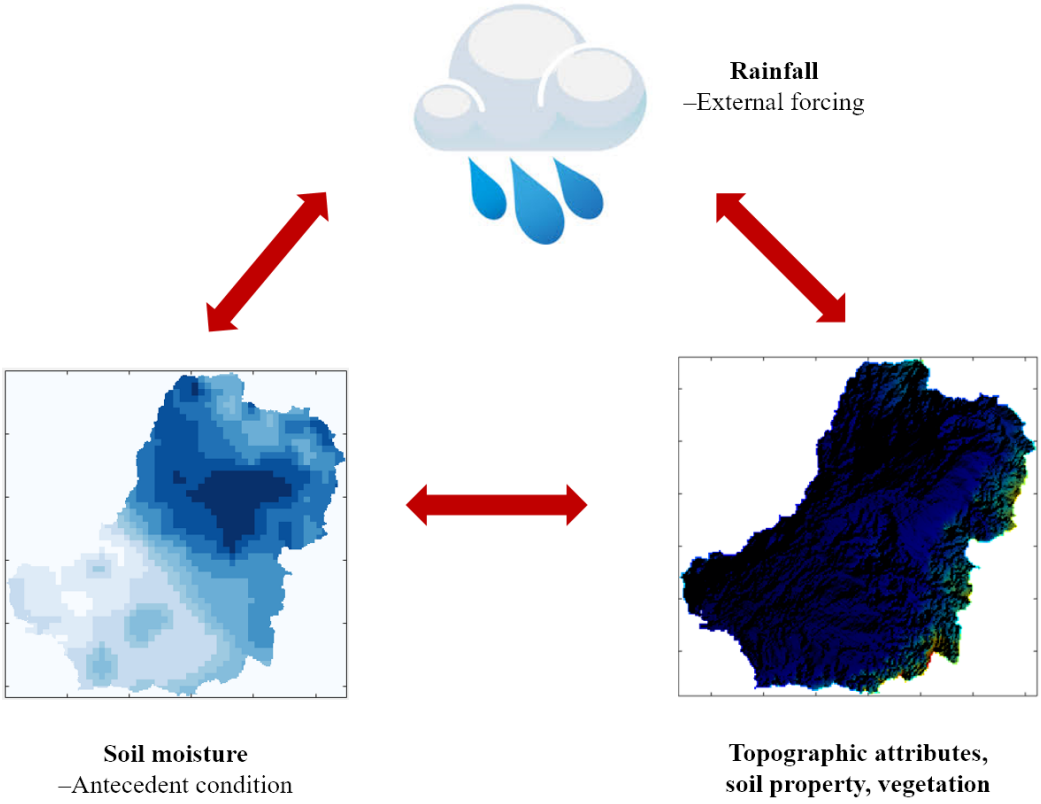


Figure 2.2. Three key sources of information for estimating flood risk in this thesis

In this thesis, a novel approach is presented for identifying flood risk in a watershed using readily available soil moisture and rainfall data when ground observations are not available (Chapter 7). To enable flood warnings to be issued, three key sources of information are considered (Figure 2.2); 1) soil moisture as antecedent conditions, 2) rainfall as an external forcing, and 3) topographic attributes defining where the two factors interact. As assessment of flood risk is performed by considering the amount of incident rainfall and its location within the watershed, how the incident rainfall interacts with antecedent soil moisture is then measured

through satellite remote sensing to generate, transfer and accumulate water on the ground, leading to flood risk.

Chapter 3 Complementarity of Alternate Soil Moisture Products

This chapter assessed two remotely sensed soil moisture products from the Advanced Microwave Scanning Radiometer 2 (AMSR2), a sensor onboard the Global Change Observation Mission 1 – Water (GCOM-W1) that was launched in May 2012. The soil moisture products were retrieved by the Japan Aerospace Exploration Agency (JAXA) algorithm and the Land Parameter Retrieval Model (LPRM) developed by the VU University Amsterdam, in collaboration with the National Aeronautics and Space Administration (NASA). The two products were compared at the global scale. In addition, the products were evaluated against field measurements from the 47 the COsmic-ray Soil Moisture Observing System (COSMOS) network stations in the United States (36 stations), Australia (7 stations), Europe (2 stations) and Africa (2 stations).

After examining the retrieval algorithms, it was hypothesized that four factors, namely, physical surface temperatures, surface roughness, vegetation and ground soil wetness conditions, may affect the quality of soil moisture retrievals. From inter-comparisons at the global scale, correlations of the two products highlight differences in the representation of the seasonal cycle of soil moisture, with negative correlations found for several regions. Correlations of the anomaly timeseries were generally strong (R>0.6) because of soil moisture sensitivity to external meteorological forcing and possibly also random noise in the satellite observations. Due to the inherent differences in spatial coverage and measurement scale of the COSMOS and satellite data, the comparisons in terms of correlation coefficients were found to be the most reliable. Both products showed rapid decreases in correlation coefficients under low mean surface temperature (< 290K), high mean Enhanced Vegetation Index (EVI) (> 0.3) and highly wetted conditions. These findings were further supported by the bias and RMSE estimates which showed that JAXA has relatively better performance under dry conditions whilst the bias and RMSE of LPRM were generally smaller than JAXA, when considered against the four variables. These results provide information on appropriate parametrizations and model selection for the

retrieval algorithms and a future research direction to improve the quality by leveraging the strengths of the JAXA and LPRM algorithms. With these, when a multi-year dataset is available, there will be more confidence in defining the seasonal cycle and the data can be decomposed to identify the anomalies where the bias is not relevant.

This chapter is an edited version of: Kim, *S.*; Liu, Y.Y.; Johnson, F.M.; Parinussa, R.M.; Sharma, A., A global comparison of alternate AMSR2 soil moisture products: Why do they differ*? Remote Sensing of Environment* 2015, 161 (0), 43-62.

3.1. Introduction

The Advanced Microwave Scanning Radiometer 2 (AMSR2) is a passive microwave sensor onboard the Global Change Observation Mission 1 – Water (GCOM-W1) satellite that was launched by the Japan Aerospace Exploration Agency (JAXA) in May 2012. AMSR2 is the successor of the successful Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E, May 2002 – October 2011), which was a passive microwave sensor widely used for the retrieval of soil moisture (Koike et al., 2004; Njoku et al., 2003; Paloscia et al., 2006). AMSR-E provided a consistent and continuous dataset for almost a decade. AMSR2 provides improved spatial resolution due to its larger reflector compared to its predecessor. Moreover, it has an additional 7.3 GHz channel that was developed for Radio Frequency Interference (RFI) mitigation and an improved calibration system (Imaoka et al., 2010). The JAXA developed a soil moisture retrieval algorithm (Fujii et al., 2009) and has made available its soil moisture product from AMSR2 since July 2012. Another algorithm, the Land Parameter Retrieval Model (LPRM), developed by the VU University Amsterdam in collaboration with the National Aeronautics and Space Administration (Owe et al., 2001), has been applied to AMSR2 passive microwave observations to derive a soil moisture product (Parinussa et al., 2014). The LPRM also produces land surface temperature (LST) and vegetation optical depth (VOD) which is an indicator of the total vegetation water content of above-ground biomass (Liu et al., 2011a; Liu et al., 2013).

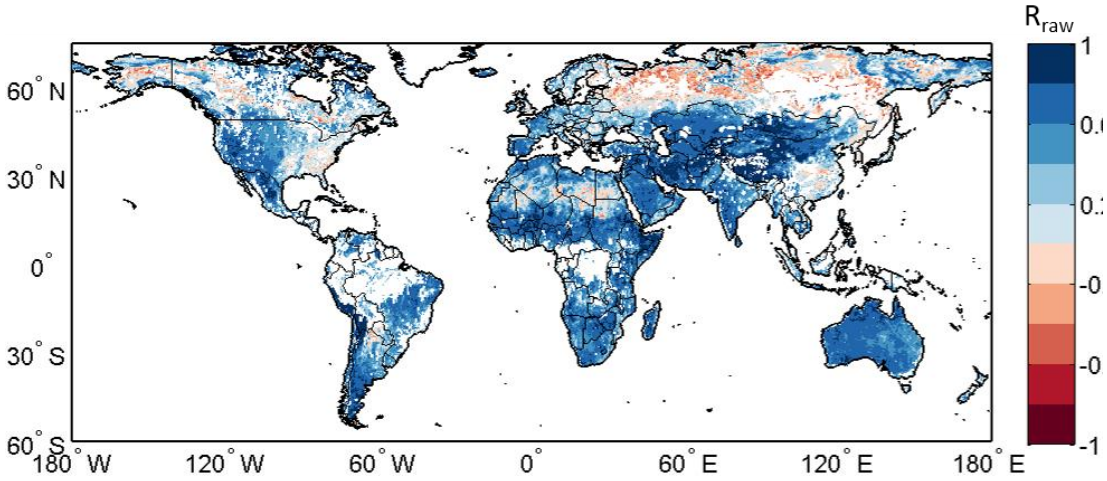


Figure 3.1. Spatial distribution of Pearson correlation coefficients (R) between the daily JAXA and LPRM soil moisture products for the period August 2012 through July 2013. The soil moisture products are from the descending overpasses of 10.7 GHz (X-band); the regions with dense forests are masked out

Given that the underlying microwave emission observations are the same for the LPRM and JAXA retrieval algorithms, an exciting opportunity is now available to assess the effects of algorithms on the resulting AMSR2 soil moisture products. A preliminary correlation analysis between the daily LPRM and JAXA soil moisture products, for one year at the global scale shows that there are a number of similarities and differences between these products (Figure 3.1).

I seek to answer the obvious question that arises from Figure 3.1 – what are the reasons for their similarities and differences? Previous studies have compared individual remotely sensed soil moisture product to ground-based measurements at a regional (Brocca et al., 2011; Draper et al., 2009; Gruhier et al., 2010; Yee et al., 2013) or global scale (Al-Yaari et al., 2014; Albergel et al., 2012b; Albergel et al., 2013). These performance verifications have generally been limited to examining differences with little attention to identifying the causes for these differences. The objectives of this chapter are twofold. First is to provide a guidance for users of these products by performing a comprehensive comparison between these two AMSR2 soil moisture products as well as with ground-based measurements for a single year. The second aim is to identify the reasons for the differences and similarities between the two remotely sensed products for better understanding and further improvements of the algorithms. While the JAXA and LPRM

algorithms start from the simple radiative transfer model (Mo et al., 1982), they use different ways to define the physical surface temperature, surface roughness, vegetation conditions and dielectric constants. The influence of these four factors will be investigated in detail.

More details of the JAXA and LPRM AMSR2 products, ground soil moisture measurements and statistical methods used in this study are described in 3.2. The results of the spatiotemporal comparisons are presented in 3.3 with an emphasis on the regions or time periods where these two products differ. The last sections discuss the results and suggestions for future research directions.

3.2. Data and Methods

Datasets used over the study period 1 August 2012 through 31 July 2013 are listed including: 1) two AMSR2 soil moisture products (the JAXA and LPRM algorithms), 2) field soil moisture measurements collected from the COSMOS and USCRN networks, 3) enhanced vegetation index (EVI) from the Moderate Resolution Imaging Spectroradiometer (MODIS) (Justice et al., 1998) to represent global green vegetation density, 4) topographic data from (the Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model (DEM), version 1.0 (GLOBE-Task-Team & Others, 1999) used as an indicator of the large-scale surface roughness, 5) soil temperature level 1 data from ERA-Interim produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Dee et al., 2011) to mask out data with freezing condition and 6) precipitation data from Tropical Rainfall Measuring Mission (TRMM, 3B42 V7) (Huffman & Bolvin, 2014) for qualitative comparisons with the temporal dynamics of the satellite soil moisture products.

3.2.1 Satellite-Based Soil Moisture Products

The LPRM product provides AMSR2 soil moisture retrievals for the 6.9, 7.3 and 10.7 GHz channels as this algorithm can be applied to low microwave frequencies (<20 GHz), whereas the JAXA product is only available for 10.7 GHz. Although the soil moisture retrievals from lower microwave frequencies are expected to be more accurate (Parinussa et al., 2011c) this study

Table 3.1. Details of satellite-derived soil moisture, field soil moisture measurements and ancillary data used in this study.

Data	Product name	Temporal resolution	Spatial resolution	Units
AMSR2 -JAXA	Level 3 geophysical parameter SMC	Daily	0.25°	m^3/m^3
AMSR2 -LPRM	Level 3 soil moisture	Daily	0.25°	m^3/m^3
COSMOS	Level 3 SM12H	Hourly	A few hundred metres	m^3/m^3
USCRN	SOIL_MOISTURE_5_DAILY (depth 5cm)	Hourly	Point measurement	m^3/m^3
MODIS EVI	MOD13C2	Monthly	0.25° (Resampled)	-
Digital Elevation Model	Global Land One-kilometre Base Elevation (GLOBE) Digital Elevation Model, Version 1.0	-	0.25° (Resampled)	-
Soil Temperature	European Centre for Medium-Range Weather Forecasts (ECMWF) re-analysis (ERA) interim, Soil Temperature Level 1	6-hourly	0.25°	K
Precipitation	Tropical Rainfall Measuring Mission (TRMM), 3-hourly product 3B42 (V7)	Daily (Aggregated)	0.25°	mm/day

focuses on the soil moisture retrievals from 10.7 GHz at a 0.25° global grid that is available from both retrieval algorithms. Another benefit is that RFI issues are less severe for the 10.7 GHz soil moisture retrievals than the 6.9 GHz over the USA (Njoku et al., 2005) where most of ground stations are distributed for this study (Figure 3.3). Differences have been noted between ascending and descending overpass soil moisture retrievals as the geophysical conditions are different at day- and night-time (Draper et al., 2009; Fujii et al., 2009; Owe et al., 2001). As night-time is more favorable to retrieve soil moisture because of the equilibrium thermal conditions vegetation and near-surface soil (De Jeu et al., 2008), only soil moisture products from

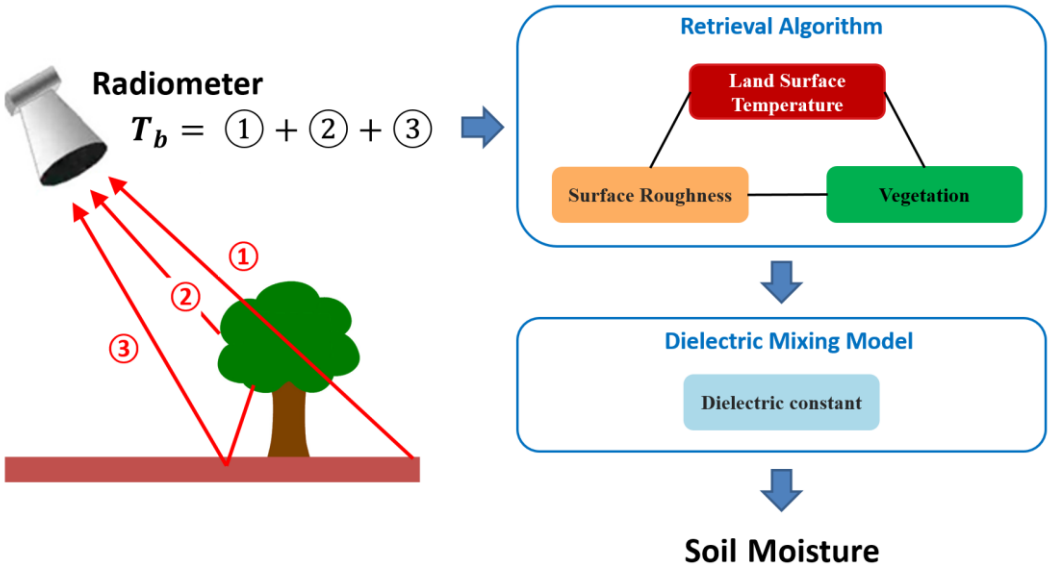


Figure 3.2. Simple radiative transfer model for retrieving soil moisture using passive microwaves consisting of three components

descending overpasses are used in the main text, while results from the ascending overpass data are presented in the appendix of this chapter.

The simple radiative transfer model (Mo et al., 1982) is the basis for both the JAXA and LPRM algorithms to retrieve soil moisture from passive microwave observations. Ignoring attenuation effects by the atmosphere, the microwave brightness temperature (T_b), observed at the satellite, consists of three components (Figure 3.2); 1) the radiation from the soil attenuated by the overlaying vegetation layer, 2) the upward radiation from the vegetation layer and 3) the downward radiation from the vegetation, reflected upward by the soil and again weakened by the vegetation canopy (Owe et al., 2001). Each term can be expressed in sequence with

$$\begin{aligned} T_{b(p)} = {} & T_s e_{r(p)} \Gamma_{(p)} + (1 - \omega_{(p)}) T_c (1 - \Gamma_{(p)}) \\ & + (1 - e_{r(p)})(1 - \omega_{(p)}) T_c (1 - \Gamma_{(p)}) \Gamma_{(p)} \end{aligned} \qquad \text{Eq. 3.1}$$

Where p represents a polarization (horizontal or vertical); T_s and T_c represent physical temperatures of the soil and vegetation canopy respectively (K); ω is the single scattering albedo; e_r is the rough surface emissivity; Γ is the transmissivity defined with vegetation optical depth (τ).

Both algorithms use brightness temperature at 10.7 GHz and 36.5 GHz channels from AMSR2, but they have differences in estimating physical surface temperature, surface roughness, vegetation dynamics and linking dielectric constants to soil moisture by dielectric mixing models by which a soil moisture retrieval algorithm based on the simple radiative transfer model can be differentiated from each other (Neelam & Mohanty, 2015).

3.2.1.1. AMSR2 – JAXA

The current JAXA algorithm (Fujii et al., 2009) is an improved version of the predecessor which was employed as the JAXA standard algorithm (Koike et al., 2004). Considering the simple radiative transfer model in Eq. 3.1, in the JAXA implementation, both T_s and T_c are fixed as 293 K (20° C) (Koike, 2013) and the emissivity is defined by using both polarizations (Wang & Choudhury, 1981) by

$$e_r = \left[1 - \left\{(1-Q)\cdot R_p + Q\cdot R_q \cdot e^{-h\cdot cos^2 u}\right\}\right] \qquad \text{Eq. 3.2}$$

where *Q* represents the polarization mixing ratio; *R* is the reflectivity expressed by the Fresnel equations in terms of the incidence angle (*u*) and the dielectric constant (*k*); *q* is the opposite polarization from *p*. The dimensionless roughness parameter (*h*) is calculated by using the variance (σ^2) of the height distribution of the land surface and the microwave wavelength (λ) as Eq. 3.3 (Choudhury et al., 1979)

$$\mathrm{h} = 4\cdot\sigma^2\cdot(2\pi/\lambda) \qquad \text{Eq. 3.3}$$

$e^{-h\cdot cos^2 u}$ in Eq. 3.2 is defined as *H* with *Q* and *H* for 10.65 GHz assumed to be constants (Koike, 2013) based on a site experiment in Mongolia, which minimized differences between satellite-derived and *in situ* values of two indices, the Polarization Index (PI) (Paloscia & Pampaloni, 1988) and the Index of Soil Moisture (ISW) (Koike et al., 1996).

To consider the vegetation, the Jackson and Schmugge model (Jackson & Schmugge, 1991) is applied for the relationship between τ and the vegetation water content (W_c) with the vegetation parameter *b*.

$$\tau = b \cdot W_c \quad \text{Eq. 3.4}$$

The values for b and ω were obtained from a ground-based vegetation observation experiment by a Ground-Based Microwave Radiometer (GBMR). One of the main improvements of the new JAXA algorithm is the use of the fractional vegetation coverage (f_c) based on the MODIS Normalized Difference Vegetation Index (NDVI). In this new algorithm, the total brightness temperature observed by the satellite is expressed as the sum of bare soil and vegetation brightness temperature proportional to f_c calculated using NDVI.

With respect to the dielectric constants, which affect the emissivity used in the Fresnel equations, the Dobson model (Dobson et al., 1985) is used based on a site experiment using GBMR (Fujii et al., 2000). Generally, emissivity tends to decrease as soil moisture increases (Gadani et al., 2011). However, it was found that under relatively dry conditions (soil moisture less than 0.1 m^3/m^3) the apparent emissivity increased when the soil moisture increased. This was thought to be due to emissivity extinction and lower physical temperature in deeper penetration depths under relative dry conditions. The four-stream fast model (Liu, 1998) is adopted in the JAXA algorithm to reflect the results of the GBMR experiment. It is expected that this will lead to higher values of soil moisture in the JAXA product under dry conditions than would otherwise be the case.

A database of lookup tables depending on f_c was thus established to obtain the W_c and the soil moisture corresponding to PI and ISW calculated from observed brightness temperatures at 10 GHz (V, H) and 36 GHz (V) (Koike, 2013). The retrieved soil moisture ranges from 0 to 0.6 m^3/m^3.

3.2.1.2. AMSR2 - LPRM

The LPRM algorithm is applicable for low microwave frequencies (<20 GHz). The soil moisture and VOD which is an indicator of the total vegetation water content of above-ground biomass (Jackson & Schmugge, 1991) are retrieved simultaneously from the dual-polarization brightness temperatures. The vertically polarized Ka-band (37 GHz, T_b(37GHz[V])) is used to estimate the physical temperature of the soil surface (T_s) following a linear relationship between T_b(37 GHz[V]

and T_s (Parinussa et al., 2011b). The LPRM assumes that T_c is equal to soil surface temperature T_s, therefore the LPRM soil moisture products from the night-time overpasses are more reliable as the minimal temperature gradients at night-time are more favorable for the retrievals (De Jeu & Owe, 2003).

The surface roughness parameter (*h*) and the polarization mixing ratio (*Q*) are estimated as constants but take different values from JAXA (Njoku & Li, 1999) and the single scattering albedo is set to a constant of 0.06 as an average value from a number of related studies (Owe et al., 2001).

The transmissivity (Γ) is defined using τ and the incidence angle (*u*) and expressed as;

$$\Gamma = e^{-\tau/\cos u} \qquad \text{Eq. 3.5}$$

where τ is expressed in terms of *k*, *u* and the Microwave Polarization Difference Index (MPDI) (Meesters et al., 2005), with MPDI is defined with microwave brightness temperatures at both polarizations, $T_b[V]$ and $T_b[H]$ as shown in Eq. 3.6 (Becker & Choudhury, 1988).

$$\mathrm{MDPI} = (\mathrm{T_b[V]} - \mathrm{T_b[H]})/(\mathrm{T_b[V]} + \mathrm{T_b[H]}) \qquad \text{Eq. 3.6}$$

All parameters in Eq. 3.1 can be expressed with the assumed constants or according to *k*, as the emissivity ($e_r(p)$) at both polarizations also can be defined in terms of the dielectric constant (*k*) using the Fresnel equation. In this case, *k* is obtained by solving Eq. 3.1 through a nonlinear iterative procedure. Consequently, the soil moisture is retrieved from *k* using the Wang-Schmugge dielectric model (Wang & Schmugge, 1980).

3.2.1.3. Key Differences Between JAXA And LPRM Algorithms

Reviewing the above descriptions, the four key differences between the two algorithms in relation to the parameterization models are summarized in Table 3.2 and their implications explained below, which are expected to affect the accuracy of the soil moisture products.

Table 3.2. Summary of main differences between the JAXA and LPRM algorithms

Parameter	JAXA	LPRM
Soil and vegetation canopy physical temperatures	$T_s = T_c$ =293 K	$T_s = T_c$, linearly related with T_b(37 GHz[V])
Surface roughness	Constants *Q* and *H*	Constants *h* and *Q*
Vegetation	$\tau = b \cdot W_c$ $f_c = f(NDVI)$ ω = 0.060 ~ 0.063 depending on polarization and frequency	τ = f(MPDI, *k*, u, ω) ω = 0.060
Dielectric mixing model	Four-stream fast model (Liu, 1998)	Wang and Schmugge (1980)

T_s: soil surface temperature, T_c: vegetation canopy temperature, *Q*: polarization mixing ratio, *H* and *h*: roughness parameters used in JAXA and LPRM, τ: optical depth, W_c: vegetation water content, *b*: vegetation parameter, f_c: fractional vegetation coverage, *k*: dielectric constant, *u*: incidence angle, ω: single scattering albedo.

1. Physical surface temperature: The JAXA algorithm assumes that both soil and vegetation canopy temperatures are constant (293 K) throughout the year globally. In contrast, the LPRM algorithm assumes that these surface temperatures are linearly related to the brightness temperature from 37 GHz vertical polarization and therefore the temperatures will vary both temporally and spatially.

2. Surface roughness: In both JAXA and LPRM algorithms, the emissivity is assumed to be related to the surface roughness, however different assumptions have been made in the way that surface roughness is defined. Both algorithms use vertical and horizontal polarizations but they use different constants to define surface roughness.

3. Vegetation: VOD is calculated differently in the two products. The JAXA algorithm uses the optical information from NDVI to calculate the fractional vegetation coverage which is then used in a linear relationship between vegetation water content (W_c) and VOD. The LPRM algorithm

uses the microwave brightness temperature (in MPDI) and the dielectric constant to estimate VOD.

4. Dielectric mixing model: Different dielectric mixing models are used in the JAXA and LPRM algorithms (Table 3.2) to relate the soil dielectric constant to the soil moisture estimate. In addition, the JAXA algorithm separately considers very dry conditions explicitly, whereas the LPRM algorithm is the same for all soil moisture levels.

3.2.2. Field Soil Moisture Measurements

The cosmic-ray method is a newly developed way to measure soil moisture representative of an area of a few hectares and to a depth in the order of a few hundred millimeters in the COsmic-ray Soil Moisture Observing System (COSMOS) (Zreda et al., 2012). Stationary cosmic-ray probes detect neutrons generated by cosmic rays in air, soil and other materials. The neutrons are mainly moderated by hydrogen atoms which are primarily located in soil water, and emitted to the atmosphere in which they are instantaneously mixed within a few hundred meters and the density is inversely correlated with soil moisture (Zreda et al., 2012). At present (2015), 95 COSMOS stations are in operation providing hourly soil moisture data; 68% of them are in the United States with the remainder in Europe, Africa, South America and Australia. Two types of data are available: 1) soil moisture for the counting time interval (1 hour) in volumetric units, and 2) a 12-hour running average by using the 12-hour robust boxcar filter to reduce the noise (Zreda et al., 2012) from the International Soil moisture Network (ISMN) (Dorigo et al., 2011) and which is used in this study.

The use of the COSMOS over different continents ensures consistency in measurement techniques and is expected to minimize the uncertainties associated with ground measurements which use various probe types, settings and methods over different regions. Additionally, COSMOS data is expected to provide an area-representative value over point measurements within a footprint of the cosmic-ray probe (Zreda et al., 2012).

Nevertheless, one known issue with cosmic-ray derived soil moisture is its uncertainties under high atmospheric water vapor (Zreda et al., 2012). Previous research has investigated the extent

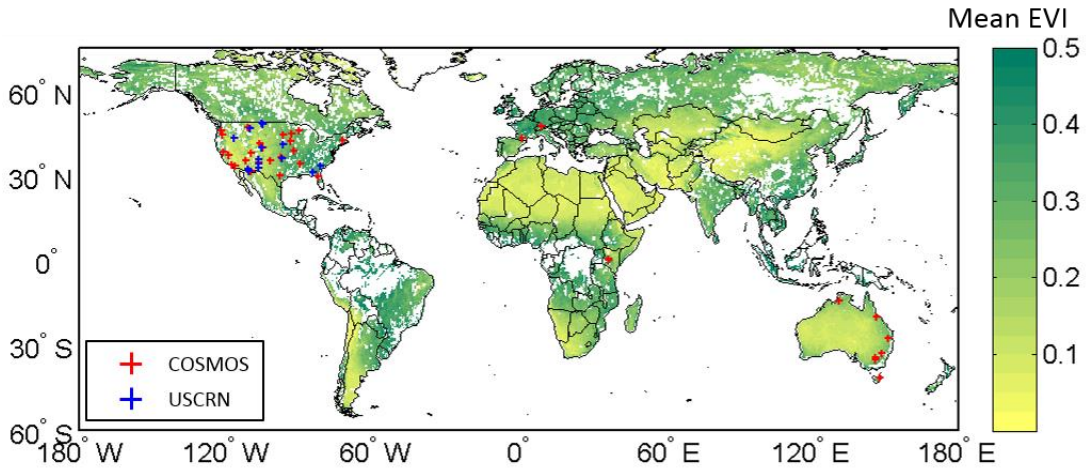


Figure 3.3. Locations of 47 COSMOS stations used in this study and coexisting 17 USCRN stations, presented with red and blue '+' symbols respectively. The background colour indicates annual average EVI during the entire study period

of the problem and has suggested corrections with respect to the atmospheric water vapor (Bogena et al., 2013; Rosolem et al., 2013). To avoid possible errors in this study from the COSMOS data a well-tested quality control (QC) procedure (Dorigo et al., 2013) has been applied, which detects unnatural increases or decreases in soil moisture and flags for those potentially doubtful observations. In this study, all flagged soil moisture values are excluded from further analysis. Accordingly, the final dataset includes field measurements from 47 stations which have at least 6-month overlapping with the satellite observations were used.

As another way of validating the COSMOS data, soil moisture measurements (soil layer depth 5 cm) from 17 stations in the U.*S*. Climate Reference Network (USCRN) (Diamond et al., 2013) are compared to COSMOS datasets when one or more USCRN stations are located within 50 km (*i.e.* adjacent grids) of the COSMOS stations (Figure 3.3). The results suggest that the temporal patterns of soil moisture measurements between these COSMOS and traditional ground stations are well correlated and provide additional confidence in the comparisons with AMSR2. Despite the advantages of COSMOS data: the consistency in measurement techniques and the wider horizontal coverage compared to point measurements, it should be noted that the difference in the measurement depths of the cosmic ray probes and the satellite derived soil moisture products still remains a limit just like all studies validating remotely sensed soil moisture, which

use *in situ* measurements as a reference. It is therefore appropriate to consider the discrepancies in absolute values of the two products, which are represented by bias and root mean square error (RMSE) in this study, as supportive metrics. However, it should be also noted that temporal correlations are less affected by the depth discrepancy problem and therefore the main conclusions in this work rely on the correlations.

The field measurement from each COSMOS station is compared with the satellite-based soil moisture grid cell in which the station is located. When two or more stations are in the same grid cell, their average is taken first. As the satellite observation is a snapshot at the overpass time (01:30 a.m. equatorial local crossing time for AMSR2 descending overpass), the field measurement recorded at a time closest to the local overpass time is used for a more direct comparison.

3.2.3. Ancillary Data

The MODIS enhanced vegetation index (EVI) provides information about global green vegetation conditions. In this study, the original monthly 0.05° resolution EVI product (MOD13C2) was aggregated to 0.25° by taking the average value of the 25 0.05° grid cells in each 0.25° grid cell and used as an indication for the vegetation condition at each site.

As an indication for surface roughness at the remotely sensed soil moisture scale (0.25°), the Global Land One-kilometer Base Elevation Digital Elevation Model (GLOBE DEM version 1.0) (GLOBE-Task-Team & Others, 1999) was used. In each 0.25° grid cell, there are 625 1km DEM grid cells. An indication of the relative surface roughness of each 0.25° grid cell was estimated as the variance (σ^2) of elevations from all 625 1km-DEM grid cells (Choudhury et al., 1979). The spatial distribution of surface roughness values at the global scale can be found in Appendix (Figure A3.2). This surface roughness proxy was then compared to the errors in soil moisture retrievals for each COSMOS station to test the hypothesis that the differences in the two products' algorithms for surface roughness leads to changes in the estimates of soil moisture. Although this surface roughness proxy does not exactly represent the overall surface roughness

affecting the microwave emission, this simplified investigation provides information on the relationship between surface roughness and the retrieval algorithms' performance.

To evaluate the temporal dynamics of the satellite-based soil moisture products, the 3-hourly precipitation product (3B42 V7) from the Tropical Rainfall Measuring Mission (TRMM) (Huffman & Bolvin, 2014) was aggregated to daily total precipitation (mm/day). This information can assist in interpreting the anomaly timeseries of soil moisture where the responses of both products are expected to be influenced by local rainfall.

It is not possible to retrieve soil moisture under frozen conditions. The LPRM applies an internal soil-freezing masking step in the product (Parinussa et al., 2011a) whereas the JAXA algorithm does not omit data under such conditions. Therefore, to ensure consistent comparisons between the two products, a mask to remove estimates under frozen conditions has been used. To construct the mask 6-hourly soil temperature of the top-layer (*i.e.* 0-0.07 m) from the ERA-Interim reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) (Dee et al., 2011) was adopted. Since the original soil temperature data is on the Universal Time Coordinated (UTC) time, they were first converted to local time based on station coordinates. Then temporal interpolation was performed between two adjacent temperature values to obtain soil temperature values at the overpass time. When the interpolated ERA-Interim soil temperature was below 0°C, the corresponding satellite-based soil moisture retrievals, field soil moisture measurements, VOD and EVI were masked out. Other masks that were applied to the data included removing grid cells along the coastline (*i.e.* center of grid cells within 25 km from the coast) to remove the influence of ocean water. Regions with dense vegetation (*i.e.* annual mean VOD greater than 0.8 at 6.9 GHz derived from the LPRM) were also excluded in this analysis (De Jeu et al., 2008).

3.2.4. Statistical Metrics

The AMSR2 products were evaluated by comparing the two products to each other as well as comparing them to the ground-based soil moisture measurements. The temporal correlations between the JAXA and LPRM products were examined over all grid cells for both raw data and

anomaly data. The anomaly data was obtained in two steps. The first step was to calculate the seasonal cycle by taking a 31-day moving average over the study period 1 August 2012 through 31 July 2013, while the data used is from mid-July 2012 through mid-August 2013. The anomaly data was calculated by removing the seasonal cycle from the raw data. It should also be noted that the 1-year analysis period did not allow conclusions on inter-annual variations and this could be a topic of interest for the future.

The seasonal cycles and anomaly timeseries from each product were compared over all grid cells by considering their correlations as well as the maximum, minimum and mean values.

At each grid cell, a paired Student's *t*-test was used with the daily timeseries to test the differences in the mean of the JAXA and LPRM soil moisture products, with a significance level of $\alpha=0.05$ adopted for this test. To compare the satellite-based and ground-based soil moisture, three statistical metrics were used, namely, the Pearson correlation coefficient (R), bias and root mean square error (RMSE).

3.3. Results

3.3.1. Inter-Comparison of JAXA And LPRM Products; How Are They Different?

The spatial patterns of mean, maximum and minimum values of the JAXA and LPRM soil moisture products and their difference are presented in Figure 3.4. The mean values of JAXA were generally lower than LPRM. Over the high latitude regions, *e.g.* Russia and Canada, both maximum and minimum values of JAXA were lower than LPRM, which results in the lower mean values. Over the other regions (except desert regions), the minimum values of JAXA and LPRM were quite similar. The lower mean values of JAXA were primarily caused by its lower maximum values. This may be related to the fact that the soil moisture range in the JAXA and LPRM algorithms is 0-0.6 m^3/m^3 and 0-1 m^3/m^3, respectively. The maximum value (0.6 m^3/m^3) of the JAXA product is a result of the pre-established lookup tables that match soil moisture and vegetation water content to microwave brightness temperature-derived indices (*i.e.* PI and ISW see section). The minimum values of the JAXA product over the desert regions (*e.g.* Sahara, Africa, Middle East, south Mongolia and central Australia) are somewhat higher than LPRM. This can be attributed to the approach used in the JAXA algorithm to estimate the soil moisture under extreme dry conditions (<0.1 m^3/m^3) as described in the section 3.2.1. With these considerable differences in the absolute values between these two products, it is not surprising to see that the paired-*t* test indicates the means of the two products are statistically different for nearly all grid cells (Figure A3.1)

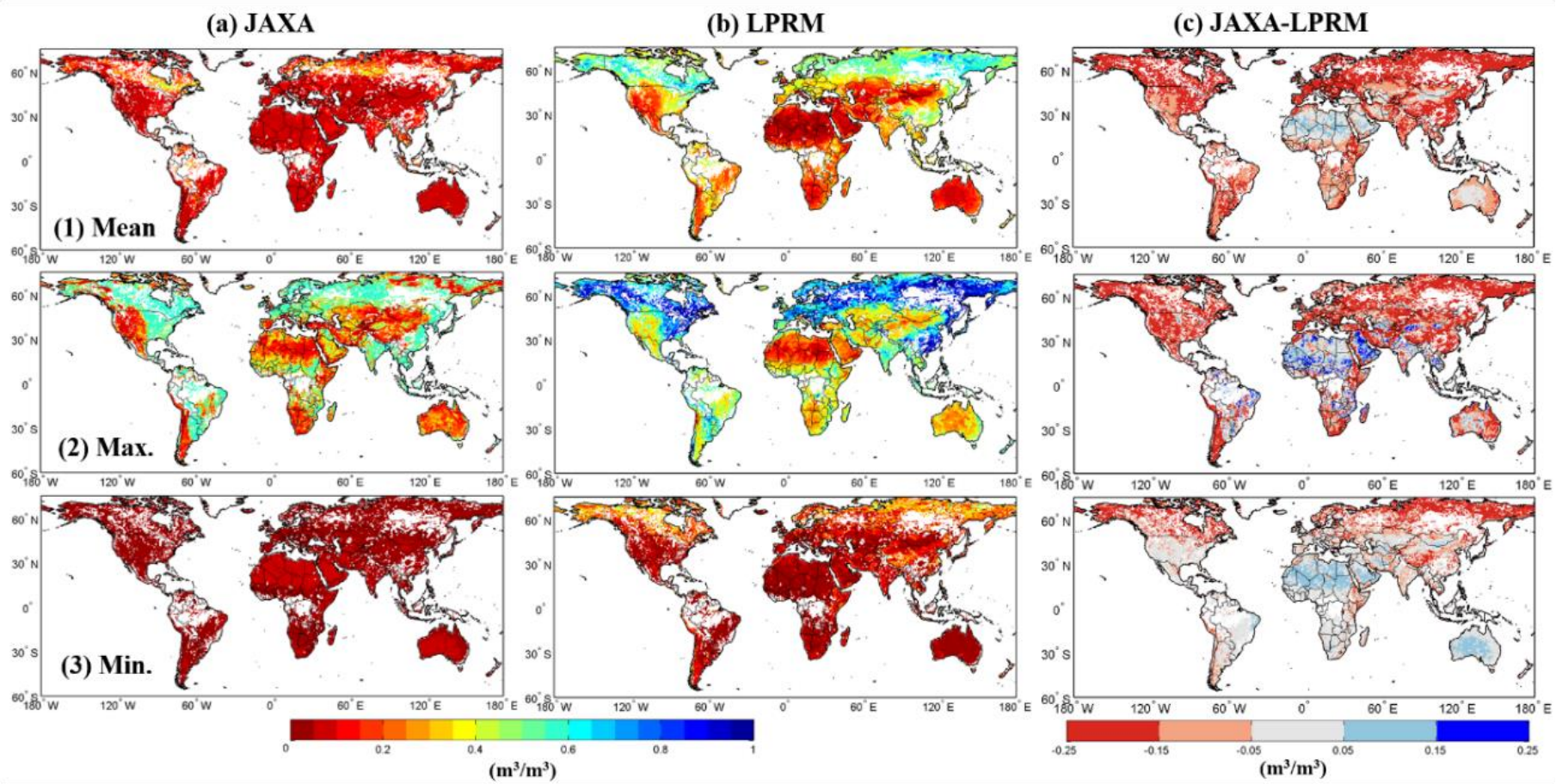


Figure 3.4. Global maps of mean (top panels), maximum (middle panels) and minimum (bottom panels) values of JAXA (left column), LPRM (middle column) and differences (*i.e.* JAXA-LPRM, right column) derived from descending overpasses.

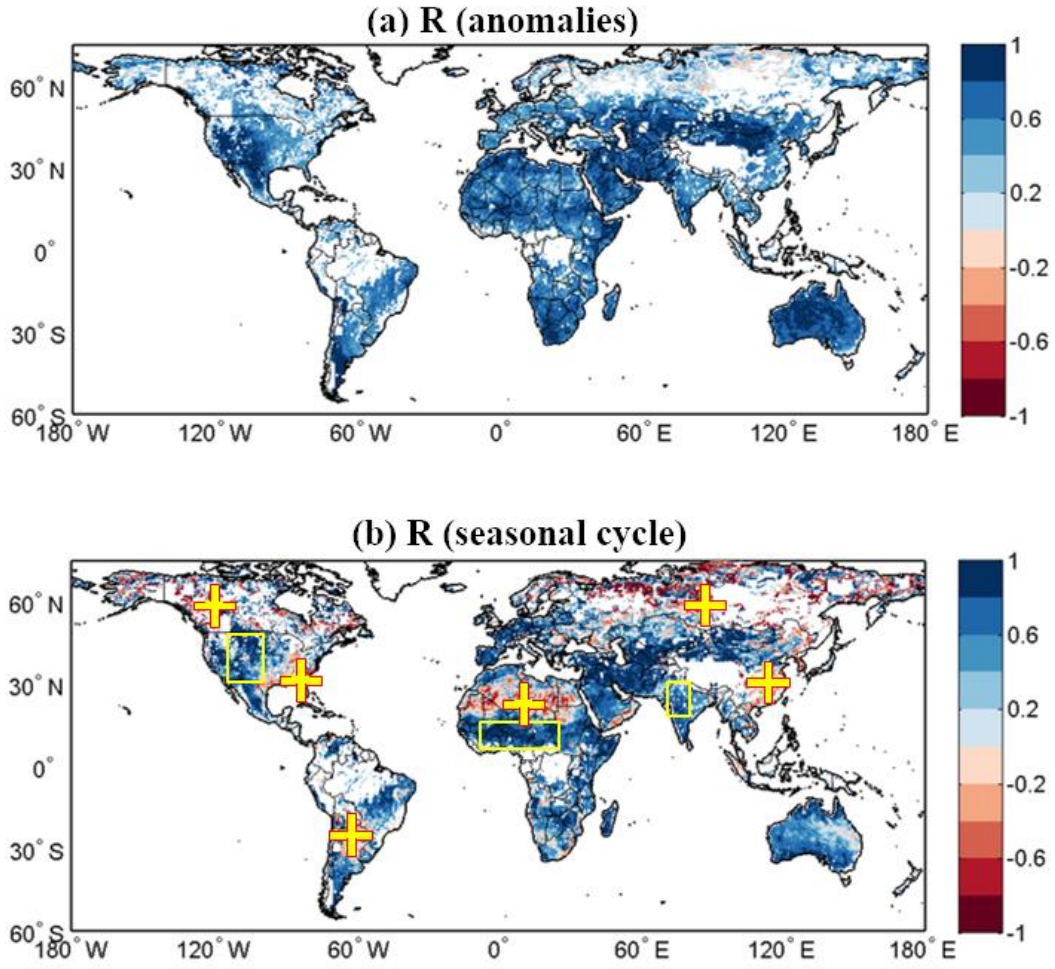


Figure 3.5. Global maps of correlation coefficients (R) between JAXA and LPRM soil moisture products derived from descending overpasses for (a) soil moisture anomalies ($R_{anomaly}$) and (b) seasonal cycle (R_{season}). Three regions with strong positive correlations in seasonal cycle (outlined by yellow boxes) are in the transition zones identified by Koster et al. (2004). Six regions with strong negative correlations (labelled as yellow crosses) are selected for further temporal comparison.

The signals in soil moisture products can be decomposed into the 1) seasonal cycle and 2) soil moisture anomalies which primarily represent responses to rainfall events. The correlation coefficients of the seasonal cycle and anomalies between the two AMSR2 soil moisture products are presented in Figure 3.5. The soil moisture anomalies of the two products are highly positively correlated (Figure 3.5a) because of similar responses to external meteorological forcing and to some extent the same random noise in the satellite observations. Over the 'transitional regions' where strong coupling between soil moisture and precipitation are expected, *e.g.* central Great

Plains of North America, Sahel and India (Koster et al., 2004), both seasonal cycle and soil moisture anomalies are highly positively correlated between these two AMSR2 products (regions highlighted with yellow boxes in Figure 3.5b). Strong negative correlation coefficients in seasonal cycles are also found over many regions, *e.g.* western Canada, Russia, southeast USA, southeast China, central South America and northern Africa (Figure 3.5b) Again, it should be noted that the 1 year timeframe in which the products were evaluated does not allow conclusions on inter-annual variations and there will be more confidence in the seasonal cycle results in future when a multi-year climatology can be constructed.

To better understand their temporal characteristics, timeseries of the JAXA and LPRM soil moisture together with soil temperature and EVI over the highlighted regions in Figure 3.5b are shown in Figure 3.6 and Figure 3.7. Over the transition regions, the overall correlations during the study period are high and the temporal dynamics as responses to precipitation events are very similar (Figure 3.6). However, the recession patterns since the end of the wet season are quite different between these two AMSR2 products. Over the Sahel region (Figure 3.6b and c), the JAXA soil moisture drops to its lowest value very quickly at the end of the wet season and remains constant throughout the following dry season, but the LPRM soil moisture declined gradually. When there is no precipitation event, the temporal patterns of the LPRM soil moisture tend to be opposite to surface temperature. The apparent difference between these two AMSR2 products under the very dry conditions may also come from their different dielectric mixing models. It is noticeable that the EVI value over India is similar between dry and wet season (Figure 3.6d). Meanwhile, the LPRM soil moisture is kept to 0.3 m^3/m^3 in the dry period whereas the JAXA soil moisture drops to 0.1 m^3/m^3.

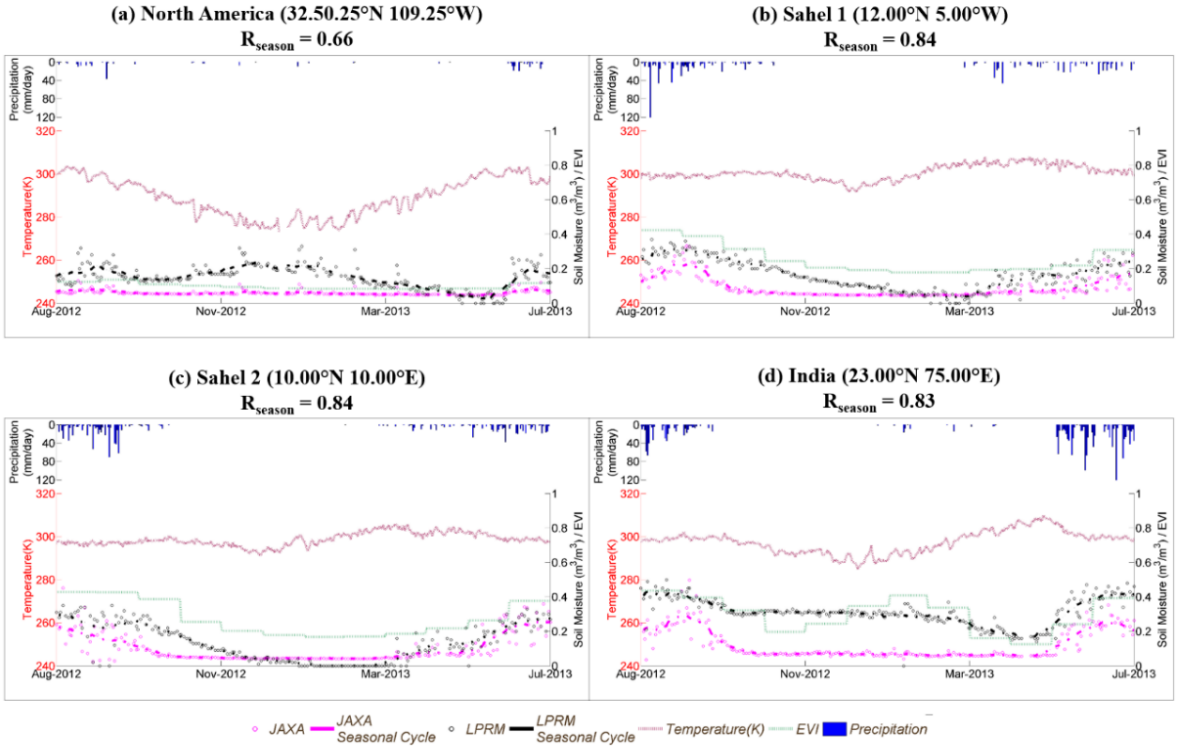


Figure 3.6. Descending pass timeseries of JAXA, LPRM, temperature and EVI at four locations in the transition zones that have strong positive correlations in the representation of the seasonal cycle; (a) the central Great Plains of North America, (b) and (c) Sahel, and (d) India.

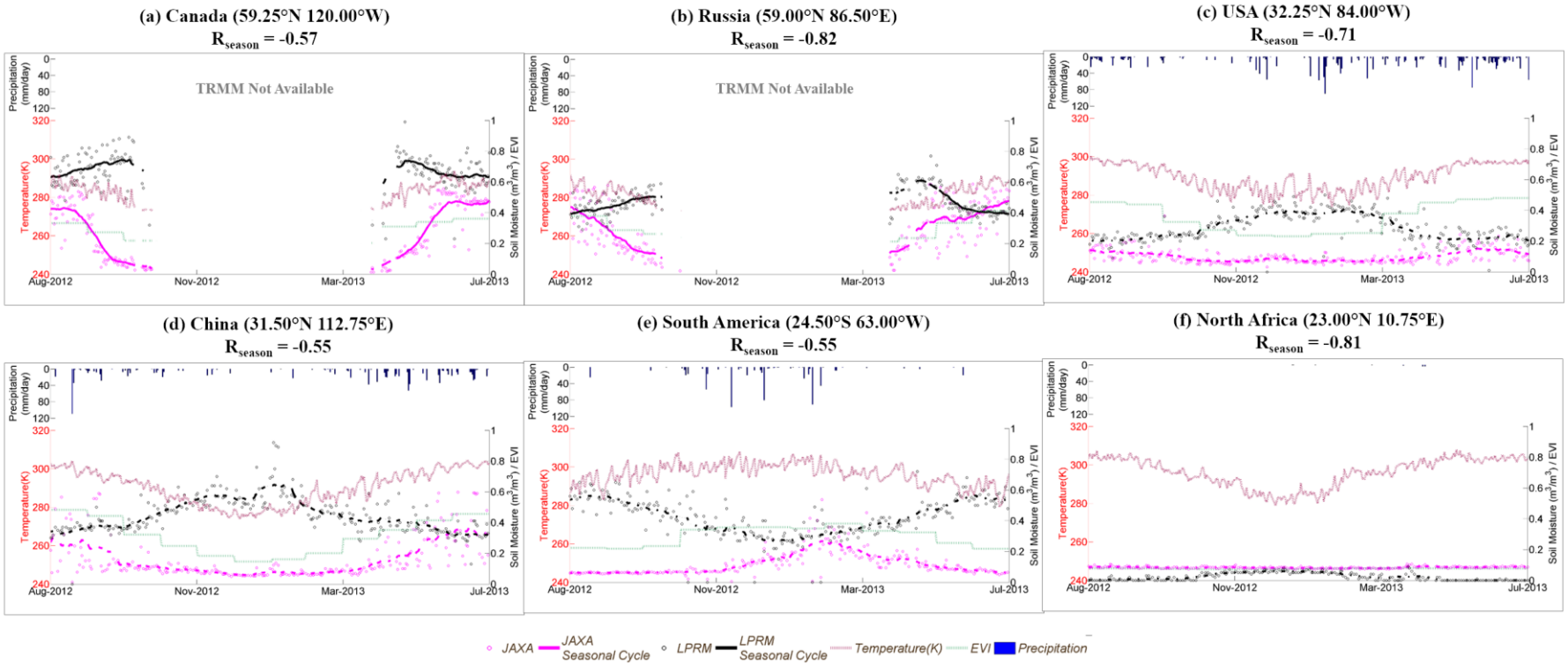


Figure 3.7. Timeseries of JAXA, LPRM, temperature and EVI at six locations that have strong negative correlations in the representation of the seasonal cycle; (a) Canada, (b) Russia, (c) USA, (d) China, (e) South America and (g) North Africa. Different scales on y-axis are used in panel f for better visualization.

Over the regions where strong negative correlations in seasonal cycles are observed, it appears that JAXA soil moisture variations are generally in proportion to temperature at the locations with strong seasonal cycles in temperature (Figure 3.7a-d). In contrast, the seasonal cycle of the LPRM soil moisture tends to be opposite to the JAXA product and temperature. When the temperature is consistently high (*e.g.* Figure 3.7e South America), the JAXA soil moisture still has a strong seasonal cycle which is similar to the EVI temporal pattern. Meanwhile, the LPRM soil moisture drops sharply with increases in EVI in austral summer (*i.e.* late 2012 to early 2013). Over the arid Sahara (Figure 3.7f) with low EVI during the whole year, there are no notable variations in the JAXA soil moisture during the whole period, while the LPRM soil moisture increases during the rainy season.

The above observations suggest that over these regions where the JAXA and LPRM soil moisture have opposite seasonal cycles, the dynamics of the JAXA soil moisture agrees with temperature and/or EVI variations whereas the seasonal cycle of LPRM soil moisture is inversely related to temperature and seems to be considerably affected when vegetation density (represented by EVI) is high (which is in line with the findings in Parinussa et al. (2011c)). It can also be seen in Figure 3.7e that when EVI is high in the austral summer that the behavior of JAXA and LPRM is quite different. The performance of the AMSR2 products and vegetation is examined in more detail in the next section for all COSMOS stations to determine if a threshold when EVI affects soil moisture can been found.

3.3.2. Comparisons Against Field Measurements

For the grid cells with particularly high correlations and those with low or even negative correlations between the JAXA and LPRM, both products are compared to field measurements using the available COSMOS stations over the USA. The USA was selected as it has one of the transition zones identified by Koster et al. (2004) and also has areas where the two products have similar and different seasonal cycles.

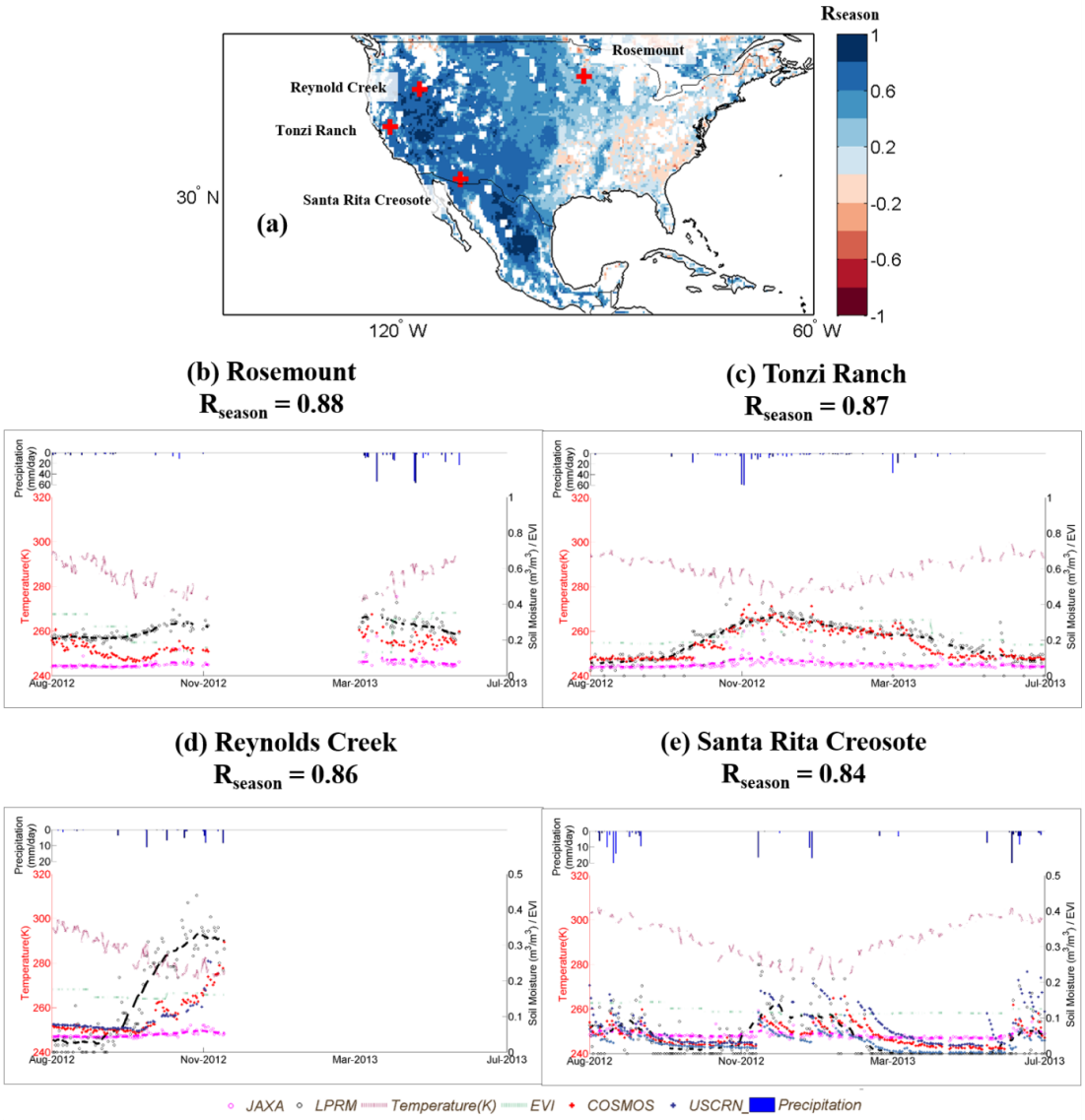


Figure 3.8. Timeseries of AMSR2 soil moisture retrievals, soil temperature, EVI, precipitation and ground soil moisture measurements for four COSMOS stations where R_{season} between the JAXA and LPRM products are highest among all grid cells with COSMOS stations. Panel (a) shows the location of these four stations and R_{season} in the background. Different scales on y-axis are used in panels d and e for better visualization.

At the grid cells with good agreement in the seasonal cycle (Figure 3.8), the LPRM soil moisture estimates were generally higher than the JAXA estimates, except for the dry season at Santa Rita Creosote (Figure 3.8e). The LPRM soil moisture had strong variations throughout the year, while the JAXA soil moisture showed little variation with a similar minimum value around 0.05 m^3/m^3 in the dry season for all four locations. Furthermore, the LPRM soil moisture

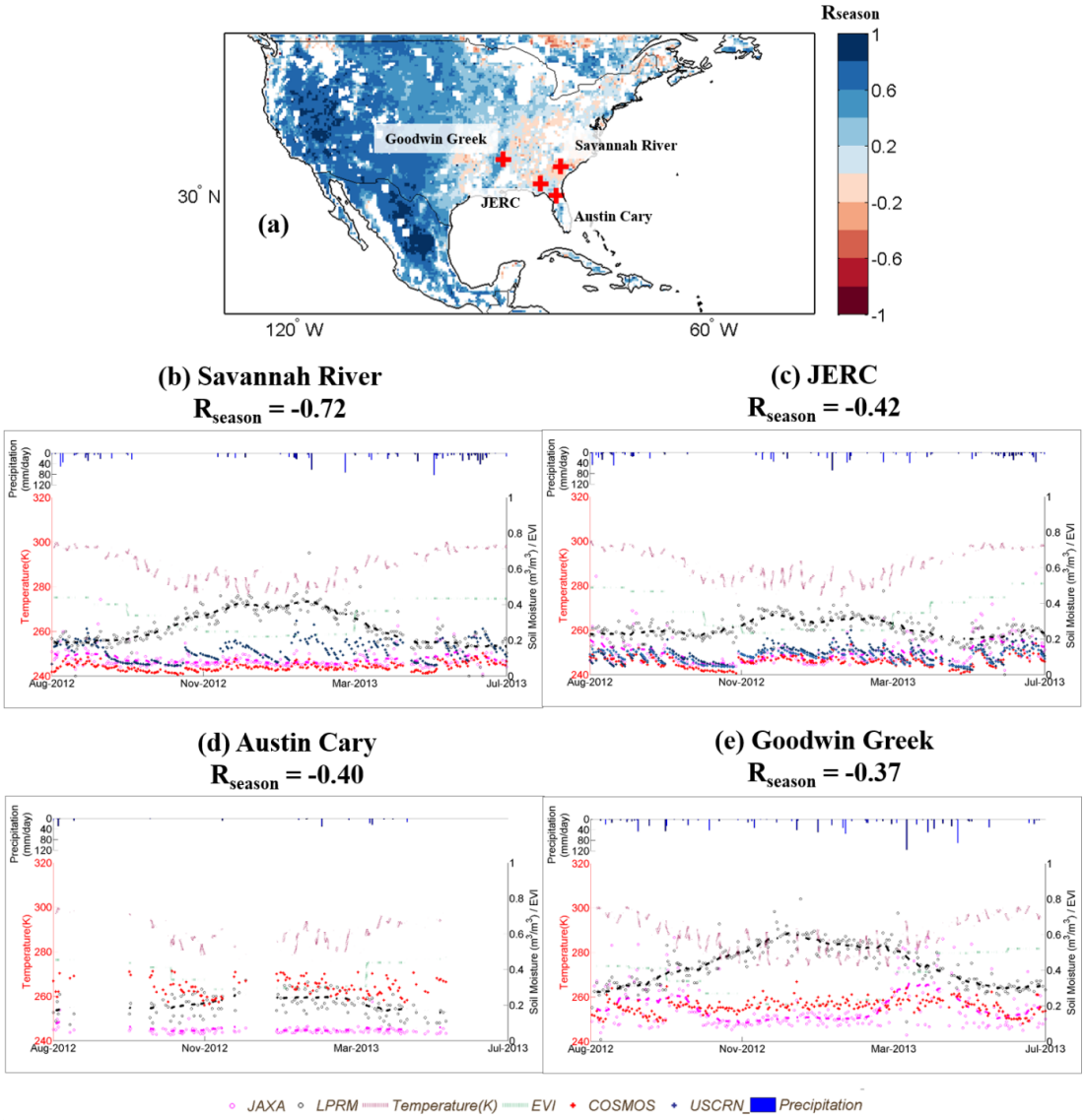


Figure 3.9. Same as Figure 3.8, but for four COSMOS stations where R_{season} between the JAXA and LPRM products are lowest among all grid cells with COSMOS stations.

followed the temporal patterns captured by the ground-based measurements better than the JAXA soil moisture retrievals.

When it comes to the grid cells with strong negative correlations in the seasonal cycle (Figure 3.9), it is worthwhile to highlight several points. 1) These stations are in forested areas and the mean EVI and VOD values are higher than the average value of all 47 COSMOS stations used in this study. 2) The dynamics of the LPRM soil moisture tended to be opposite to field measurements and/or the JAXA soil moisture (*i.e.* Savannah River which has the highest mean VOD among them). 3) the LPRM soil moisture values tend to be higher than the JAXA soil

moisture, which is consistent with the global mean values shown in Figure 3.4. 4) It appears that the variance of the LPRM soil moisture product is much larger than the JAXA soil moisture, as the former has much more scatter around the seasonal cycle compared to the latter. 5) At Savannah River and JERC station which are in relatively dry conditions, the JAXA product shows fairly good correspondences to the field measurements. It has been previously found that the LPRM product has difficulties in this region (Hain et al., 2011).

Expanding the comparisons from the 8 stations included in Figure 3.8 and Figure 3.9, the error statistics at all 47 COSMOS stations were considered. For each station the error statistics (bias, RMSE and R) between both AMSR2 products and ground-based measurements are plotted against four independent variables, namely, annual mean temperature, log (*h*) representing coarse scale surface roughness, annual mean EVI and mean value of ground soil moisture (dots in Figure 3.10). It should be noted that only dates which are coincident in all datasets were used for calculating the error. To better visualize and understand the patterns in these scatterplots, smooth curves were added for each plot using robust local regression method. The regression used weighted linear least squares with a span of 50% with lower weights assigned to outliers and zero weight for data lying outside six standard deviations from the mean (lines in Figure 3.10). The results from this analysis are discussed in the following subsections.

1. Annual mean temperature: Biases in the JAXA product are primarily negative, which means that the values are smaller than the field measurements, whereas the LPRM product biases are generally positive. There is no significant trend in the relationship of RMSE with mean temperature for either product, although there is a sharp decrease in the correlations (R_{JAXA} and R_{LPRM}) when mean temperature decreases below 290 K. However, the rate of decrease in the LPRM correlation-temperature relationship is smaller. This may be due to the large scatter in the LPRM correlation-temperature relationship compared to the JAXA case.

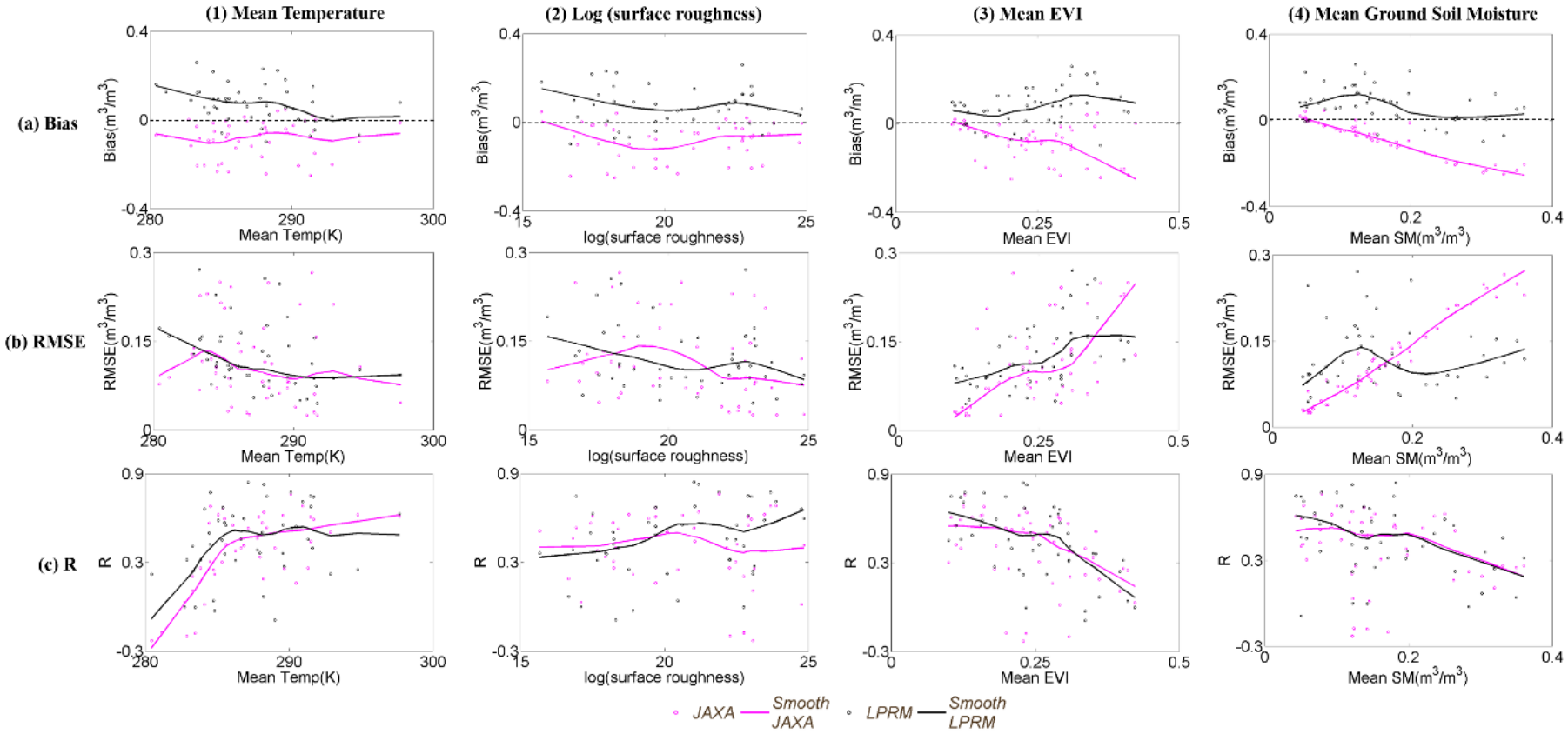


Figure 3.10 Scatterplots of (y-axis) bias, RMSE and correlation coefficients (R) between AMSR2-based raw soil moisture and field measurements against (x-axis) mean temperature, coarse scale surface roughness (log (h)), mean EVI and mean ground soil moisture. A robust local regression method is used for smoothing data.

2. Surface roughness: The pattern of biases of the JAXA and LPRM products are very similar when plotted against the coarse scale surface roughness. When considering RMSE, the largest errors in the JAXA product are seen at around log (surface roughness) = 20 which is also the case for bias, while no significant trend is observed in the RMSE value of LPRM. There is a lot of scatter in the distribution of correlation coefficients against changes in log (*h*). However, in general, there are improvements in LPRM correlations as the surface becomes rougher but JAXA correlations decreases with the rougher surface conditions. Further investigations are required to assess why LPRM correlations improved and JAXA correlations degrade with increasing surface roughness and how this knowledge could be used to improve both products.

3. Annual mean EVI: The bias in the JAXA product when considered with respect to annual mean EVI is larger than for LPRM, particularly at higher EVI values (*i.e.* mean EVI > 0.3). It is observed that the RMSE of JAXA increases strongly when mean EVI is higher than 0.3. The correlations of both product show a strong degradation in performance as mean EVI increases, with steeper decline in both correlation coefficients, occurring for mean EVI greater than 0.3.

4. Mean field soil moisture: There is a very clear relationship between errors in both products and mean field soil moisture. As mean field soil moisture increases the JAXA product increasingly underestimates the soil moisture (negative bias). Correlations of both products steadily decrease as mean field soil moisture increases. For the LPRM, bias and RMSE are relatively constant for changes of mean field soil moisture. The exception is found for the LPRM product under dry conditions (*i.e.*, field soil moisture < 0.2 m^3/m^3) where soil moisture is overestimated and for these sites the JAXA is found to have smaller errors.

3.4. Discussion

The primary aim of this study was to provide guidelines for using and improving the JAXA and LPRM AMSR2 soil moisture products. In soil moisture retrieval algorithms based on the radiative transfer equation (Mo et al., 1982), it is very difficult to estimate parameters such as single scattering albedo (ω), polarization mixing ratio (*Q*) and roughness parameter (*h*) at the global scale due to the lack of experimental data to calculate them. To solve this, the parameters are

Table 3.3. Summary of relative performance AMSR2-JAXA and LPRM for RMSE and correlation coefficients with mean temperature, log (*h*), mean EVI and mean ground soil moisture. The product with generally better results for all 47 COSMOS stations is listed. If the performance of both products was similar this is denoted by "Similar"

Variables	Range	RMSE	Correlation Coefficients
Mean soil temp. (K)	< 290	Similar	LPRM
	> 290	Similar	JAXA
Log (h)	> 20	*	LPRM
Mean EVI	< 0.30	JAXA	Similar
	> 0.30	LPRM	Similar
Mean ground soil moisture (m^3/m^3)	dry (< 0.20)	JAXA	Similar
	wet (> 0.20)	LPRM	Similar

* JAXA product shows better RMSE out of log (*h*) 18 to 22

generally chosen based on past research and limited site experiments (Fujii et al., 2009; Njoku et al., 2003; Njoku & Li, 1999; Owe et al., 2001). Uncertainties in the parameterizations have been used for error estimations of soil moisture retrievals (Parinussa et al., 2011c). Although these parameterization choices will lead to biases and possible errors in the modelled dynamic ranges, it has been proved that the soil moisture retrievals generally perform well in terms of temporal variability. In many applications and particularly climate studies, the temporal variability is likely to be the most important statistic and scaling approaches have been proposed to adjust the range of soil moisture (Draper et al., 2009; Entekhabi et al., 2010b). Therefore, in this study the temporal variability of soil moisture was firstly focused. The second area of focus was to examine the difference in values resulting from the two different algorithms by using raw soil moisture retrievals. This second focus assessed the global parameterizations and suggested areas where improvements can be made in the parameterizations through improved understanding of how the two approaches affect the accuracy of soil moisture retrievals.

In the analysis, the two AMSR2 soil moisture products were compared with the field soil moisture from 47 COSMOS stations. Four primary factors were identified between the JAXA and

LPRM retrieval algorithms, namely, physical surface temperature, surface roughness, vegetation density and quantitative soil wetness conditions, which were expected to affect the accuracy and precision of the satellite products. A summary of the performance of both products with respect to the independent variables is presented in Table 3.3 based on the results in Figure 3.10 and discussions in the previous section. Five main conclusions were drawn. 1) The JAXA algorithm generally underestimated the ground soil moisture, whereas the LPRM algorithm tends to overestimate soil moisture. The distributions of bias and RMSE of the LPRM product are relatively insensitive to changes of the four independent variables, whereas the JAXA soil moisture showed larger degradations in performance above certain thresholds. 2) Correlation coefficients between AMSR2 products and ground measurements decreased when the mean temperature decreased below approximately 290 K, but the degree of decline in the LPRM product was smaller than the JAXA product. 3) Even though there was large variance in the errors when shown against surface roughness, in general the LPRM correlations increased as the surface became rougher whilst the JAXA correlations decreased. 4) The performance of JAXA is affected in areas with dense vegetation, particularly for mean EVI greater than 0.30. Correlation coefficients of both products decline for mean EVI is higher than 0.3, which is in line with the findings in Parinussa et al. (2011c). 5) Distributions of bias and RMSE of LPRM are relatively insensitive to variation of mean ground soil moisture; however, the JAXA algorithm performs better in dry condition (< 0.2 m^3/m^3). Correlations for both products gradually reduce as mean ground soil moisture increases.

These results suggest possible areas for future improvement of soil moisture retrievals. Nevertheless, there are several limitations with the analyses of this study. The main factors that will have affected the results are: 1) use of data in a single year to calculate the statistics and shortage of ground stations to derive general conclusions due to the short period of data available, and 2) discrepancies in the spatial scale of AMSR2 and COSMOS soil moisture measurements.

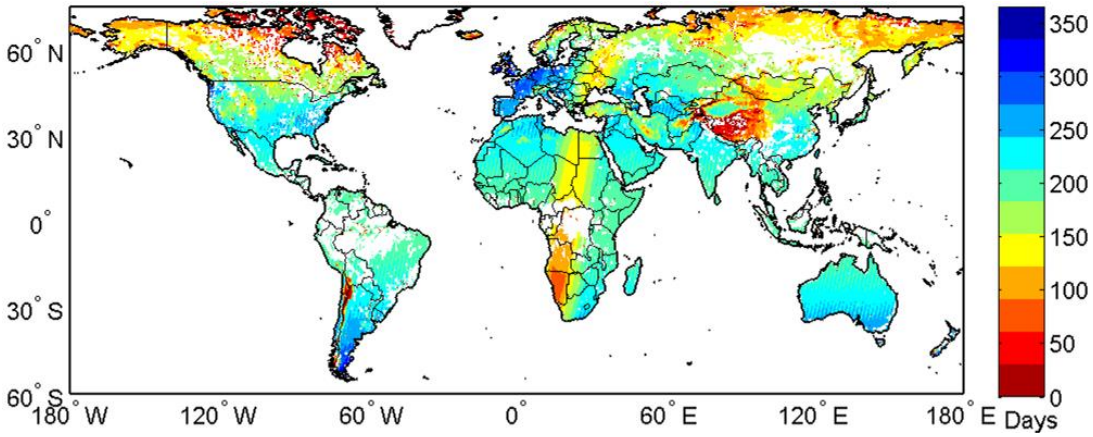


Figure 3.11. Number of days available to calculate correlations between the JAXA and LPRM soil moisture estimates.

As the AMSR2 data are made available from July 2012, its temporal coverage is less than 2 years at the time of writing (2014). Therefore, there are not many ground stations with sufficient overlap with the study period. Secondly the short record length means that it is not able to examine inter-annual variations. Furthermore, the data mask using the 6 hourly soil temperature data reduced the number of available data. This insufficient temporal coverage and the differences in temporal resolution (*i.e.*, hourly to monthly) likely leads to more uncertainty in the results. Future research using multi-year data can potentially address this issue and provide further insights into the difference and similarity between these two AMSR2 retrieval algorithms and products.

Figure 3.11 shows the available number of days in common between the two data sets that was used to calculate the correlation coefficients. Most of the ground data is in the USA where it can be seen that there is generally a relatively low number of available records, with the number of available days decreasing as the latitude increases and the chance of freezing conditions increases, leading to more masking. Also, evident in Figure 3.11 is a strip across Africa, where there are extremely low numbers of days in common between the two data sets. Interestingly this effect is not evident in the ascending pass data (Figure A3.6), which indicates that it is due to the different processing of the descending pass between the JAXA and LPRM algorithms. The JAXA algorithm separates a swath when a UTC scanning time passes 00:00 AM and then regards the remain part as the next day retrieval. However, the LPRM algorithm

includes the swath by the south pole in the same day of retrieval. Consequently, the numbers of paired observations between the JAXA and LPRM products decrease along the strip.

The spatial resolution of the AMSR2 soil moisture is 0.25° grid and the area-averaged value over each grid represents soil moisture to a depth of few millimeters to centimeters. On the other hand, COSMOS soil moisture is an area-representative value within a diameter of a few hundred meters and to a depth of a few hundred millimeters. For these reasons, some part of the mismatch between the AMSR2 and COSMOS soil moisture products can be attributed to these differences in horizontal and vertical support which is translated into the evaluation statistics. This is a general problem of all validation studies using *in situ* measurements that have different horizontal and vertical coverages compared with remotely sensed soil moisture. It is therefore appropriate to see bias and RMSE as supportive metrics showing the relative differences of the two retrievals with the reference dataset and consider correlation coefficients as the main metric.

As noted in the previous section, there is considerable scatter in the ground comparisons shown in Figure 3.8, Figure 3.9 and Figure 3.10. This scattering issue has been reported in previous work (Draper et al., 2009). In general, this scatter has been explained by the systematic differences between the satellite-derived and ground soil moisture. There are also uncertainties in averaging all swath data with variations due to the progression of satellite orbit, soil parameters and different spatial coverage. In this study, some further possible reasons for the scatter in the relationships have been identified. First, the number of data to calculate the statistics is reduced by masking out the data under the frozen and forested conditions. Second, there are differences in the temporal resolutions of the used data (hourly to monthly).

3.5. Conclusions

Most previous studies validating satellite-based soil moisture products have focused on comparisons with ground data at the regional scale which have a limited range of climatic and vegetative properties, or have only considered the verification without exploring the reasons for discrepancies or good agreements. This study provides an important contribution to this area

by considering global performance of the two satellite soil moisture products by comparisons with ground data under various factors as well as considering the reasons for that performance.

Due to the different spatial coverage and measurement scales of the COSMOS and satellite data, the results of correlation coefficients between them are the most reliable. In this regard, it was found that both products show rapid decreases in correlation coefficients under low mean soil temperature (< 290K), high mean EVI (> 0.3) and highly wetted conditions. In support of these correlation results, it was found that the JAXA product shows relatively better performance in bias and RMSE under dry conditions, and the bias and RMSE of LPRM were generally smaller than those of JAXA.

The results from this study suggest areas that improvements in the algorithms could be made. Firstly, the different retrievals from the two algorithms along with the relationships of soil moisture with the four external variables (*i.e.* mean temperature, surface roughness, mean EVI and mean ground soil moisture) provides information on appropriate parametrizations and model selection. Another possibility is a combined product which would leverage the strengths of the JAXA and LPRM algorithms, and this would provide improvements in temporal correlations after scaling to adjust the dynamic range of the retrievals. With these, an extended work with use of a multi-year data will be conducted in the future, by which there will be more confidence in defining the seasonal cycle and the data can be decomposed to identify the anomalies where the bias is not relevant.

3.6. Appendix

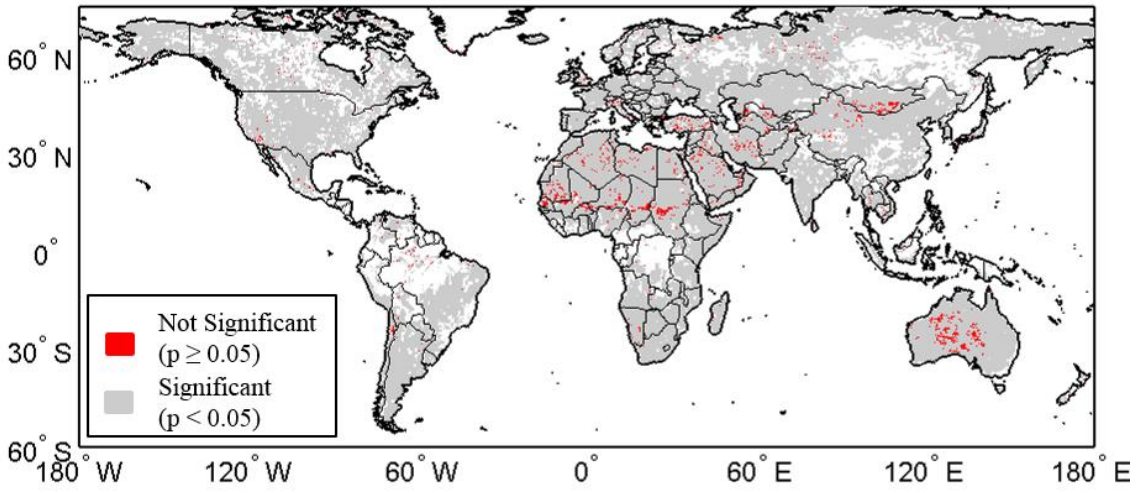


Figure A3.1. Global map (descending) of *p* values from a paired *t*-test for a null hypothesis, $\mu_{JAXA} = \mu_{LPRM}$ with α=0.05. Over the desert regions, soil moisture values from both AMSR2 products are consistently low during the entire year, and their difference is very small and not statistically significant.

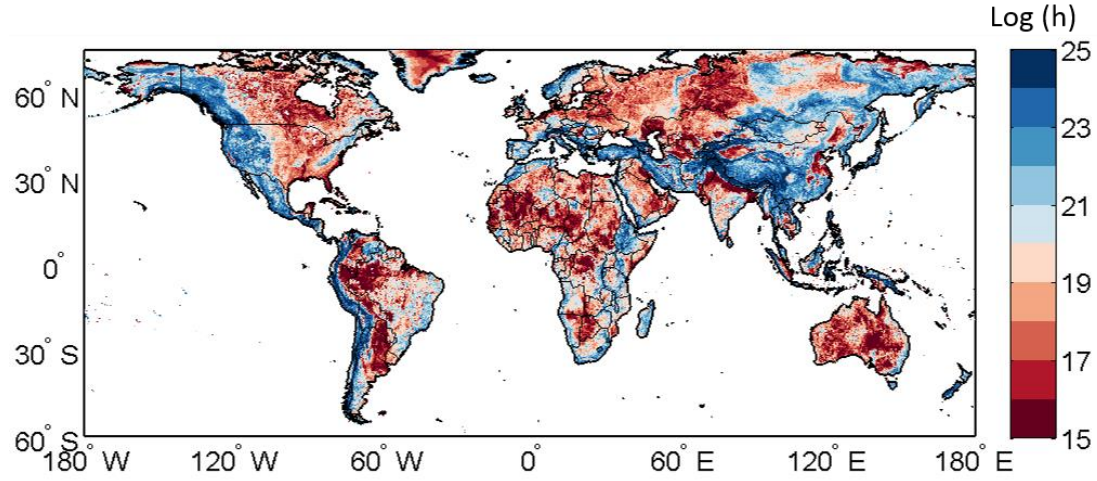


Figure A3.2. Global map of coarse scale surface roughness (log (*h*)) derived from 1-km DEM.

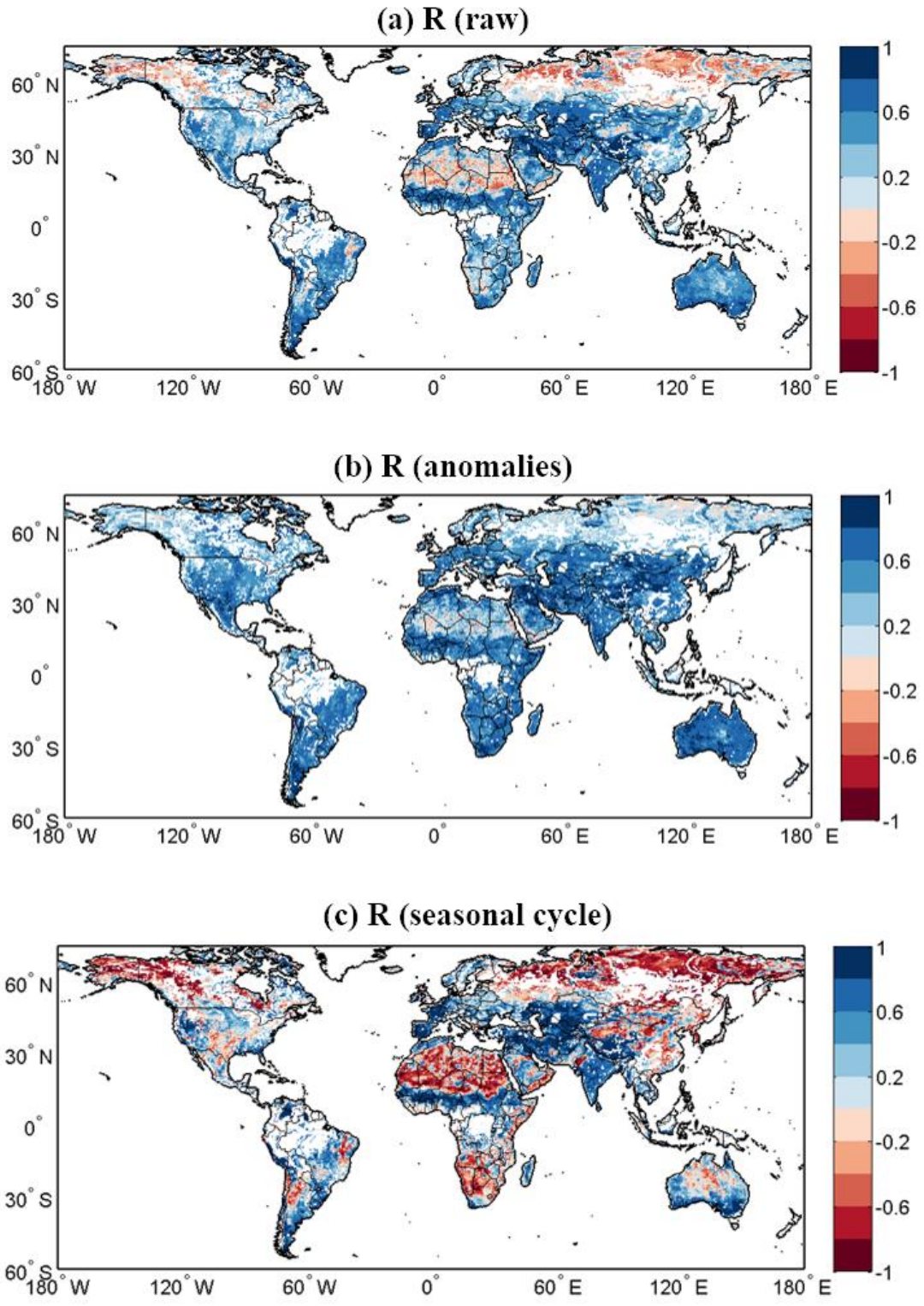


Figure A3.3. Global maps of correlation coefficients (R) between the JAXA and LPRM soil moisture products derived from ascending overpasses of 10.7 GHz (X-band) for the period 01/08/2012 to 31/07/2013. (a) Raw soil moisture (R_{raw}), (b) anomalies ($R_{anomaly}$) and (c) seasonal cycle (R_{season}).

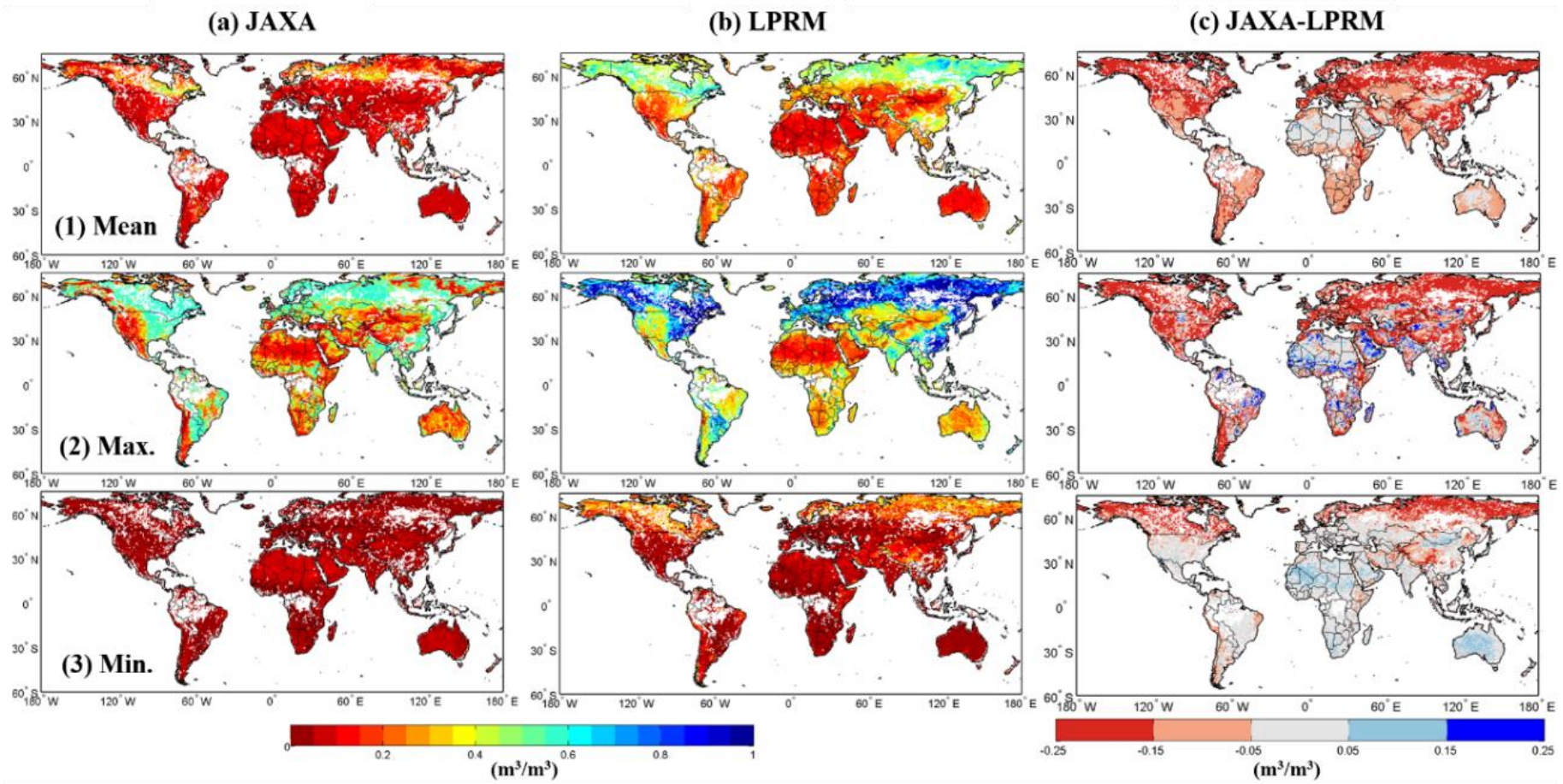


Figure A3.4. Global maps of mean (top panels), maximum (middle panels) and minimum (bottom panels) values of JAXA (left column), LPRM (middle column) and differences (*i.e.*, JAXA-LPRM, right column) derived from ascending overpasses.

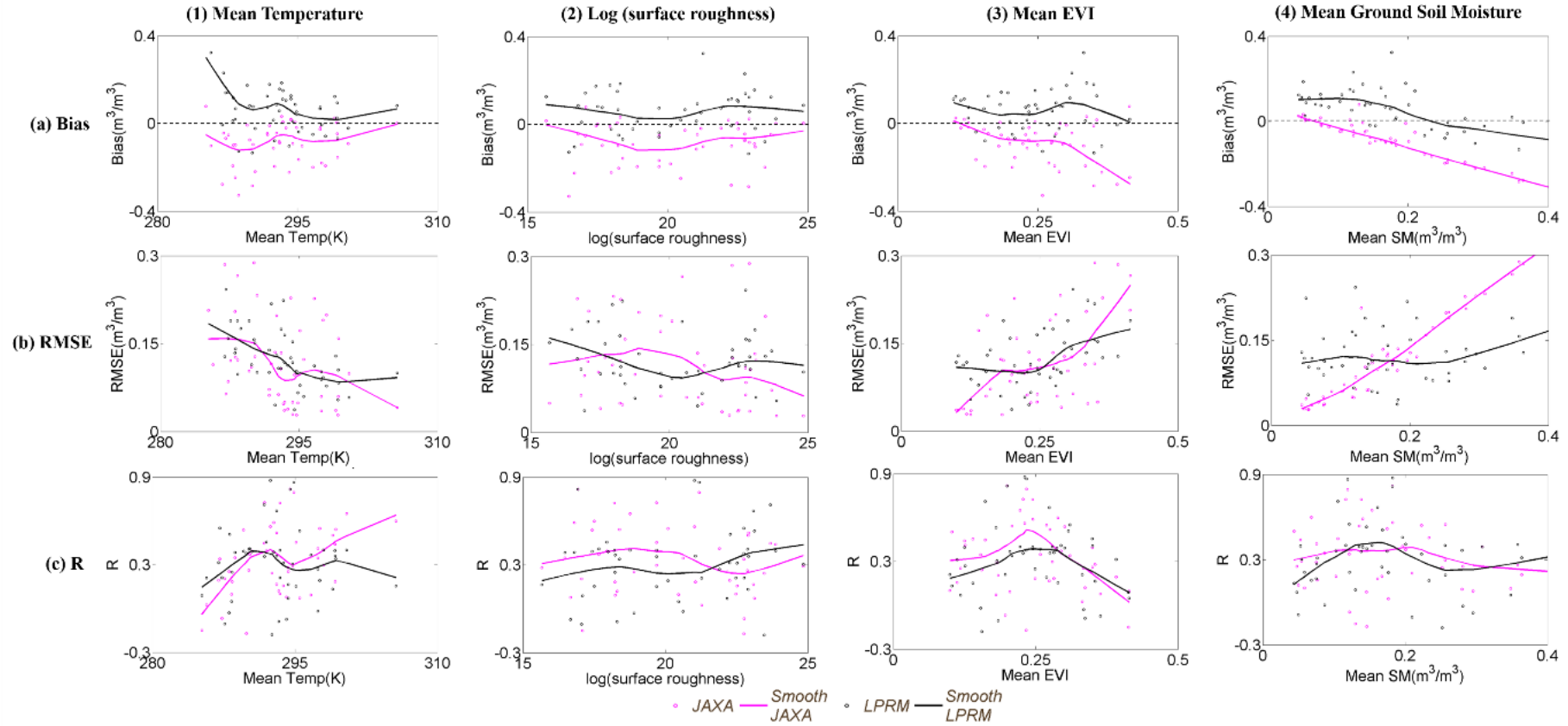


Figure A3.5. Scatterplots of (*y*-axis) bias, RMSE and correlation coefficients (R) between AMSR2-based soil moisture and ground-based measurements against (*x*-axis) mean temperature, roughness (log (*h*)), mean EVI and mean ground soil moisture. A robust local regression method is used for smoothing data.

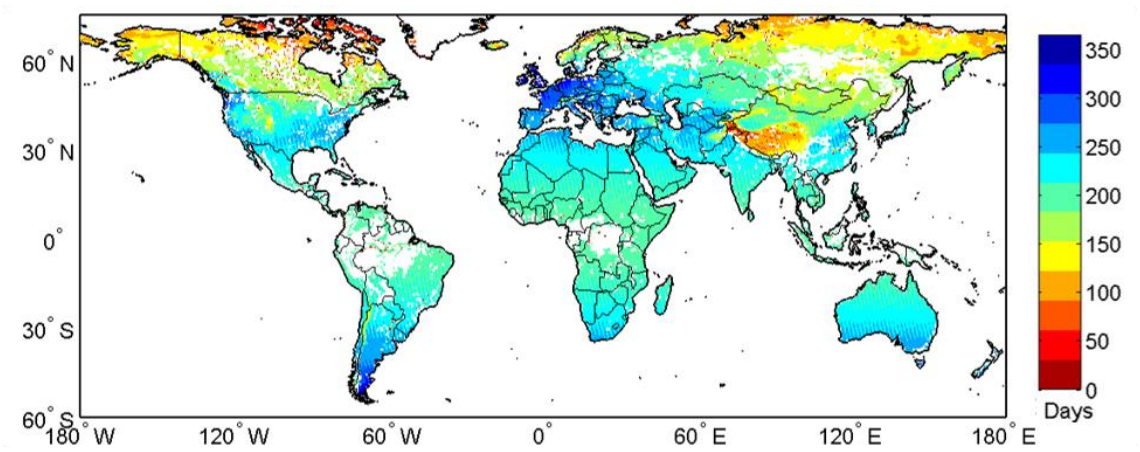


Figure A3.6. Number of available days to calculate correlation coefficients between the JAXA and LPRM soil moisture based on ascending data

Chapter 4 Combining Multiple Soil Moisture Retrievals Based on Maximizing Temporal Correlation: Static Approach

A method for linearly combining two microwave satellite soil moisture products by maximizing the temporal correlation with a reference dataset has been developed. The method was applied to two global soil moisture datasets, the JAXA and LPRM products, retrieved from AMSR2 observations for the period 2012-2014. A global comparison revealed superior results of the combined product compared to the individual products against the reference dataset of ERA-Interim volumetric water content. The global mean temporal correlation coefficient (R) between the combined product and this reference was 0.52 which outperforms the individual JAXA (0.35) as well as the LPRM (0.45) product. Additionally, the performance was evaluated against in situ observations from the ISMN. The combined dataset showed a significant improvement in temporal correlation coefficients in the validation compared to JAXA and minor improvements for the LPRM product. The mean R for the combined dataset was 0.56, 0.35 for JAXA and 0.56 for LPRM over the in situ stations. The average improvements were marginal compared to LPRM but this is most likely due to the uneven distribution of the selected in situ stations.

This chapter is an edited version of: Kim, *S.*; Parinussa, R.M.; Liu, Y.Y.; Johnson, F.M.; Sharma, A., A framework for combining multiple soil moisture retrievals based on maximizing temporal correlation. *Geophysical Research Letters* 2015, 42 (16), 2015GL064981.

4.1 Introduction

It is common to consider the temporal correlation of the satellite soil moisture and reference data sets to characterize temporal dynamics because it is insensitive to bias. In addition, scaling approaches have been developed to adjust the dynamic range of soil moisture (Liu et al., 2011b; Reichle et al., 2004; Yilmaz & Crow, 2013), the dynamic range of the two products being compared is not such an issue and therefore the temporal correlation can be considered as a suitable metric (Entekhabi et al., 2010b). To this end, this study develops a methodology to

improve the temporal correlation coefficients of satellite soil moisture products compared to a reference data.

As presented in Chapter 3 , the JAXA and LPRM algorithms share a common background in the radiative transfer model, but surface temperature, roughness and vegetation are treated differently within both algorithms and they also use different dielectric mixing models to convert the dielectric constant into soil moisture. Consequently, it was shown that the performance of the two products is complementary in many locations in terms of bias, RMSE and, most importantly correlation coefficients against a reliable reference (e.g. in situ measurements).

The main objective of this chapter is to present a method for improving temporal correlation with by linearly combining these AMSR2 soil moisture products of which performances are complementary. The method applies optimal weights to the AMSR2 products and then combines them into one product which has improved temporal correlation by leveraging the strengths of both the AMSR2 products. The weights are calculated based on statistics of two AMSR2 products; specifically, variance and correlation coefficients against a reference data set, and provide information on their relative strengths against each other.

4.2 Data

4.2.1 Soil Moisture Data Sources

This chapter used the daily JAXA (Ver. 1.0) and LPRM soil moisture products from AMSR2 at 0.25° for the two-year period from 1 August 2012 through 31 July 2014. It is used the AMSR2 soil moisture products based on the X-band brightness temperatures because it is available from both algorithms. This means that the proposed approach will be tested by combining two evenly matched products in terms of their ability to retrieve soil moisture rather than the advantages from the different wavelengths. However, there is no reason that the combination approach could not be used with C-band observations as well as the X-band ones.

The LPRM relies on an internal algorithm that retrieves LST by using brightness temperatures observed at the Ka-band (36.5 GHz) channel (Holmes et al., 2009); however, the JAXA algorithm assumes that surface temperature is constant throughout the year (Koike, 2013). To account for

frozen conditions, the corresponding AMSR2 soil moisture retrievals (both LPRM and JAXA) were masked out when the ERA-Interim soil temperature in the top soil layer (0 – 0.07m) was below freezing.

Table 4.1. Details of datasets used in this study.

Data source	Variable	Temporal resolution	Spatial resolution	Units
AMSR2-JAXA	Level 3 geophysical parameter SMC	Daily	0.25°	m^3/m^3
AMSR2-LPRM	Level 3 Surface Soil Moisture (X-band)	Daily	0.25°	m^3/m^3
AMSR2-LPRM	Vegetation optical depth (C-band)	Daily	0.25°	-
ERA-Interim	Soil water contents level 1 (0-0.07m depth)	6-hourly	0.25°	m^3/m^3
ERA-Interim	Soil temperature level 1 (0-0.07m depth)	6-hourly	0.25°	K
MERRA-Land	Top soil layer soil moisture consent (SFMC)	Hourly	0.25° (Resampled)	m^3/m^3
ISMN	In-situ measured soil moisture from 8 networks	Hourly	Point	m^3/m^3
ESA CCI	topographic complexity, wetland fraction	-	0.25°	%

To calibrate the weights for the combination approach, a reliable reference soil moisture product is required. This study used the volumetric water content in the topmost soil layer (0 - 0.07 m) from the ERA-Interim reanalysis at 0.25° as the reference. Data at the local satellite overpass time was obtained by linearly interpolating the 6-hourly reanalysis. The final combination depends on the choice of reference data set, which is an arbitrary but unavoidable choice. In this study, ERA-Interim was chosen because of its availability over the entire study period, global coverage and temporal consistency. However, it should be noted that the combination approach has the flexibility to be applied with any other appropriate reference data set such as other reanalysis or satellite soil moisture data. In case of using other satellite soil moisture data, products from the same sensor but different frequencies or algorithms are preferred. This is because the products will then have the same swath pattern, leading to a suitable number of paired observations and the same scan time, so that the weather conditions are constant. In consideration of this, the Modern-Era Retrospective Analysis for Research and

Applications Land (MERRA-Land) (Reichle et al., 2011) was also used as a reference for comparing with results using ERA-Interim.

To validate the combined product, an independent data source is required. In this work, ground based observations from the ISMN were used to evaluate the performance of the combined remotely sensed soil moisture product. The reason for using the ISMN was that the AMSR2 products were available more than two years at the time of writing (April 2015), covering more numbers of stations from 8 networks. For minimizing systematic differences from the various probe types and methods, a strict pre-process was implemented as described below. Details of data used for this study are summarized in Table 4.1.

4.2.2 Data Preprocessing

Several preprocessing steps are required to allow the three datasets (remote sensing, reanalysis and ground based observations) to be compared. Commonly applied masking procedures (De Jeu et al., 2008; Liu et al., 2011b) were adopted as follows to remove conditions unfavorable for soil moisture retrievals. First, land adjacent to the ocean or large lakes that may be sensitive to open water fluctuations was masked (*i.e.* center of grid cells within 25 km from the coast or large lake). Next, as the emitted soil (moisture) radiation is totally masked under dense canopies, these regions were excluded when annual mean vegetation optical depth is greater than 0.8 at 6.9 GHz from the LPRM (De Jeu et al., 2008). Results from the soil moisture retrievals from the night-time observations (descending satellite path) are presented here as they are generally considered to be of higher quality than the day-time observations (ascending satellite path) because of the equilibrium thermal conditions vegetation and near-surface soil (De Jeu et al., 2008). Results from these day-time observations are available in the Appendix.

For the ground-based observations, the data set was filtered to ensure only high-quality stations were used in validation. The ISMN provides access to 915 ground stations which have at least 365 days overlapping with the study period and measurement depths of 10 cm or shallower, available over 10 different monitoring networks, mainly located in the USA and Europe. Systematic differences between remotely sensed observations and the ISMN are known

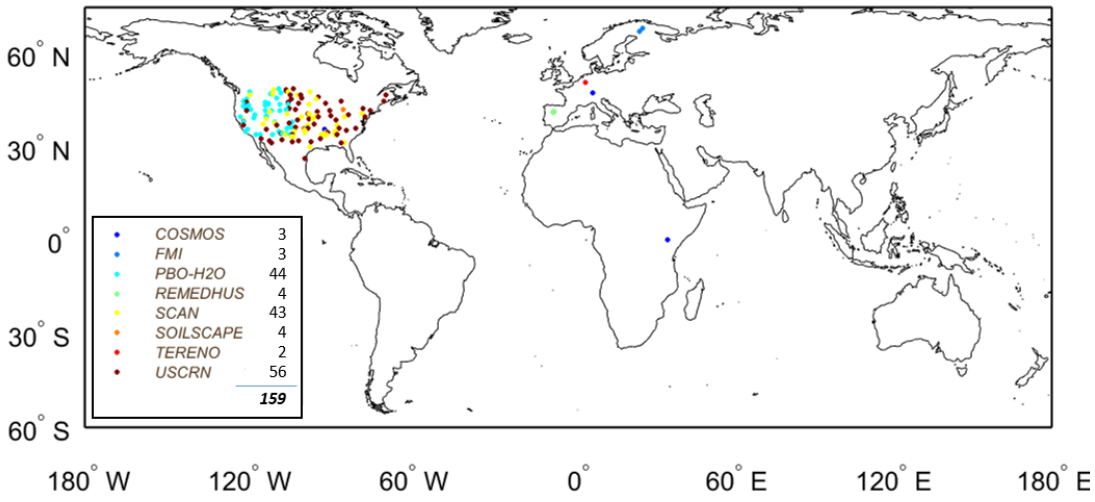


Figure 4.1. Locations of 159 *in situ* stations from 8 networks, selected from the ISMN, that were used for evaluating performances of combination using the soil moisture retrievals from the night-time observations (descending satellite path).

to be caused by their different measurement depths and the spatial heterogeneity of soil moisture that affects the comparison of point observations against coarse scale satellite footprint (Crow et al., 2012; Gruber et al., 2013). To overcome the first issue, it was only considered *in situ* measurements from the shallowest soil layer in locations with multiple measurement depths (*e.g.* 10 cm and 5 cm) and applied the standard quality control (QC) flags from the ISMN (Dorigo et al., 2013). These QC flags exclude doubtful observations such as spikes and sudden breaks in the ground based observations. Second, to evaluate the suitability of the ground stations for satellite validation purposes, topographic complexity (TC) and wetland fraction (WF) data from the European Space Agency Climate Change Initiative (ESA CCI) (Liu et al., 2012; Liu et al., 2011b; Wagner et al., 2012) were used. Only ground stations in grids with low WF (<10%) and TC (<10%) were considered for further analysis as these are known as factors contaminating the microwave signal (Draper et al., 2012). Furthermore, a simple test was performed to check the spatial representativeness of two or more ground-based stations within a coarse scale satellite footprint according to Dorigo et al. (2014). The area representativeness was defined as the average of three correlation coefficients between *in situ* data and ERA-Interim data ($R_{ERA\text{-}in\ situ}$), the JAXA and LPRM products ($R_{JAXA\text{-}in\ situ}$ and $R_{LPRM\text{-}in\ situ}$). The station with

the highest area representativeness was selected. Additionally, when two or more correlation coefficients including $R_{ERA\text{-}in\ situ}$ are significantly ($p \leq 0.05$) negative at a station, the station was considered as unrepresentative and discarded. The Snow Telemetry (SNOTEL) and Atmospheric Radiation Measurement (ARM) networks were excluded since their primary purposes are not soil moisture measurements. For the soil climate analysis network (SCAN), only stations which have suitable topography and vegetation characteristics, and which were used for the Soil Moisture and Ocean Salinity (SMOS) data assimilation experiments (De Lannoy & Reichle, 2015), were adopted in this study. Finally, to ensure statistically robust results, stations that have at least 100 observations over the entire study period have been only used. The geographic distribution of the selected 159 stations over eight networks used in this study is shown in Figure 4.1.

4.3 Methodology

The original linear combination of forecasts was developed by Bates and Granger (1969) and created a product that minimized the mean square error (MSE) from two parent forecasts. The rationale behind forecast combination is that if one forecast is based on variables or information that the other forecast has not considered, or/and the forecast makes a different assumption about the form of the relationship between the variables, then the combined product is likely to have lower overall error than any of the individual components. A follow-up study (Granger & Ramanathan, 1984) showed that the idea could be extended to multiple forecasts and the weights optimized by using a restricted least squares regression that assumes the total sum of weights is restricted to 1. During the past decades, the combination concept has been widely applied to various disciplines dealing with timeseries of forecasts or other products (Clemen, 1989; Timmermann, 2006; Wasko et al., 2013). For a recent example, Khan et al. (2014) proposed an approach combining five global sea surface temperature forecasts by seasonality-based dynamic weighting factors. The following is a summary of the original forecast combination. Given two sets of unbiased forecasts or products ($\boldsymbol{F_1}$ and $\boldsymbol{F_2}$, $n\times1$) a combination ($\boldsymbol{F_c}$) can be expressed with a weight (w), ranging from 0 to 1, as

$$\boldsymbol{F_c} = w\boldsymbol{F_1} + (1-w)\boldsymbol{F_2} \quad \text{Eq. 4.1}$$

where *n* is the length of the forecasts. The optimal weight *w*, which leads to the minimum MSE for the combined F_c against a reference, is expressed as

$$w = \frac{\sigma_{\varepsilon_2}^2 - \rho_\varepsilon \sigma_{\varepsilon_1} \sigma_{\varepsilon_2}}{\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 - 2\rho_\varepsilon \sigma_{\varepsilon_1} \sigma_{\varepsilon_2}} \quad \text{Eq. 4.2}$$

where $\sigma_{\varepsilon_1}^2$ and $\sigma_{\varepsilon_2}^2$ are the error variances of $\boldsymbol{F_1}$ and $\boldsymbol{F_2}$ against the reference respectively, and ρ_ε is the correlation between the two sets of errors.

As the remotely sensed soil moisture products are retrieved by various algorithms which are based on different variables, information and assumptions (Kim et al., 2015a), it is valid to extend the combination scheme to soil moisture datasets. As discussed previously, the temporal correlation coefficient is the most important indicator of the utility of remotely sensed soil moisture products. However existing combination approaches focus on the minimization of the MSE. Minimizing the MSE will also improve the temporal correlation coefficients because MSE consists of three components contributed by bias (differences in temporal mean), variance (dynamic range) and correlations (temporal pattern) (Su et al., 2013b). However, given the importance of the temporal patterns in soil moisture analysis, it is aimed to answer the following question: Can a combination framework to solely maximize the temporal correlation coefficients rather than minimizing MSE and applied to the remotely sensed soil moisture products?

One can consider that two vectors of unbiased soil moisture retrievals $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ (*n*×1) are linearly combined into $\boldsymbol{\theta_c}$ by applying a weighting factor *w*, 0 to 1 such that

$$\boldsymbol{\theta_c} = w\boldsymbol{\theta_1} + (1-w)\boldsymbol{\theta_2} \quad \text{Eq. 4.3}$$

The Pearson correlation coefficient (R) between $\boldsymbol{\theta_c}$ and a reference ($\boldsymbol{\theta_{Ref}}$) can be expressed as a function of *w* according to the definition of *R* and Eq. 4.3, and this is an optimization problem according to

$$\text{Maximize } R = f(w) = \frac{E(\boldsymbol{\theta_c} - \mu_C)(\boldsymbol{\theta_{Ref}} - \mu_{Ref})}{\sigma_c \sigma_{Ref}} \quad \text{Eq. 4.4}$$

$$\text{Subject to } 0 \leq w \leq 1$$

where μ_C and μ_{Ref} are the mean values, and σ_c and σ_{Ref} are the standard deviations of $\boldsymbol{\theta_c}$ and $\boldsymbol{\theta_{Ref}}$ respectively. For this case with only two parent products which have positive correlation coefficients against the reference, differentiating Eq. 4.4 with respect to *w*, the maximum R between the combined product and the reference is found when:

$$w = \frac{\sigma_2(R_{1\cdot Ref} - R_{1\cdot 2} \cdot R_{2\cdot Ref})}{\sigma_1(R_{2\cdot Ref} - R_{1\cdot 2} \cdot R_{1\cdot Ref}) + \sigma_2(R_{1\cdot Ref} - R_{1\cdot 2} \cdot R_{2\cdot Ref})} \quad \text{Eq. 4.5}$$

where each σ presents the standard deviation of each product, and R is the temporal correlation coefficient between two products. The details of the derivation of Eq. 4.5 are in the a appendix of this chapter. When applying the weighting factor which is calculated by Eq. 4.5 and in the constrained range (0 to 1), the correlation coefficient between the combined product and the reference will always be larger than or equal to the correlation coefficients between each original product and the reference. In the case of a negative correlation coefficient for either of the parent products, the weights can be optimized numerically to maximize R and ensure that the weights are between zero and one.

To apply the proposed methodology, the first requirement is to remove the systematic differences among the datasets. This was achieved by normalizing the two parent products (*i.e.* JAXA and LPRM) against the chosen reference through (Draper et al., 2009).

$$\boldsymbol{\theta}_{\mathbf{NORMAL}} = (\boldsymbol{\theta}_{\mathbf{RAW}} - \overline{\theta_{\mathrm{RAW}}}) \times \frac{\mathrm{std}(\theta_{\mathrm{REF}})}{\mathrm{std}(\theta_{\mathrm{RAW}})} + \overline{\theta_{\mathrm{REF}}} \quad \text{Eq. 4.6}$$

where, $\boldsymbol{\theta_{NORMAL}}$ is the normalized soil moisture, $\boldsymbol{\theta_{RAW}}$, raw soil moisture, $\boldsymbol{\theta_{REF}}$, reference soil moisture, $\bar{\theta}$, mean of $\boldsymbol{\theta}$, and std, standard deviation. When the two parent products are normalized against the same reference, Eq. 4.5 can be simplified to

$$w = \frac{R_{1\cdot Ref} - R_{1\cdot 2} \cdot R_{2\cdot Ref}}{R_{2\cdot Ref} - R_{1\cdot 2} \cdot R_{1\cdot Ref} + R_{1\cdot Ref} - R_{1\cdot 2} \cdot R_{2\cdot Ref}} \qquad \text{Eq. 4.7}$$

The optimal weighting factor is calculated through Eq. 4.7. Finally, the parent products are combined into a single product by using the calculated optimal weighting factor.

4.4 Results and Discussion

4.4.1 Global Optimal Weighting Factors

The proposed correlation-based combination approach was applied to the remotely sensed soil moisture products from AMSR2 over the 2-year study period. As discussed, the volumetric water content in the topmost soil layer (0 - 0.07 m) of ERA-Interim was chosen as the reference data set. Based on this reference data, the global optimal weighting factors were determined using the approach presented in the previous section. The global map of the optimal weighting factors for the two remotely sensed soil moisture products from AMSR2 is presented in Figure 4.2.

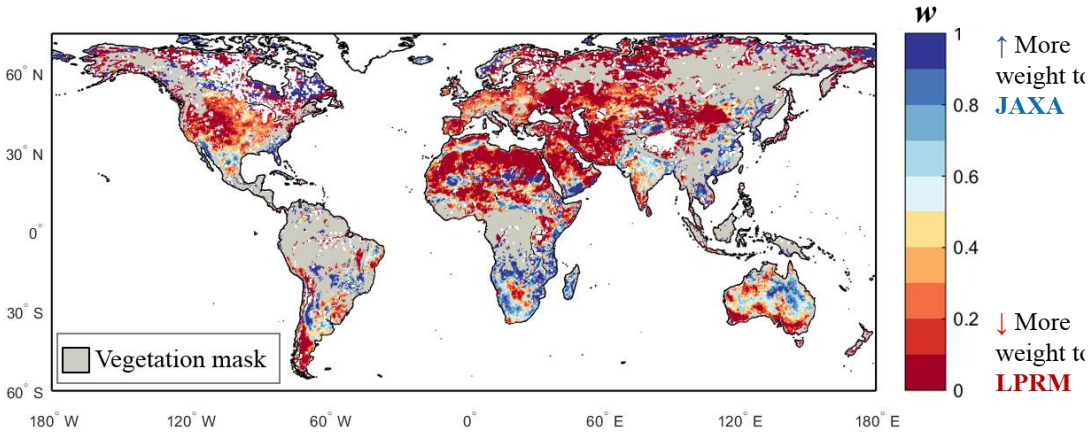


Figure 4.2. The spatial distribution of the optimal weights for the JAXA and LPRM soil moisture products using ERA Interim as the reference.

The dark red color indicates that most the weight comes from the LPRM product whereas the dark blue color is where the JAXA product is most effective. For the in-between colors of the combination product are a more even mix of the two parent products. This global map of optimal weights provides information on the relative strengths (and weaknesses) of the remotely sensed

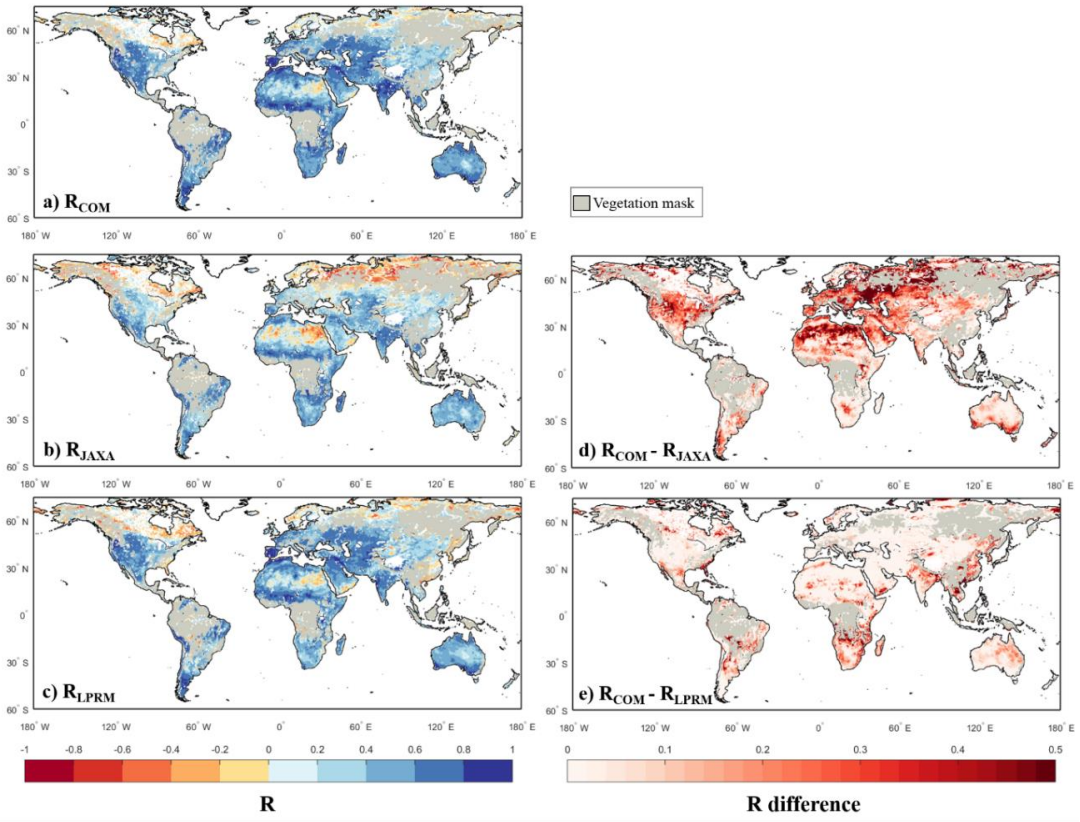


Figure 4.3. Spatial distribution of Pearson's correlation coefficients between the reference (ERA-Interim) and a) the combined product (R_{COM}), b) the JAXA product (R_{JAXA}) and c) the LPRM product (R_{LPRM}). Panel d) shows the differences in between correlation coefficients of the combined and JAXA products (R_{COM} minus R_{JAXA}), and e), the combined and LPRM products (R_{COM} minus R_{LPRM}).

soil moisture products against each other under the assumption that ERA-Interim reanalysis soil moisture represents the true soil moisture. As shown in Figure 4.2, the JAXA product has strengths over humid subtropical regions, such as the southeastern USA, northern India and the southeastern Asia. The LPRM product generally shows strengths in the more temperate areas which is in line with the results from previous studies (Crow et al., 2010; Dorigo et al., 2010).

After applying the weighting factors, the two AMSR2 products are combined into a single product, and correlation coefficients of the combination product and the two parent products against the reference are presented in Figure 4.3. The correlation coefficients for the combined product are higher in all cases than either of the parent products (Figure 4.3d and e), demonstrating that the combination method assures improved (or at least equal) performance.

The global mean of correlation coefficients is 0.35 for the JAXA product, 0.45 for the LPRM product and 0.52 for the combined product.

Again, the results and actual values of the weights in Figure 4.2 and Figure 4.3 depend on the choice of reference data set which is an assumed truth. The optimal combination is considered independently for each location so a reference data set with smaller spatial domain or even at a point would still allow a combination product to be constructed. The presented combination approach is general enough to be applied with any other appropriate reference data set, which could be reanalysis, ground based or remotely sensed data sets, and should therefore be considered as a flexible tool that can be applied given a chosen reference data set. The analysis was repeated using MERRA-Land top soil layer soil moisture content (SFMC) - for a sensitivity analysis. As presented Table A4.1, it showed (somewhat) different performances in the temporal correlation (mean R) and weights (mean w) but these differences simply reflect the relative differences (and similarities) of the chosen references. Combination results using MERRA-Land as the reference are summarized in the appendix through Figure A4.5 to Figure A4.10. Also, results based on cross validating the combination product against a different reference product are presented through Figure A4.11 to Figure A4.14. Even in cross validation the proposed temporal combination approach provides superior results compared to either of the parent products.

4.4.2 Evaluation Against *In Situ* Soil Moisture

The global maps of correlation coefficients (Figure 4.3) show an improved match for the combined product compared to the two individual remotely sensed soil moisture products. Of interest is how the combination approach compares to an independent validation dataset. To achieve this goal, the two remotely sensed soil moisture products (JAXA and LPRM) and the combination were evaluated through the set of ground based observations from the 159 stations. Figure 4.4 presents the evaluation results.

The rationale of the scatterplot (Figure 4.4a) is that stations that fall below the 1:1 line (red line) indicate locations where the performance of the combined product is better than the

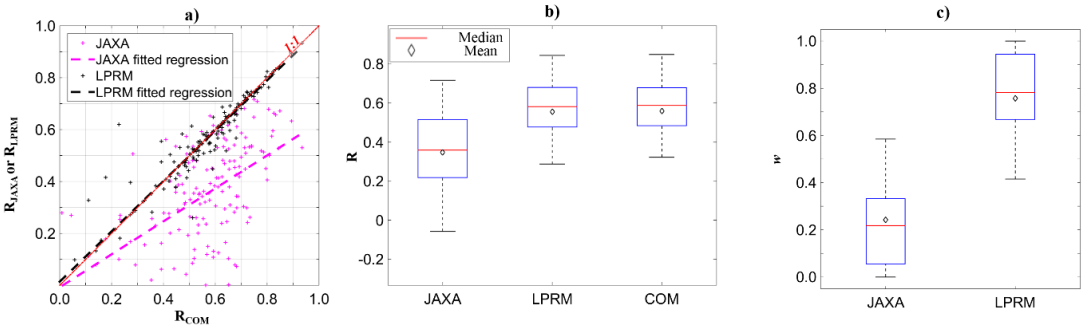


Figure 4.4. Results for evaluating improvements in correlation coefficients through combinations. a) Scatter plot showing correlation coefficients of the JAXA and LPRM products (R_{JAXA} and R_{LPRM} on y-axis respectively) against correlation coefficients of the combined product (R_{COM} on x-axis). b) Boxplots for three sets of correlation coefficients for the JAXA, LPRM and combined products against the reference. c) Boxplot for weighting factors (*w*) from all *in situ* stations.

parent product. The dashed lines are fitted linear regressions for the JAXA (magenta) and LPRM (black) products and present the overall performance of the combined product versus the parent products. For both products, the combination is generally better than either product when considered over all the validation ground stations. In Figure 4.4b, the mean R_{JAXA} is 0.35 and the mean R_{LPRM} is 0.56 over the *in situ* stations, and after combining the two AMSR2 products at each location, the mean R_{COM} is calculated as 0.56. The average improvements compared to the LPRM product are marginal but this is most likely due to the uneven distribution of the selected *in situ* stations. As shown in Figure 4.1, they are mostly distributed over the USA and the northern Europe where the LPRM product shows better performances than the JAXA product (Figure 4.2). If there was sufficient ground data in regions where the JAXA product shows better performance, such as the southeastern USA, northern India and the southeastern Asia, more balanced results would be expected. The mean weights applied to each product further support this finding. When using the ERA-Interim as the reference, the mean weight for the JAXA product over the grid cells with *in situ* stations is 0.24 (Figure 4.4c) whereas its global mean weight is 0.35. In addition, over the USA (30°N-50 °N, 70°W-125°W), the regional mean weight is 0.26 for the JAXA product. This shows that the JAXA performance is generally worse in the areas well sampled by the *in situ* data. More improvements are evident in the ascending

data where the LPRM product has relatively lower quality than the descending data (De Jeu et al., 2008). The results from the ascending data are presented and summarized in the appendix.

The points above the 1:1 line in Figure 4.4a represent locations where the correlation with the validation data is degraded after the combination. This is a result of significant differences between the ERA-Interim data and the *in situ* data at these locations. In general, however, under both the calibration and validation settings the correlations for the combined product are improved compared to both the AMSR2 products.

4.5 Conclusions

In this study, a correlation-based combination approach was developed which maximizes the temporal correlation coefficient of soil moisture retrievals. This method was applied to two remotely sensed soil moisture products from the AMSR2 sensor over a 2-year study period. Based on the method developed here, the JAXA and LPRM soil moisture products were combined into a single product that showed superior results against the chosen reference soil moisture dataset after applying the optimal weighting factors (Figure 4.3). The spatial distribution of optimal weighting factors provides information on the relative strengths of the JAXA and LPRM products at the global scale. These weighting factors provide an important step forward in terms of how best to combine soil moisture products and improve over taking a simple average of the two products (Liu et al., 2011b). The performance of the combined product was validated against *in situ* observations from the ISMN. The combined product showed a significant improvement compared to the JAXA product and marginal improvements compared to the LPRM product. An important factor to consider is the unevenly distributed *in situ* stations for the evaluation due to the ISMN coverage.

The actual weighting factors presented depend on the choice of reference data (ERA-Interim in this study) and different results would be obtained if a different reference data set was chosen. Sensitivity testing with MERRA-Land demonstrated that the method continues to perform well even with a different reference data set choice. The main contribution of this work is to provide a method that can be used to improve soil moisture products compared to any choice of

reference data set. And of particular relevance for soil moisture, the new method focuses on temporal correlations which have not previously considered in combination approaches. The comparison of the combined product using *in situ* stations showed that improvements in correlations were also obtained when considered in a validation setting. The uneven distribution of *in situ* stations due to the ISMN coverage led to small but consistent improvements for the combined product over the LPRM product and larger improvements for the JAXA product; a different set of *in situ* stations would lead to different results.

Even though the new correlation-based combination approach has been demonstrated to operate robustly, there are a few future extensions that could provide even better combined products. First, as demonstrated with the MERRA-Land analyses it is useful to test different reference datasets to understand the sensitivities in the combined product. Second, the linear rescaling method (Draper et al., 2009) was used in this study to only adjust the mean and standard deviation of satellite soil moisture datasets against the reference. It may be worth investigating in future work whether other rescaling methods (*e.g.* Cumulative Distribution Function matching (Brocca et al., 2011; Liu et al., 2011b)) could lead to further improvements to the combined satellite product. The demonstration of the concept here has used just two parent datasets but there is no limit to the number of products that could be combined. In the context of soil moisture this is an exciting area of future research due to the number of satellite-based soil moisture products from different remote sensing techniques or different retrieval algorithms that are being developed. Any combined product would be expected to reflect the varying strengths of these techniques and algorithms. Another area for potential development is time-varying weights as well as the spatial weighting developed in this paper. Dynamic weights would take account of the time-varying performances of different soil moisture products due to the time-varying land surface features (e.g. soil temperature and vegetation). If weights can be updated in close to real-time such a development could be particularly useful for operational flood forecasting or other forecast problems. However, the reference datasets used for the linear combination (i.e. ERA-Interim and MERRA-Land) are not available at the time of soil

moisture retrievals from AMSR2, and it usually takes a few months to be available. Therefore, it is necessary to develop a method estimating optimal weights without such reference data but using other information available in near real-time such as regional vegetation and temperature. Lastly, it is required to investigate how the optimal weights are related to the important components in the retrieval algorithms such as soil temperature, surface roughness, vegetation and soil moisture status as mentioned in Chapter 3 . With this, data users or algorithms developers can understand better which product is preferred in time and space and why.

4.6 Appendix

Derivation of equation for optimal weighting factor (Eq. 4.5)

Two sets of unbiased soil moisture retrievals $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ (n×1) are linearly combined into $\boldsymbol{\theta_c}$ by applying a weighting factor w, 0 to 1.

$$\boldsymbol{\theta_c} = w\boldsymbol{\theta_1} + (1-w)\boldsymbol{\theta_2} \tag{1}$$

The Pearson correlation coefficient (R) between $\boldsymbol{\theta_c}$ and a reference ($\boldsymbol{\theta_R}$) can be expressed as a function of w according to the definition of R and equation (3), and this is an optimization problem of the following function.

$$\begin{gathered}\text{Maximize } R = f(w) = \frac{E[(\boldsymbol{\theta_c} - \mu_C)(\boldsymbol{\theta_R} - \mu_R)]}{\sigma_c \sigma_R} \\ \text{Subject to } 0 \leq w \leq 1\end{gathered} \tag{2}$$

Where μ_C and μ_R are the mean values, and σ_c and σ_R are the standard deviations of $\boldsymbol{\theta_c}$ and $\boldsymbol{\theta_R}$ respectively. From equation (1),

$$\begin{aligned}\mu_C &= \mathrm{E}[\boldsymbol{\theta_C}] \\ &= \mathrm{E}[w\boldsymbol{\theta_1} + (1-w)\boldsymbol{\theta_2}] \\ &= w\mu_1 + (1-w)\mu_2\end{aligned} \tag{3}$$

$$\begin{aligned}\sigma_c^2 &= \mathrm{Var}(\boldsymbol{\theta_C}) \\ &= \mathrm{E}[(\boldsymbol{\theta_c} - \mu_C)^2] \\ &= \mathrm{E}\left[\boldsymbol{\theta_C^2}\right] - \mu_C{}^{\boldsymbol{2}} \\ &= \mathrm{E}[(w\boldsymbol{\theta_1} + (1-w)\boldsymbol{\theta_2})^2] - (w\mu_1 + (1-w)\mu_2)^2 \\ &= [\mathrm{Var}(\boldsymbol{\theta_1}) + \mathrm{Var}(\boldsymbol{\theta_2}) - 2Cov(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})] \cdot w^2 - 2 \\ &\qquad \cdot [Var(\boldsymbol{\theta_1}) - Cov(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})] \cdot w + \mathrm{Var}(\boldsymbol{\theta_2})\end{aligned} \tag{4}$$

Therefore, from equation (3) and (4),

$$f(w) = \frac{E[(\boldsymbol{\theta_c} - \mu_C)(\boldsymbol{\theta_R} - \mu_R)]}{\sigma_c \sigma_R}$$

$$= \frac{[Cov(\boldsymbol{\theta_1}, \boldsymbol{\theta_R}) - Cov(\boldsymbol{\theta_2}, \boldsymbol{\theta_R})] \cdot w + Cov(\boldsymbol{\theta_2}, \boldsymbol{\theta_R})}{\sqrt{\begin{matrix}\mathrm{Var}(\boldsymbol{\theta_R}) \cdot [\mathrm{Var}(\boldsymbol{\theta_1}) + \mathrm{Var}(\boldsymbol{\theta_2}) - 2Cov(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})] \cdot w^2 \ldots \\ -2 \cdot \mathrm{Var}(\boldsymbol{\theta_R}) \cdot [Var(\boldsymbol{\theta_1}) - Cov(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})] \cdot w + \mathrm{Var}(\boldsymbol{\theta_R}) \cdot \mathrm{Var}(\boldsymbol{\theta_2})\end{matrix}}} \quad (5)$$

$$= \frac{A \cdot w + B}{\sqrt{C \cdot w^2 + D \cdot w + E}}$$

Where,

$\mathrm{A} = Cov(\boldsymbol{\theta_1}, \boldsymbol{\theta_R}) - Cov(\boldsymbol{\theta_2}, \boldsymbol{\theta_R})$

$\mathrm{B} = Cov(\boldsymbol{\theta_2}, \boldsymbol{\theta_R})$

$\mathrm{C} = \mathrm{Var}(\boldsymbol{\theta_R}) \cdot [\mathrm{Var}(\boldsymbol{\theta_1}) + \mathrm{Var}(\boldsymbol{\theta_2}) - 2Cov(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})]$

$\mathrm{D} = -2 \cdot \mathrm{Var}(\boldsymbol{\theta_R}) \cdot [Var(\boldsymbol{\theta_1}) - Cov(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})]$

$\mathrm{E} = \mathrm{Var}(\boldsymbol{\theta_R}) \cdot \mathrm{Var}(\boldsymbol{\theta_2})$

Differentiating equation (5) with respect to *w*,

$$f'(w) = \frac{A}{\sqrt{C \cdot w^2 + D \cdot w + E}} - \frac{(B + A \cdot w)(D + 2 \cdot C \cdot w)}{2 \cdot (C \cdot w^2 + D \cdot w + E)^{2/3}} \quad (6)$$

Therefore, the optimal weighting factor is calculated by letting equation (6) 0 and simplified as

$$w = \frac{2 \cdot A \cdot E - B \cdot D}{2 \cdot B \cdot C - A \cdot D}$$
$$= \frac{\sigma_2(\rho_{1R} - \rho_{12} \cdot \rho_{2R})}{\sigma_1(\rho_{2R} - \rho_{12} \cdot \rho_{1R}) + \sigma_2(\rho_{1R} - \rho_{12} \cdot \rho_{2R})} \quad (7)$$

Table A4.1. Combination results in terms of mean R and *w* from the ascending and descending data with ERA-Interim and MERRA-Land as references. Generally, the combination performance is more prominent for the JAXA and the ascending data than the LPRM and the descending data

Scale	**Statistics**	**Product**	**Descending**		**Ascending**	
			ERA-Interim	**MERRA-Land**	**ERA-Interim**	**MERRA-Land**
Global	Mean R	Combined	0.52	0.51	0.50	0.43
		JAXA	0.35	0.31	0.33	0.26
		LPRM	0.45	0.44	0.36	0.30
	Mean *w*	JAXA	0.37	0.31	0.46	0.43
		LPRM	0.63	0.69	0.54	0.57
In situ stations	Mean R	Combined	0.56	0.45	0.53	0.52
		JAXA	0.35	0.34	0.34	0.34
		LPRM	0.56	0.45	0.49	0.49
	Mean *w*	JAXA	0.24	0.22	0.26	0.29
		LPRM	0.76	0.78	0.74	0.71

Table A4.2. Cross-validation results (*i.e.* correlation coefficients) from combination using the ascending and descending data with ERA-Interim and MERRA-Land which are interchangeably used for combination and validation.

Satellite path	**Descending**		**Ascending**	
Combination Reference	ERA-Interim	MERRA-Land	ERA-Interim	MERRA-Land
Validation Reference	MERRA-Land	ERA-Interim	MERRA-Land	ERA-Interim
Combined	0.47	0.49	0.39	0.46
JAXA	0.31	0.35	0.26	0.33
LPRM	0.44	0.45	0.30	0.36
Remark	Figure A4.11	Figure A4.12	Figure A4.13	Figure A4.14

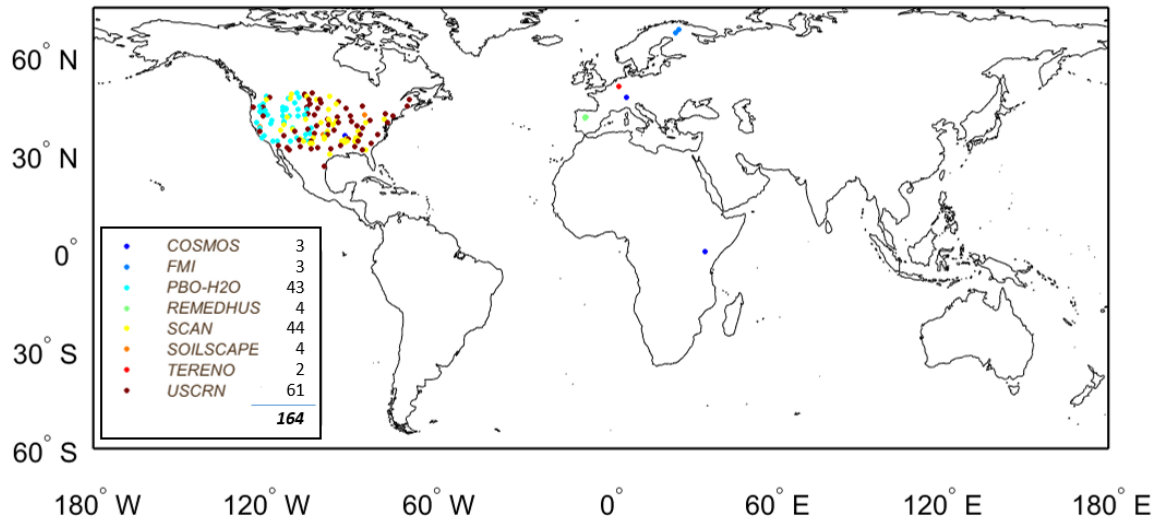


Figure A4.1. Locations of 164 *in situ* stations from 8 networks, selected from the ISMN, that were used for evaluating performances of combination using the soil moisture retrievals from the day-time observations (ascending satellite path). The reason for the different number of stations compared to Figure 4.1 is that number of observations at day-time is generally larger than ones at night-time due to moderately higher temperature reducing data mask.

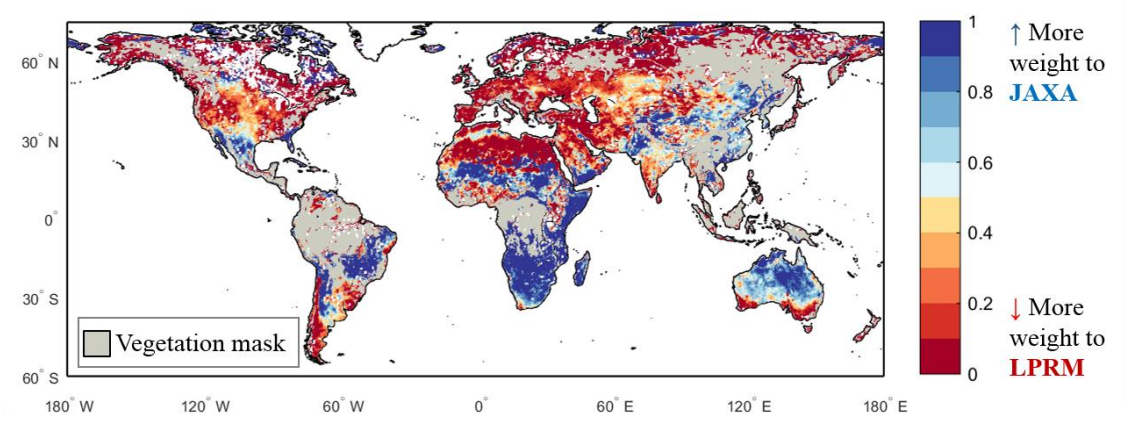


Figure A4.2. The spatial distribution of the optimal weights for the JAXA and LPRM soil moisture products at the day-time (ascending satellite path) using ERA-Interim soil water contents level 1 as the reference.

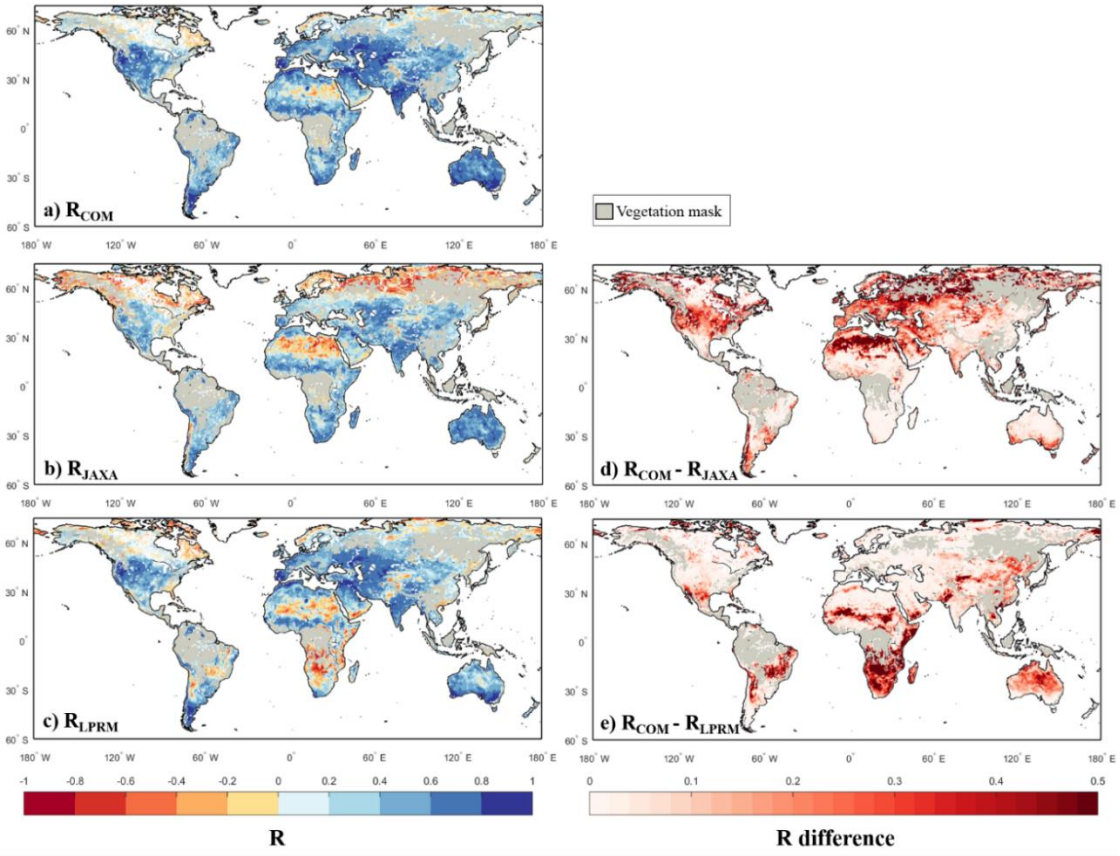


Figure A4.3. Results of combination using datasets at the ascending satellite path with ERA-Interim soil water contents level 1 as the reference: Spatial distribution of Pearson's correlation coefficients between the reference and a) the combined product (R_{COM}), b) the JAXA product (R_{JAXA}) and c) the LPRM product (R_{LPRM}). Where, the global mean of R_{COM} is 0.50, R_{JAXA}, 0.33 and R_{LPRM}, 0.36 respectively. Panel d) shows the differences in between correlation coefficients of the combined and JAXA products (R_{COM} minus R_{JAXA}), and e), the combined and LPRM products (R_{COM} minus R_{LPRM}).

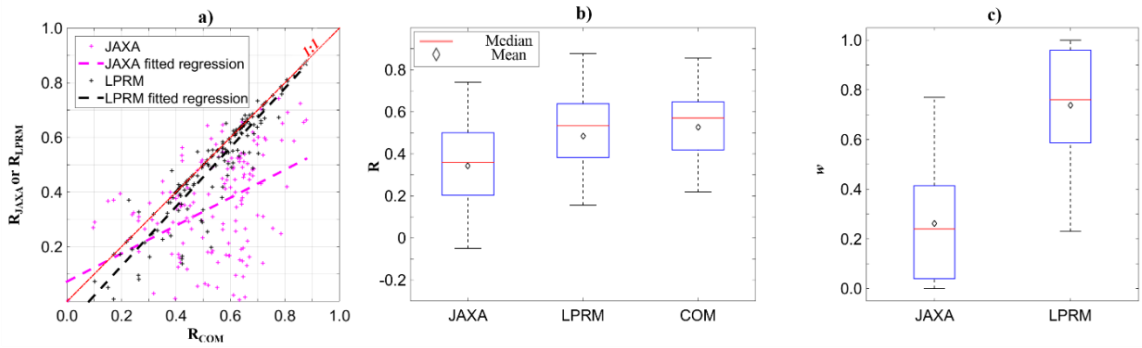


Figure A4.4. Results for evaluating improvements in correlation coefficients through combinations of data at the ascending satellite path using ERA-Interim soil water contents level 1 as the reference. a) Scatter plot showing correlation coefficients of the JAXA and LPRM products (R_{JAXA} and R_{LPRM} on y-axis respectively) against correlation coefficients of the combined product (R_{COM} on x-axis). b) Boxplots for three sets of correlation coefficients for the JAXA, LPRM and combined products against the reference. Where, the mean of correlation coefficients for the JAXA product is 0.34, the LPRM product, 0.49 and the combined product, 0.53 respectively. c) Boxplot for weighting factors (*w*) from all *in situ* stations. Where, the mean of weighting factors is 0.26 for the JAXA product and 0.74 for the LPRM product.

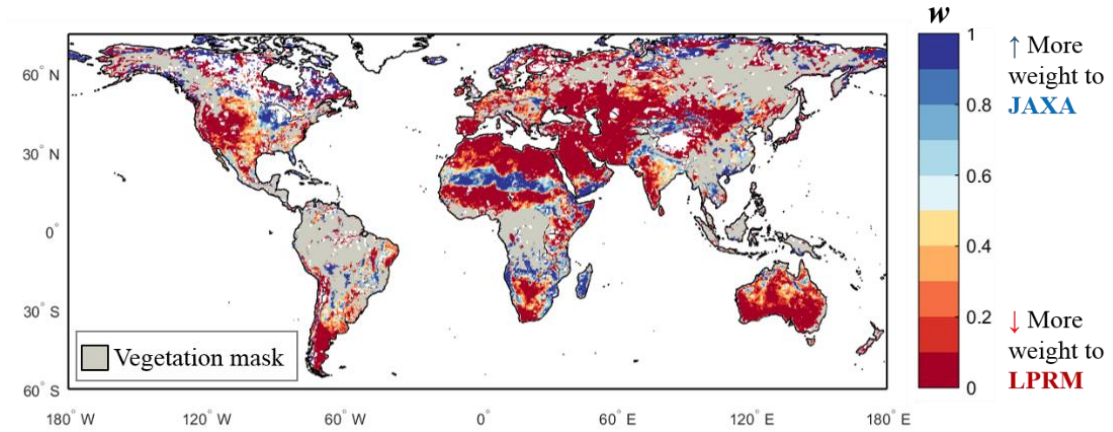


Figure A4.5. The spatial distribution of the optimal weights for the JAXA and LPRM soil moisture products at the night-time (descending satellite path) using MERRA-Land top soil layer soil moisture consent as the reference.

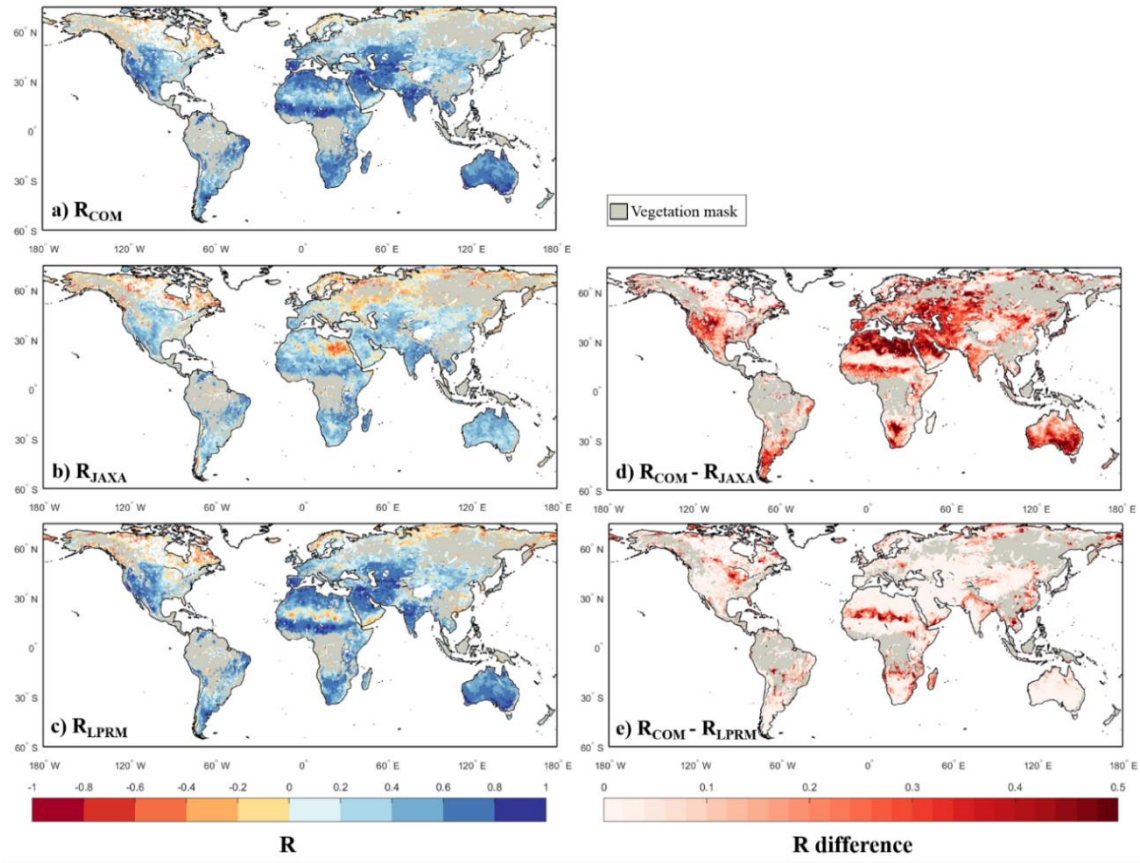


Figure A4.6. Results of combination using datasets at the descending satellite path with MERRA-Land top soil layer soil moisture content as the reference: Spatial distribution of Pearson's correlation coefficients between the reference and a) the combined product (R_{COM}), b) the JAXA product (R_{JAXA}) and c) the LPRM product (R_{LPRM}). Where, the global mean of R_{COM} is 0.51, R_{JAXA}, 0.31 and R_{LPRM}, 0.44 respectively. Panel d) shows the differences in between correlation coefficients of the combined and JAXA products (R_{COM} minus R_{JAXA}), and e), the combined and LPRM products (R_{COM} minus R_{LPRM}).

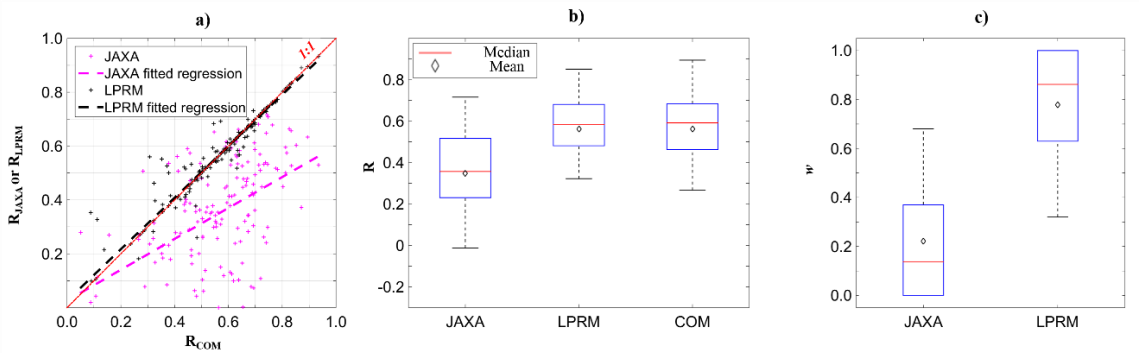


Figure A4.7. Results for evaluating improvements in correlation coefficients through combinations of data at descending satellite path using MERRA-Land top soil layer soil moisture content as the reference. a) Scatter plot showing correlation coefficients of the JAXA and LPRM products (R_{JAXA} and R_{LPRM} on y-axis respectively) against correlation coefficients of the combined product (R_{COM} on x-axis). b) Boxplots for three sets of correlation coefficients for the JAXA, LPRM and combined products against the reference. Where, the mean of correlation coefficients for the JAXA product is 0.34, the LPRM product, 0.45 and the combined product, 0.45 respectively. c) Boxplot for weighting factors (*w*) from all *in situ* stations. Where, the mean of weighting factors is 0.22 for the JAXA product and 0.78 for the LPRM product.

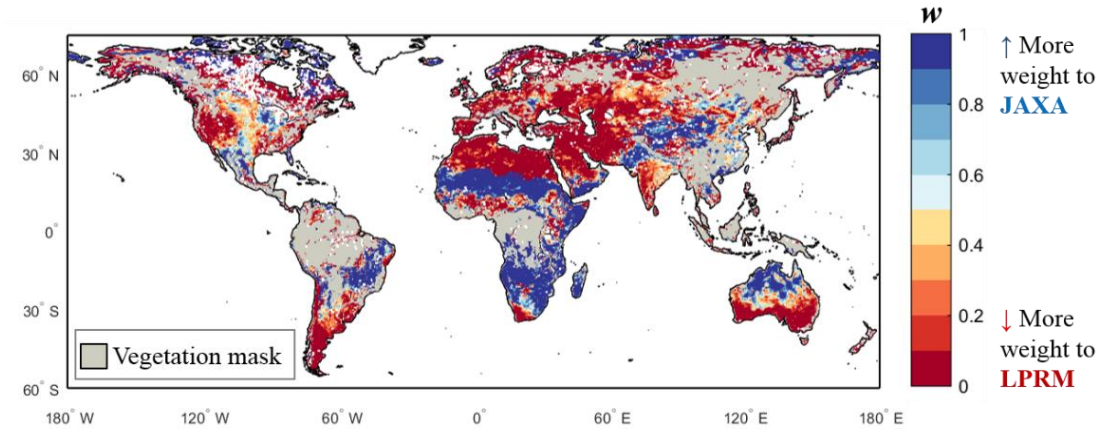


Figure A4.8. The spatial distribution of the optimal weights for the JAXA and LPRM soil moisture products at the day-time (ascending satellite path) using MERRA-Land top soil layer soil moisture consent as the reference.

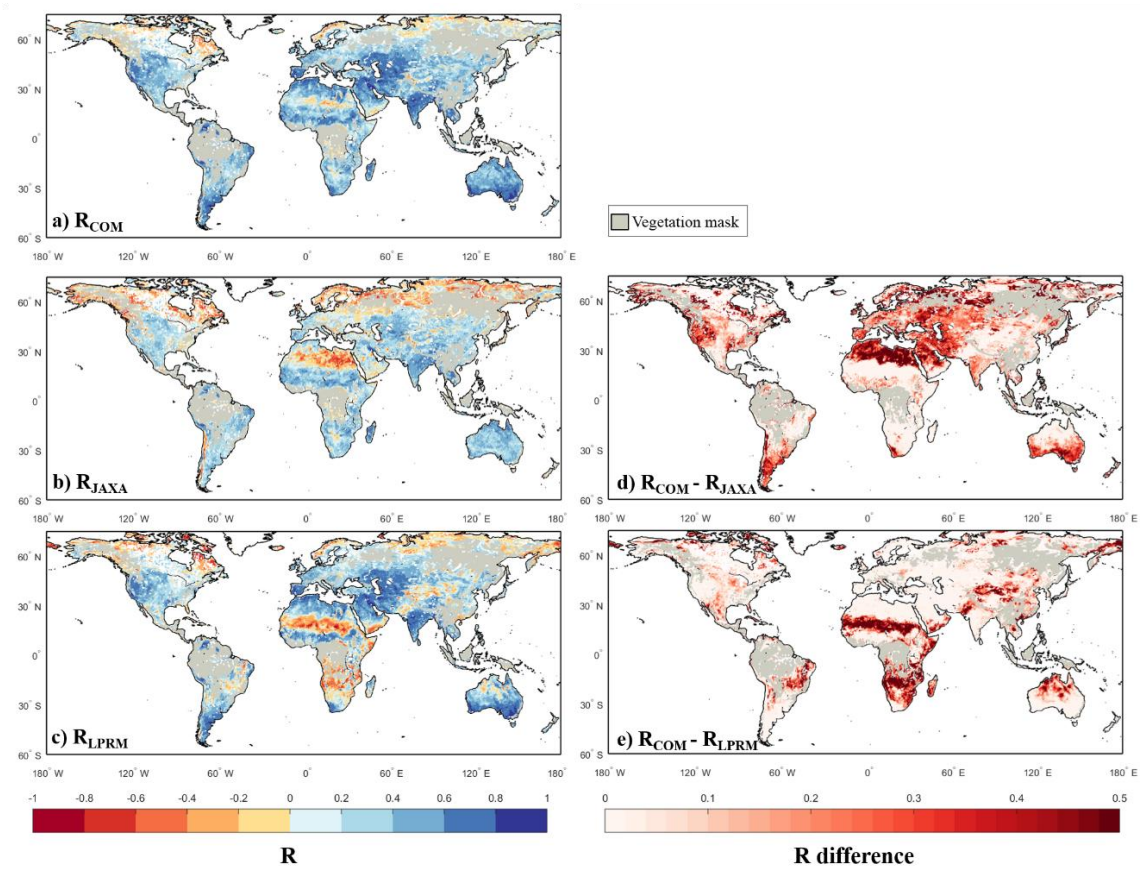


Figure A4.9. Results of combination using datasets at the ascending satellite path with MERRA-Land top soil layer soil moisture content as the reference: Spatial distribution of Pearson's correlation coefficients between the reference and a) the combined product (R_{COM}), b) the JAXA product (R_{JAXA}) and c) the LPRM product (R_{LPRM}). Where, the global mean of R_{COM} is 0.43, R_{JAXA}, 0.26 and R_{LPRM}, 0.30 respectively. Panel d) shows the differences in between correlation coefficients of the combined and JAXA products (R_{COM} minus R_{JAXA}), and e), the combined and LPRM products (R_{COM} minus R_{LPRM}).

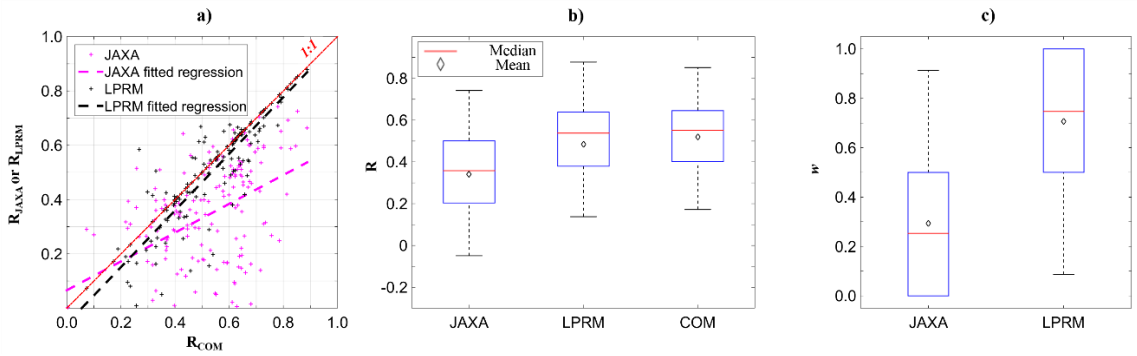


Figure A4.10. Results for evaluating improvements in correlation coefficients through combinations of data at descending satellite path using MERRA-Land top soil layer soil moisture content as the reference. a) Scatter plot showing correlation coefficients of the JAXA and LPRM products (R_{JAXA} and R_{LPRM} on y-axis respectively) against correlation coefficients of the combined product (R_{COM} on x-axis). b) Boxplots for three sets of correlation coefficients for the JAXA, LPRM and combined products against the reference. Where, the mean of correlation coefficients for the JAXA product is 0.34, the LPRM product, 0.49 and the combined product, 0.52 respectively. c) Boxplot for weighting factors (*w*) from all *in situ* stations. Where, the mean of weighting factors is 0.29 for the JAXA product and 0.71 for the LPRM product.

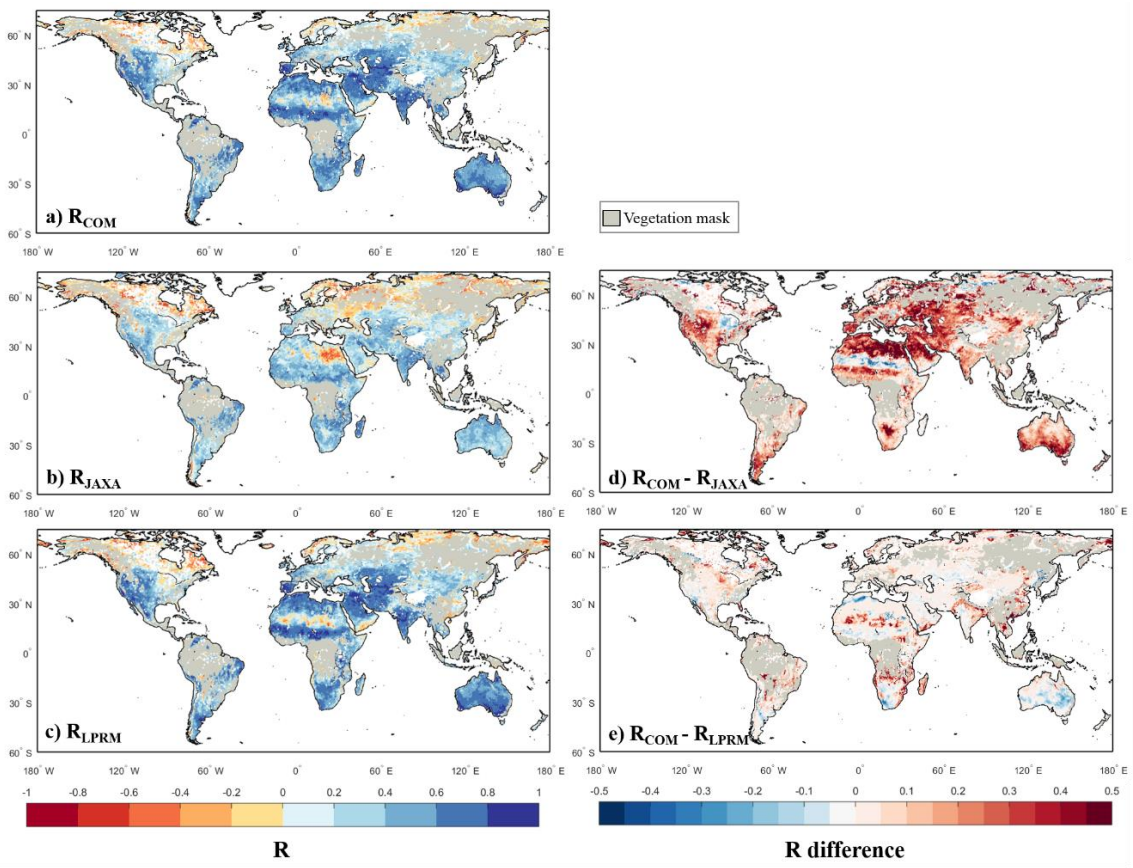


Figure A4.11. Cross-validation results of combination using data at the descending satellite path with ERA-Interim soil water contents level 1 as the reference: Spatial distribution of Pearson's correlation coefficients between MERRA-Land top soil layer soil moisture content and a) the combined product (R_{COM}), b) the JAXA product (R_{JAXA}) and c) the LPRM product (R_{LPRM}). Where, the global mean of R_{COM} is 0.47, R_{JAXA}, 0.31 and R_{LPRM}, 0.44 respectively. Panel d) shows the differences in between correlation coefficients of the combined and JAXA products (R_{COM} minus R_{JAXA}), and e), the combined and LPRM products (R_{COM} minus R_{LPRM}).

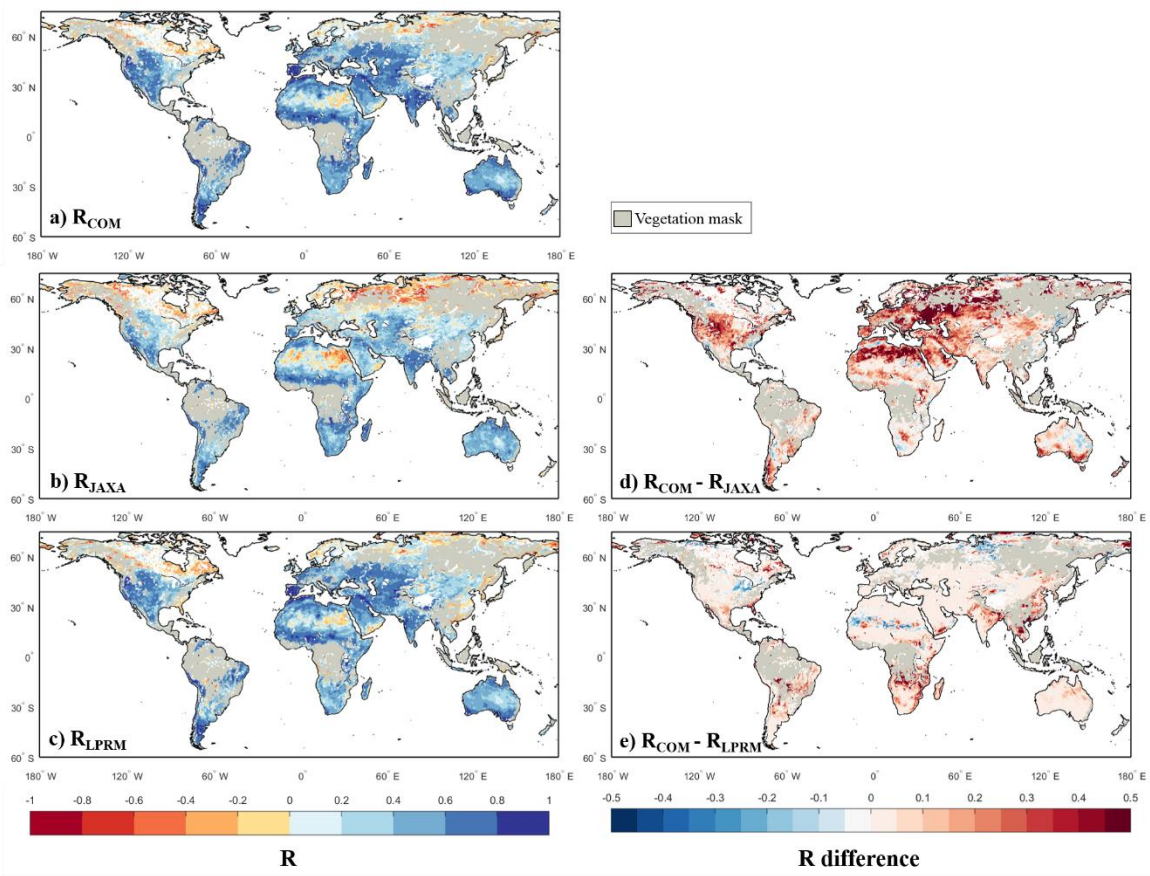


Figure A4.12. Cross-validation results of combination using data at the descending satellite path with MERRA-Land top soil layer soil moisture content as the reference: Spatial distribution of Pearson's correlation coefficients between ERA-Interim soil water contents level 1 and a) the combined product (R_{COM}), b) the JAXA product (R_{JAXA}) and c) the LPRM product (R_{LPRM}). Where, the global mean of R_{COM} is 0.49, R_{JAXA}, 0.35 and R_{LPRM}, 0.45 respectively. Panel d) shows the differences in between correlation coefficients of the combined and JAXA products (R_{COM} minus R_{JAXA}), and e), the combined and LPRM products (R_{COM} minus R_{LPRM}).

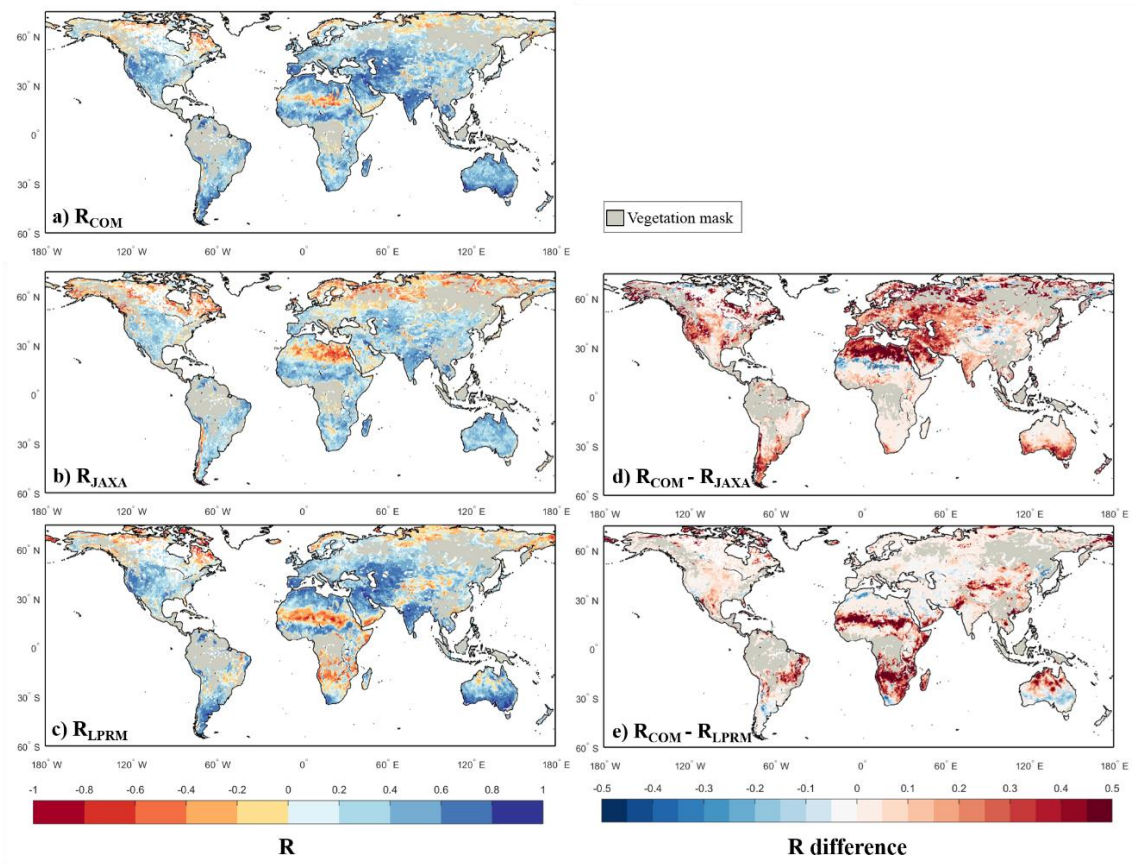


Figure A4.13. Cross-validation results of combination using data at the ascending satellite path with ERA-Interim soil water contents level 1 as the reference: Spatial distribution of Pearson's correlation coefficients between MERRA-Land top soil layer soil moisture content and a) the combined product (R_{COM}), b) the JAXA product (R_{JAXA}) and c) the LPRM product (R_{LPRM}). Where, the global mean of R_{COM} is 0.39, R_{JAXA}, 0.26 and R_{LPRM}, 0.30 respectively. Panel d) shows the differences in between correlation coefficients of the combined and JAXA products (R_{COM} minus R_{JAXA}), and e), the combined and LPRM products (R_{COM} minus R_{LPRM}).

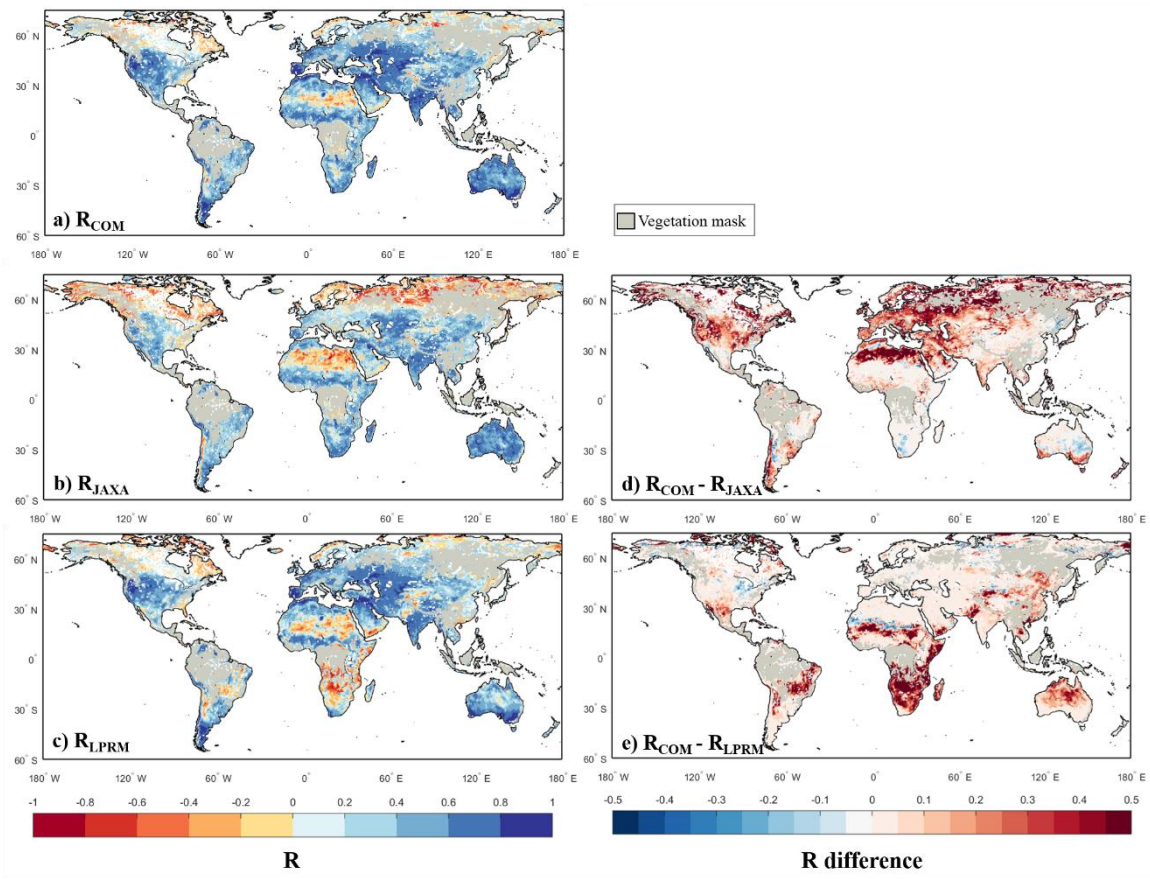


Figure A4.14. Cross-validation results of combination using data at the ascending satellite path with MERRA-Land top soil layer soil moisture content as the reference: Spatial distribution of Pearson's correlation coefficients between ERA-Interim soil water contents level 1 and a) the combined product (R_{COM}), b) the JAXA product (R_{JAXA}) and c) the LPRM product (R_{LPRM}). Where, the global mean of R_{COM} is 0.46, R_{JAXA}, 0.33 and R_{LPRM}, 0.36 respectively. Panel d) shows the differences in between correlation coefficients of the combined and JAXA products (R_{COM} minus R_{JAXA}), and e), the combined and LPRM products (R_{COM} minus R_{LPRM}).

Chapter 5 Combining Multiple Soil Moisture Retrievals Based on Maximizing Temporal Correlation: Dynamic Approach

This chapter improved the static linear combination method of Chapter 4 to allow for a non-static or nonstationary model combination as the basis for improving remotely-sensed surface soil moisture. Previous research had noted that two soil moisture products retrieved using the JAXA and LPRM algorithms from the same AMSR2 sensor are spatially complementary in terms of R against a suitable reference over a fixed period. Accordingly, a linear combination was proposed to maximize R using a set of spatially-varying, but temporally-fixed weights. Even though this approach showed promising results, there was room for further improvements, using non-static or dynamic weights that take account of the time-varying nature of the combination algorithm being approximated. The dynamic weighting was achieved by using a moving window. A number of different window sizes was investigated. The optimal weighting factors were determined for the data lying within the moving window and then used to dynamically combine the two parent products. Improved performance was shown for the dynamically-combined product over the static linear combination. Generally, shorter time windows outperformed the static approach, and a 60-day time window was suggested to be the optimum. Results were validated against in situ measurements collected from 124 stations over different continents. The mean R of the dynamically-combined products was found to be 0.57 and 0.62 for the cases using the European Centre for ERA-Interim and MERRA-Land reanalysis products as the reference, respectively, outperforming the statically-combined products (0.55 and 0.54).

This chapter is an edited version of: Kim, *S*.; Kim, *S*.; Parinussa, R.; Liu, Y.; Johnson, F.; Sharma, A., Merging Alternate Remotely-Sensed Soil Moisture Retrievals Using a Non-Static Model Combination Approach. *Remote Sensing* 2016**,** *8* (6), 518.

5.1 Introduction

This chapter is an extension of the previous combination scheme described in Chapter 4 , and the main objective is to provide further improvements of the performance of the combination approach that was presented. As discussed in the previous chapters, land surface features affecting the soil moisture retrieval (i.e. soil temperature and vegetation) vary through time and accordingly can cause temporally different performances. It is focused here on dynamic weights that take account of the time-varying performances of the different soil moisture products. The previous static combination uses the entire timeseries of data to calculate a single, constant weight for each point in space. The dynamic combination aims to use only a part of the timeseries, which has more useful information for representing temporal variability at a point in time. This has been shown to be successful in other applications; for example, Terui and van Dijk (2002) applied a time-varying weight for combining timeseries, which is the sum of the weight at the previous time step and an error following a Gaussian distribution. The work in Chowdhury and Sharma (2009) used time-varying error-based weights for combining river flows in arid areas. In addition, Khan et al. (2014) combined five global sea surface temperature forecasts by applying seasonally-based weights. The contribution of this study is to use time-varying weights to maximize the correlation of soil moisture products against a chosen reference. A second aim of the study is to determine the optimal size of the time window to calculate the combination weights using data at a daily time step.

Datasets used in this study, as well as data preprocessing are described in 5.2. The methodology for the dynamic combination is presented in 5.3. In 5.4, results from the global experiments are presented, as well as further results from a simulation experiment. Finally, the combined products are compared against *in situ* observations. In 5.5 and 5.6, the results are summarized and discussed, and suggestions are put forward for future research.

5.2 Data and Processing

5.2.1 Data

The JAXA (ver. 1.0) and LPRM soil moisture products retrieved from the X-band 10.7 GHz brightness temperature, available from both algorithms. It is not limited to use the same frequency for both products when combining them. As it is known that night-time has more favorable conditions for soil moisture retrieval (De Jeu et al., 2008), only soil moisture products from descending overpasses (01:30 a.m. equatorial overpass time) were used. Finally, the study was done over a two-year study period (1 January 2013–31 December 2014). The additional data prior to and after the analyzed period are used in the time windows for the dynamic combinations. All analyses were performed at the standard spatial resolution of 0.25° (approximately 25 km).

A reference soil moisture dataset is required in the combination approach described in Chapter 4 , as well as in the next section of this paper and is assumed to represent ground truth and the target for combining the products. In this study, it was also used as the reference to determine the optimal window size globally, as well as to check the performance of the dynamic combination approach the two different reanalysis products *i.e.* the volumetric water content of the topmost layer (0–0.07 m) from ERA-Interim and volumetric soil moisture content in the top layer (0–0.02 m) of MERRA-Land. Daily values at each grid cell were selected from the series of the reanalysis products. The selected value is the one that is temporally closest to the AMSR2 scan time over the grid on the day.

To independently evaluate the improvements in the combined product compared to the individual parent product, *in situ* measurements from the ISMN were used. The AMSR2 scan time was also applied to select daily observations that are closest to the AMSR2 observations. To ensure that the *in situ* data are of high quality for validation, the same checks and filters with Chapter 4 were applied to the datasets used in this study of which main characteristics are summarized along the with data sources in Table 5.1.

Table 5.1. Details of datasets used in this study.

Data Source	Dataset	Temporal Resolution	Spatial Resolution	Units
AMSR2-JAXA	Level 3 geophysical parameter SMC	Daily	0.25°	m^3/m^3
AMSR2-LPRM	Level 3 Surface Soil Moisture X-band	Daily	0.25°	m^3/m^3
AMSR2-LPRM	Vegetation optical depth C-band	Daily	0.25°	-
AMSR2	Scan time	Daily	0.25°	Seconds
ERA-Interim	Soil water contents Level (1 0–0.07 m depth)	6 hours	0.25°	m^3/m^3
ERA-Interim	Soil temperature Level (1 0–0.07 m depth)	6 hours	0.25°	K
MERRA-Land	Top soil layer soil moisture consent (SFMC)	Hourly	0.25° Resampled	m^3/m^3
ISMN	*In situ* measured soil moisture from 124 stations in 10 networks	Hourly	Point	m^3/m^3
S_a CCI	Topographic complexity, wetland fraction	-	0.25°	%

5.2.2 Data Preprocessing

The data preprocessing steps in 4.2.2 were applied to all the remote sensing datasets, reanalysis and *in situ* observations. Only *in situ* stations that have more than 100 paired observations with the RS products were included to ensure statistical robustness. Thus, 124

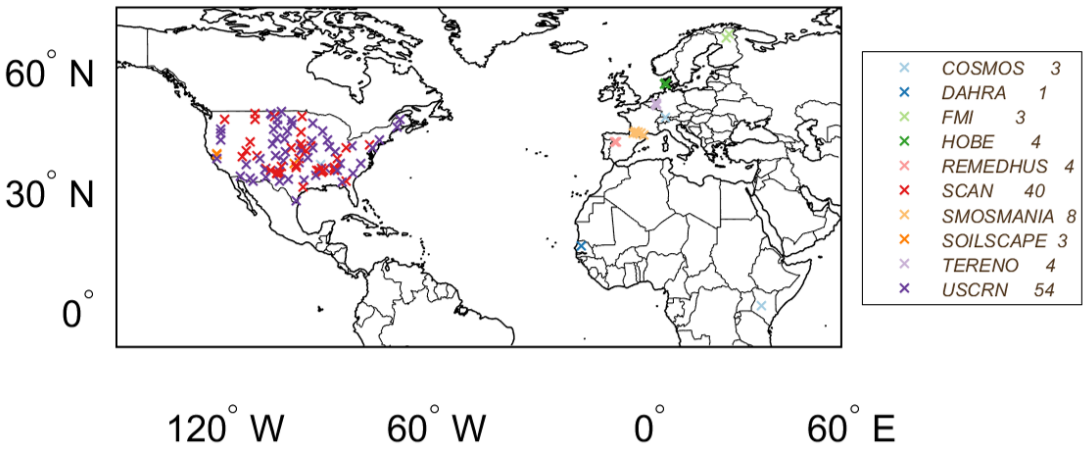


Figure 5.1. Locations of 124 ground stations from 10 networks used for the comparison with combined products.

stations from 10 networks were used for evaluation purposes, and their spatial distribution is presented in Figure 5.1. Even though such strict filtering processes were applied, it should be noted that there are likely to still be systematic differences among the datasets (Draper et al., 2009).

5.3 Methodology

The static approach in Chapter 4 that uses the entire datasets to calculate the weights will be further developed by using a dynamic segment of the datasets instead. This is based on the premise that different data segments are affected in different respects resulting from the system state being measured. Hence, one can assume that if this state-dependent behavior can be identified, combination weights may favor one retrieval algorithm over another for alternate states. This provides the basis for the dynamic combination approach presented here. The dynamic data segment is likely more dependent on and/or correlated with a spatiotemporal point to be combined and so is regarded to have more related information for calculating weights. For this, it is hypothesized that the information in a narrower range around the spatiotemporal point is more effective and sensitive for explaining temporal variability than the entire dataset. In this case, it is important to appropriately define the range of datasets by considering dimensional nearness to the point (Sharma & Mehrotra, 2014).

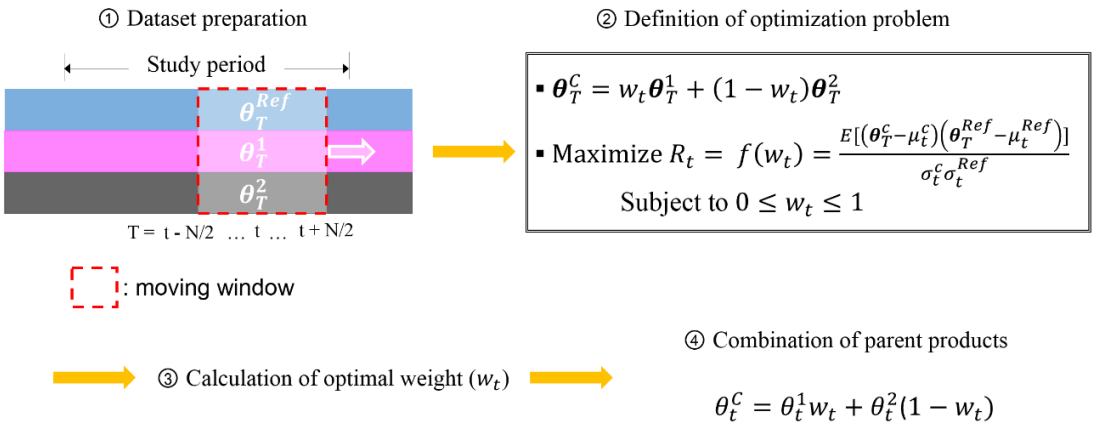


Figure 5.2. Schematic diagram for dynamic linear combination. T denotes the period defined by the window (*i.e.*, T = (t − N/2) :(t + N/2)). Therefore, a bold symbol that has T as its subscript means a vector in the period T, and a non-bold symbol with t as the subscript represents a value at the point in time t.

In this study, a temporally-moving window, which is centered on the point of interest at time *t* and has a size *N*-days, is applied to define the temporal nearness. As presented in Figure 5.2, the dynamic linear combination uses a part of the datasets in a period *T* by the moving window, to calculate the optimal weight for the point in time *t* (w_t), and the linear combinations are successively performed at points in time by using the calculated time-varying weights.

5.4 Results

5.4.1 Global Data Combination with Various Scenarios

For most locations, timeseries of the AMSR2 products are generally not continuous over the seasons due to freezing conditions and the revisit pattern of AMSR2. This could directly result in a shortage of observations within a moving window leading to non-significant values of R. For this reason, a minimum number of observations was determined based on a two-tailed *t*-test with a significance level (α) for R as

$$t_{\alpha/2} = R \cdot \sqrt{\frac{n-2}{1-R^2}} \qquad \text{Eq. 5.1}$$

In this study, α = 0.05, the corresponding $t_{\alpha/2} = t_{0.025}$ = 2.020 and a target R = 0.4 were used, and the minimum number of observations (*n*) is determined to be 25. As it approximately needs 50–75 days for acquiring the minimum observations (*i.e.*, 25) due to the revisit time of AMSR2 (*i.e.*, 2–3 days), 60 was set as the minimum window size (N) for the dynamic combination.

The initial experiment starts with using the volumetric water content of ERA-Interim as the assumed reference dataset. Both AMSR2 soil moisture products were then combined using the static combination and the dynamic combination with a range of window sizes (N = 60, 90 and 120 days). Then, the combined products were compared to the reference dataset. The combination results are presented in Figure 5.3a–f, which show global maps of R of the parent and combined products against the reference. Additionally, the box plot in Figure 5.3g shows the distributions of the global R for the various scenarios (see Figure A5.1 for results using MERRA-Land as the reference).

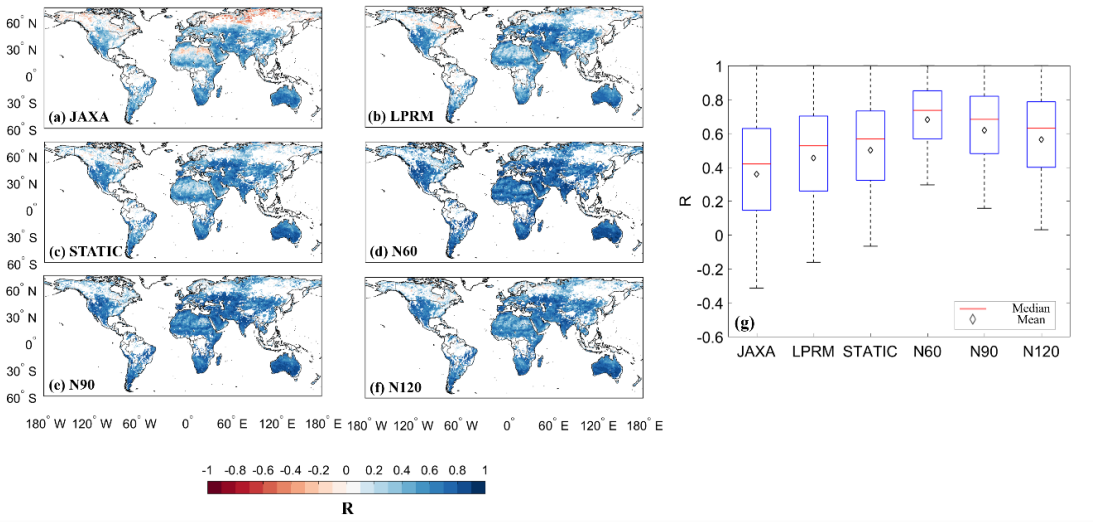


Figure 5.3. Results from experiments that uses ERA-Interim as the reference for various window sizes, N60, N90 and N120. Each panel shows the R between the reference and (a) JAXA, (b) LPRM, (c) static, (d) N60, (e) N90 and (f) N120; the more bluish colours in the maps indicate higher R against the reference; the overall performance for the various scenarios is summarized in the boxplot (g).

The spatial distributions of R tend to be improved (*i.e.*, more bluish colors) from the parent (Figure 5.3a, b) to statically- (Figure 5.3c) and dynamically-combined products (Figure 5.3d–f), and the tendency is clearly shown in the box plot of Figure 5.3g. The dynamically-combined products are consistently better than both parents and the statically-combined product. This result supports the hypothesis that data within a specific period provide better information for calculating weights maximizing R. It is also clear that the shorter window sizes in the dynamic combination outperform the longer window sizes. Hence, the N60 window size was selected for further analysis. In other words, the linear combination makes the parent products be close to the reference dataset in terms of temporal correlation, and the dynamic approach enforces the function with a shorter window size. To assess the sensitivity of this finding to the choice of reference dataset, this N60 scenario was again applied using the top soil moisture layer of MERRA-Land as the reference dataset. Figure 5.4 presents the differences of this N60 experiment using ERA-Interim (top panels) and MERRA-Land (bottom panels) as the reference, respectively.

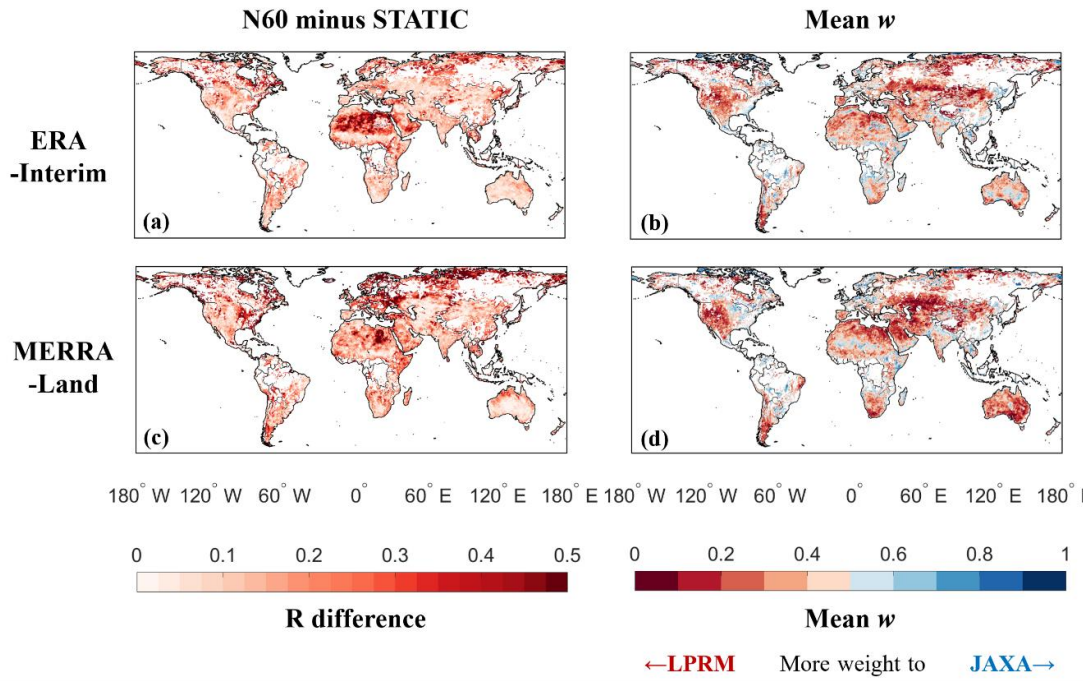


Figure 5.4. Comparison between combined soil moisture products using ERA-Interim (top panels) and MERRA-Land (bottom panels) as the reference, respectively. (a, c) The differences in R between the static and N60 products against each reference (*i.e.*, R of N60 minus R of static); and (b, d) the mean weights that were used for the dynamic combination using each reference over the two-year study period.

As shown in Figure 5.4a, c, the two dynamic products lead to improvements at different locations due to the different spatial patterns in agreement between each reference dataset and the parent products. Thus, the spatial distribution of the averaged weighting factors used for the dynamic combination is also different (Figure 5.4b, d). The differences are prominently contrasted over the desert regions, such as the Sahara, Middle East and central Australia. For comparison, results from the N90 and N120 are shown in Figure A5.2 and Figure A5.3, and global maps are also presented in Figure A5.4 showing standard deviations of optimal weights from the two references (*i.e.* ERA-Interim and MERRA-Land) and three window sizes (*i.e.* N60, N90 and N120). To better understand these differences and how the chosen reference interacts with the parent products, an additional experiment is now setup in a controlled environment using simulated datasets.

5.4.2 A Simulation Experiment

The experiments using actual datasets were performed in the previous section 5.4.1 and two findings for the static and dynamic approaches were explained; 1) The dynamically combined products are consistently better than both parents and the statically combined product; 2) The shorter window sizes in the dynamic combination outperform the longer window sizes.

To corroborate the findings in terms of the parent product quality and the window size N, the dynamic linear combination was assessed against the static approach in a simulation experiment that uses three periodical datasets generated by

$$\boldsymbol{\theta} = A \cdot \sin(2\pi \cdot F \cdot \boldsymbol{t}) + M \qquad \text{Eq. 5.2}$$

where $\boldsymbol{\theta}$ is simulated soil moisture (m^3/m^3), *A* is amplitude 0.2 (m^3/m^3), *F* is frequency 1/365, *t* is a point in time (daily) and *M* is mean soil moisture (0.4 m^3/m^3).

The parameters used in Eq. 5.2 were chosen for the simulated datasets to be within the plausible soil moisture range (0–0.6 m^3/m^3, (Dorigo et al., 2011)). From the three generated datasets, two were designated as the parent products, and the remaining one was chosen to be the reference. To introduce different correlation coefficients among the parent products and the reference, which eventually govern the optimal weights, randomly-generated white noise from −0.2 to +0.2 was added to each of the parent products. Next, an integer between 30 and 360 was also randomly selected for the size of moving window (N). After that, a series of dynamic combinations and a static combination were performed using data equivalent to a two-year period, and the correlations between the dynamically- ($\mathrm{R_{dyn}}$) and statically- ($\mathrm{R_{sta}}$) combined products were calculated and compared.

A correlation-based Euclidean distance was calculated for each simulation to represent the similarity/difference between the parent products and the reference. As there are two parent products for the combination, they are best summarized as a vector of correlations, (R_1, R_2) T, and then, the Euclidean distance (ξ) is calculated as

$$\xi = \mathrm{sqrt}((1 - R_1)^2 + (1 - R_2)^2) \qquad \text{Eq. 5.3}$$

Through this procedure, 1000 simulations were performed, and the results provide further information on the performance of the dynamic combination approach in terms of the window size (N) and also the influence on the results of the quality of the parent product(s).

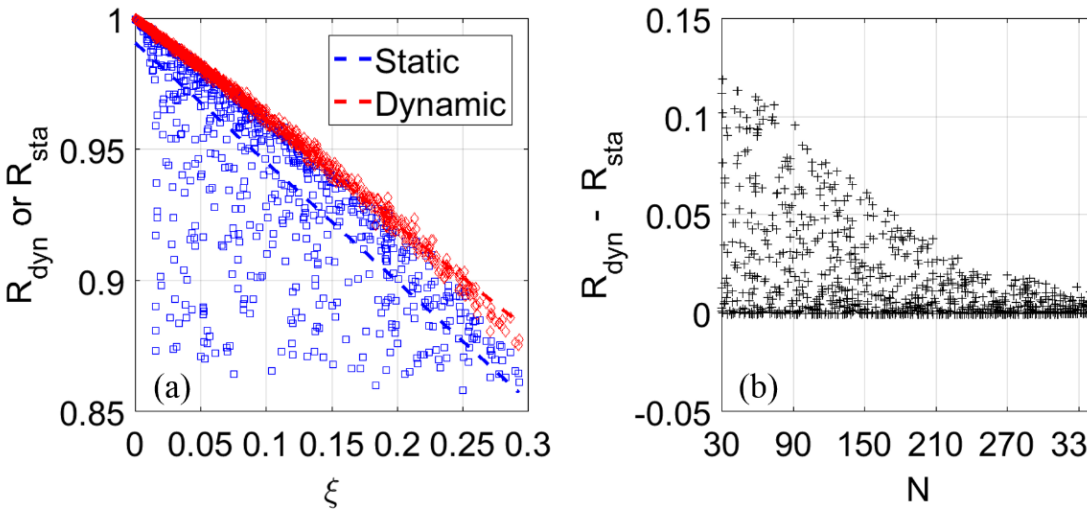


Figure 5.5. Results from the simulation experiment. (a) The x-axis indicates Euclidean distances (ξ) calculated by Eq. 5.3, representing the qualities of the parent products, and the y-axis, R_{dyn} or R_{sta}. The dashed two lines present the linear regression of all results from the dynamic and static combinations, respectively. (b) The x-axis indicates N sizes, the y-axis differences between R_{dyn} and R_{sta} (*i.e.*, $R_{dyn} - R_{sta}$).

The results from the simulation experiment show that: 1) the dynamic performance is consistently better than that of the static approach (Figure 5.5a); and 2) short window sizes provide the largest contrast between the static and dynamic approaches. Sufficiently, large window sizes will yield identical results as the static combination approach (Figure 5.5b). These results are in line with the results obtained through the global datasets, as was summarized in Figure 5.3. Figure 5.5a also shows that both performances tend to decrease with increasing Euclidean distance, in which the two linear regression lines imply strong dependence of both combination performances on initial quality of parent products.

The final performance of the combination depends on the relative differences between the selected reference and the parent products. Therefore, an important consideration is that the quality of the combined product heavily relies on the quality of the reference that is assumed to

represent the truth. Therefore, it is necessary to further investigate how the reference difference affects the combination performances from the static and dynamic approaches. To better understand the associated qualities of the different products, they were compared against *in situ* measurements from the ISMN.

5.4.3 Comparison Against *In Situ* Observations

In this section, the dynamically-combined products with various scenarios are compared against *in situ* measurements from the ISMN, demonstrating how much the results rely on the quality of the reference dataset. The R between the *in situ* measurements and the adopted reference dataset demonstrates the impact of the R on the combination approach. For this comparison, the (N60) dynamic weighting approach was compared to the static weighting approach, as well as the parent products (Figure 5.6).

Data points in Figure 5.6 that are below the 1:1 line represent locations where the dynamic approach performed better than the static one. The overall relation between these approaches for all sites in the ISMN is represented by the linear regression lines. Those regression lines show that the performance of the dynamic combination is generally better than the static approach, particularly when the initial R between the ISMN dataset and the reference data is high R > 0.4. An interesting feature obvious in Figure 5.6 is the qualitative differences in static and dynamic combination performances, where agreement between the reference and *in situ* measurements was expressed with colors from red to blue (0–1). The superiority in the dynamic approach is obvious when there is a better agreement between the reference and *in situ* measurements (bluish marks under the 1:1 line). However, degraded performance for the dynamic approach is observed when the reference and *in situ* datasets do not agree (reddish marks above the 1:1 line). This suggests that the reference quality is important to consider when applying the dynamic approach. The dynamically-combined product tends to be closer to the reference; therefore, if the reference data are different from the *in situ* data, then the dynamic product will tend to deviate more from the *in situ* data, as well. In the case of ERA-Interim, the mean R of the static combinations is 0.55, which is similar to the LPRM product (0.55), but the dynamically-

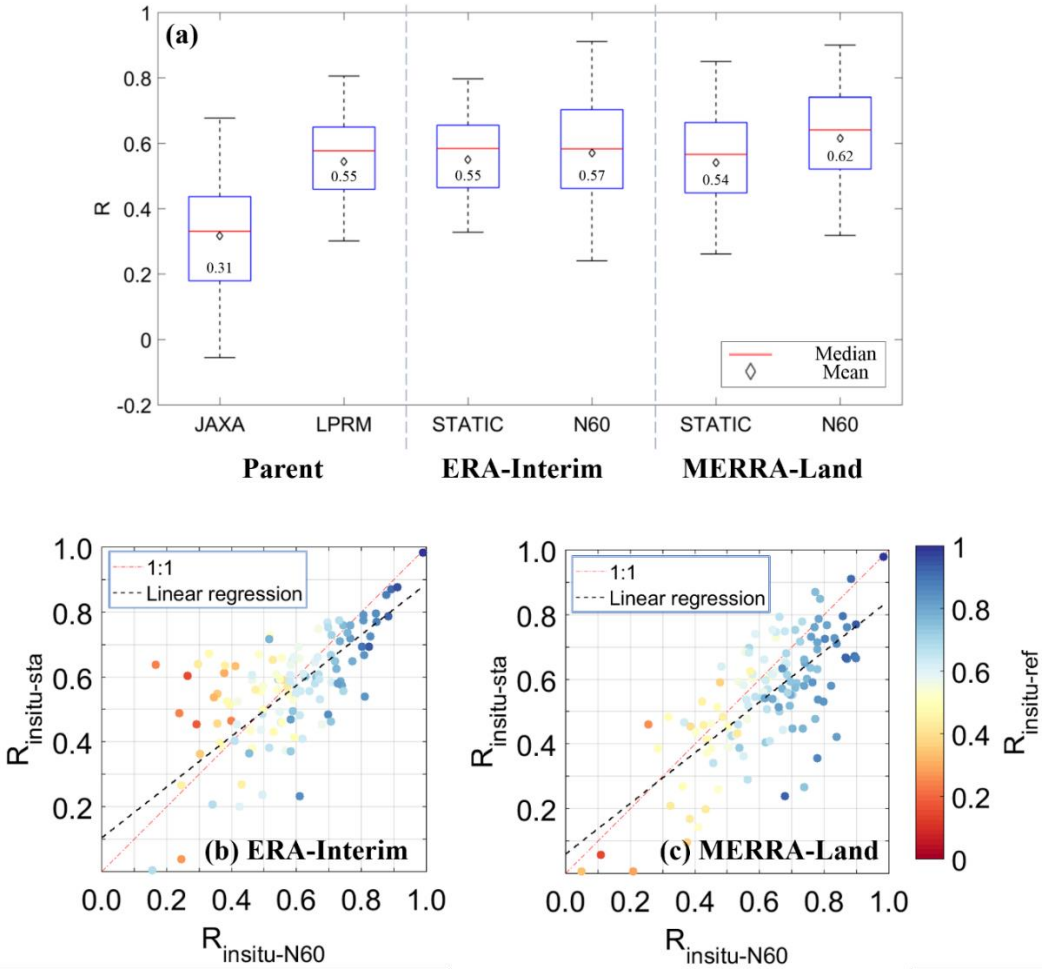
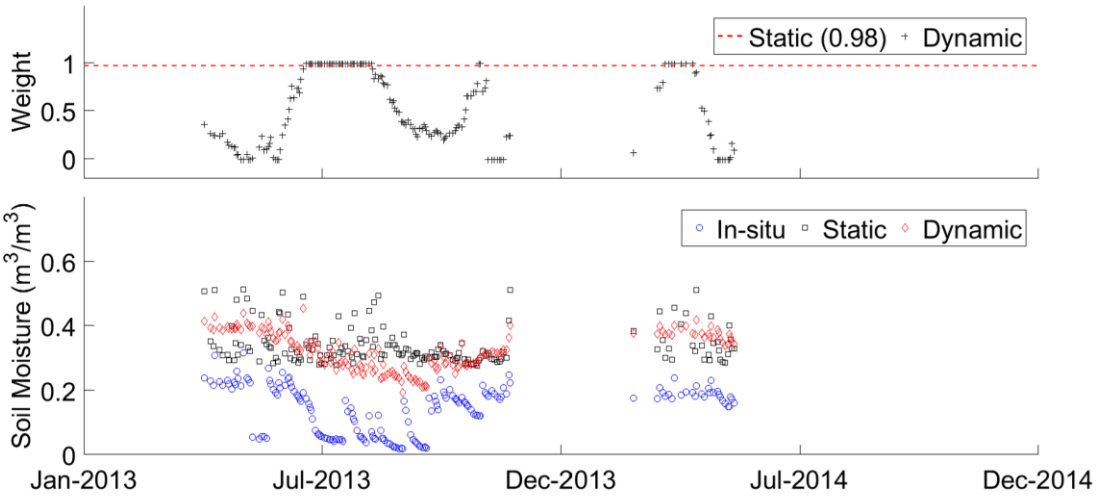
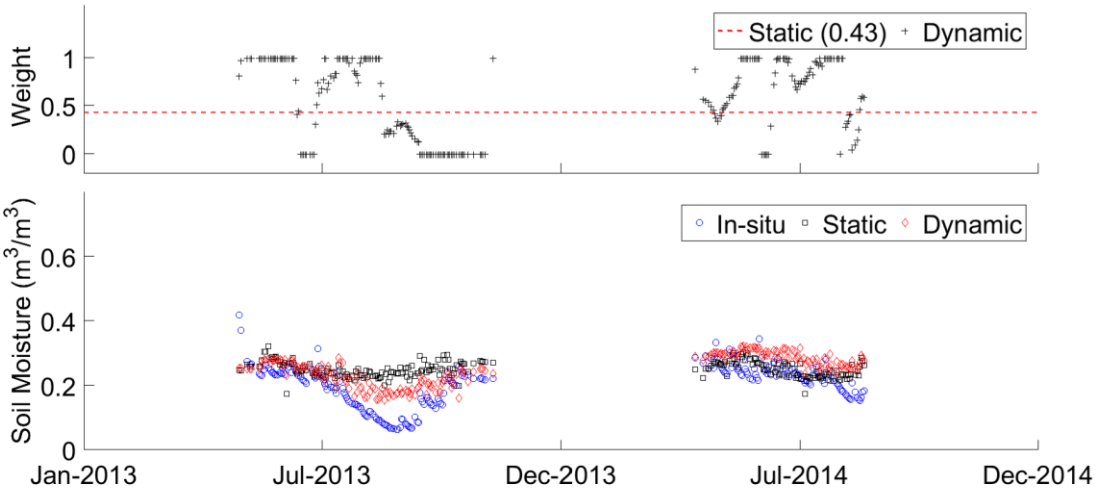


Figure 5.6. (a) Box plots showing combination performances against *in-situ* measurements with the N60 and the two references. The labels on the x-axis indicate parent or statically-/dynamically-combined products with the references, and the y-axis R between the product and the *in-situ* measurements. The value in each box is the mean of R. Comparison against *in-situ* measurements from the ISMN for dynamically combined products using the N60 and (b) ERA-Interim and (c) MERRA-Land as the reference, respectively. The x-axis presents R between a dynamic product and the *in-situ* measurements from a station, the y-axis R between a static product and the *in situ* measurements.

combined product (0.57) outperforms both parent products. For the case of MERRA-Land, the dynamic approach provides better performance (0.62) than the static approach (0.54) and all other products including the parent. From the results, it would be possible to select the static or dynamic product based on a threshold for the reference. In other words, it could be better to choose the static product if the reference quality is lower than the threshold or uncertain, which is further discussed in the next section.



(a) SCAN_SandyRidge



(b) USCRN_Sandstone-6-W

Figure 5.7. Dynamic and static combination results using MERRA-Land as the reference at (a) Sandy Ridge station in Soil Climate Analysis Network and (b) Sandstone-6-W station in U.*S*. Climate Reference Network. Each panel shows static/dynamic weights (top), as well as timeseries of statically- and dynamically-combined soil moisture products (bottom).

As an example of the static and dynamic combination behavior, Figure 5.7 shows combination results using MERRA-Land as the reference at two ground stations where the dynamic approach outperforms most the static approach among the 124 stations.

In the case of Sandy Ridge station (Figure 5.7a), the static weight is 0.98, and R between the statically-combined product and *in situ* measurements is 0.27. However, it sharply increases to 0.74 when applying the time-varying weights. For the case of Sandstone-6-W station (Figure

5.7a), the static weight is 0.43, and R between the statically-/dynamically-combined product and *in situ* measurements is 0.36 and 0.78, respectively.

5.4.4 Influence of the Quality of the Parent Products and Reference

The correlation-based Euclidean distance, used to assess the quality of the simulated datasets in the section 5.4.2, was extended to assess the quality of the parent products and reference against the *in situ* measurements.

Eq. 5.3 is expanded to include three products (*i.e.*, two parent products and the reference dataset), and the vector of correlations at each station against *in situ* measurements is defined as $(R_{insitu\text{-}JAXA}, R_{insitu\text{-}LPRM}, R_{insitu\text{-}ref})^T$. Therefore, the extended Euclidean distance (ξ) is determined by:

$$\xi = \mathrm{sqrt}(\left(1 - \mathrm{R}_{\mathrm{insitu-JAXA}}\right)^2 + (1 - \mathrm{R}_{\mathrm{insitu-LPRM}})^2 + (1 - \mathrm{R}_{\mathrm{insitu-ref}})^2) \quad \text{Eq. 5.4}$$

At every station, R was determined through the static and dynamic approaches against corresponding *in situ* measurements ($\mathrm{R}_{\mathrm{insitu-sta}}$ and $\mathrm{R}_{\mathrm{insitu-dyn}}$) and was plotted against the Euclidean distances (Figure 5.8).

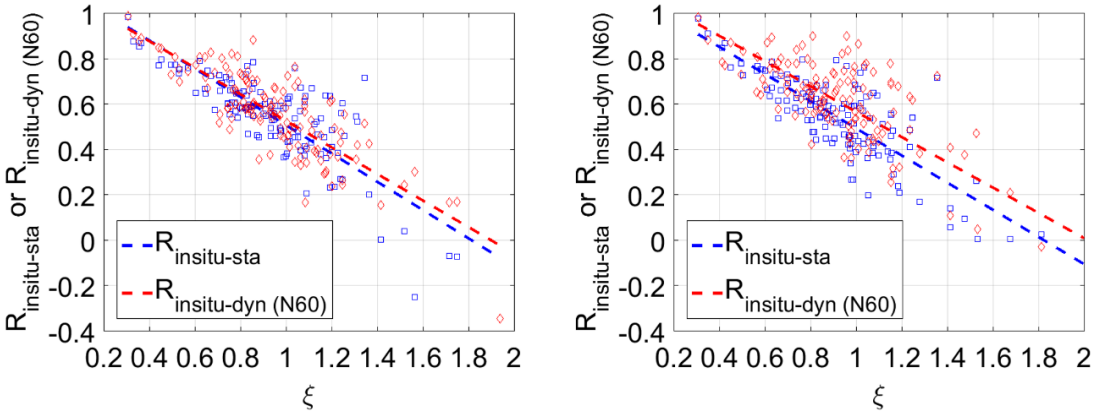


Figure 5.8. Combination performances with the quality of parent products and reference against *in situ* measurements. The *x*-axis for each panel presents the Euclidean distances (ξ) calculated by Equation (8), and the *y*-axis $\mathrm{R}_{\mathrm{insitu-sta}}$ or $\mathrm{R}_{\mathrm{insitu-dyn}}$. Linear regression lines represent the tendencies of both cases.

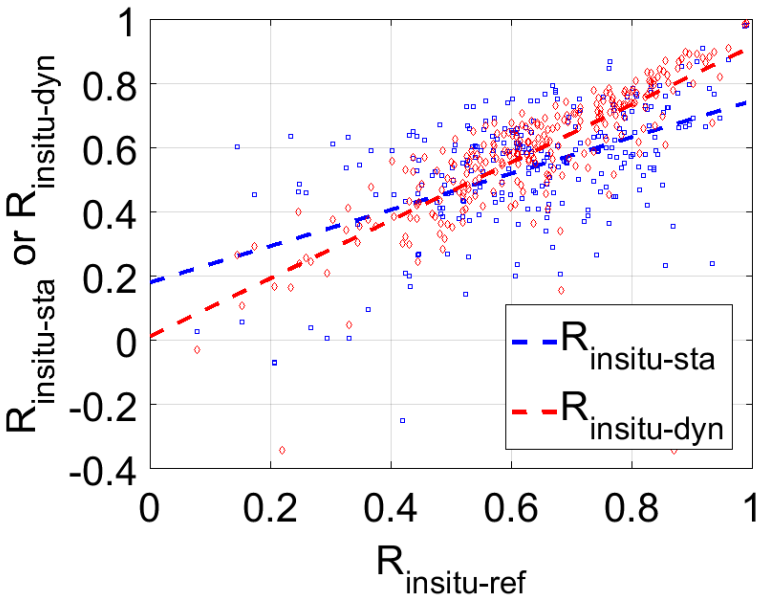


Figure 5.9. Combination performances with reference quality against *in situ* measurements. The *x*-axis presents R between *in situ* measurements and the references ($R_{insitu-ref}$), the *y*-axis R between *in situ* measurements and statically-/dynamically-combined products ($R_{insitu-sta}$ and $R_{insitu-dyn}$). Linear regression lines are added for representing the average tendencies of both cases.

The dynamic approach shows generally better performances than the static approach, and this tendency is more conspicuous for MERRA-Land (Figure 5.8b) than for ERA-Interim (Figure 5.8a). It is also shown that $R_{insitu-sta}$ and $R_{insitu-dyn}$ tend to decrease with increases of ξ, which means that a good agreement between the parent products and reference is an important precondition for both combination methods to provide improvements over the parent. As before, it is necessary to consider the quality of available reference datasets when deciding which one to use for these combination approaches.

Figure 5.9 shows all $R_{insitu-sta}$ and $R_{insitu-dyn}$ from the two combination scenarios plotted against $R_{insitu-ref}$. The performance of the dynamic combination is clearly better than those of the static combination with higher $R_{insitu-ref}$. When $R_{insitu-ref}$ is around 0.6 or greater, then the dynamic combination is the superior approach. This suggests that a threshold value for $R_{insitu-ref}$ can be used to select when to use the dynamic or static combination approach. Namely, the static product is selected when $R_{insitu-ref}$ at a station is equal to or less than the threshold and vice versa, even though such approach can be only applied in areas

where *in situ* measurements are available. That is to say, it can be optionally determined whether or not to apply the dynamic combination approach when a validation for the reference is supported. For example, Albergel et al. (2012a) evaluated ERA-Interim soil moisture using *in situ* measurements from stations, and based on the results, one can have a basis for applying the static or dynamic combination approach over certain regions. In addition, the suggestion can be extended through large verification techniques, such as triple collocation (Gruber et al., 2016).

5.5 Discussion

This study presented an approach to dynamically combine RS soil moisture products through a linear combination approach that maximizes R. The rationale of the proposed method is to use information from two parent products that are spatially and temporally complementary to each other. These parent products were linearly combined by applying a set of weights, which are governed by the correlation among the parent products and a reference.

For this, it was hypothesized that datasets within a temporal moving window, centered around a point in time and having size N, provide better information to calculate the optimal weight and lead to a better combination. Accordingly, it applied to the experiments using the two AMSR2 soil moisture products and the simulated datasets. The experiments led to three main conclusions about the dynamic combination methodology. 1) The dynamically-combined product is consistently better than the statically-combined and parent products when the reference quality is fairly good ($R > 0.6$); 2) better performances came out with shorter window sizes for the dynamic combination, and the N60 was selected as the optimum for combining these two AMSR2 products; 3) the performances of the dynamic and static approaches tend to decrease with the decreases of parent product quality against the reference.

In reality, the quality of a chosen reference is spatiotemporally variable (Dee et al., 2011; Reichle et al., 2011), and this is likely to be the most important issue to be considered for combining parent products. To investigate this, experiments with two combination scenarios, the N60 with two references (*i.e.*, ERA-Interim and MERRA-Land), were performed and compared against the assumed truth, *i.e.*, *in situ* measurements from the ISMN. The results

showed that the performance indeed relies on reference quality and a good quality of reference is essential for a good performance from the dynamic approach. It was found that R between a reference and *in situ* data should be at least moderately positive (0.6). When the correlations are lower, the static combination is likely to be a more reliable choice based on the results from Figure 5.9.

5.6 Conclusions

Based on the results, there are a few possibilities for future extensions. First, the spatiotemporally-varying weights could provide information to improve the parent products, as well as the retrieval algorithms. The results here provide the JAXA and LPRM algorithms with clues on promising areas for further improvement by highlighting times of year and areas of the globe where the other product has superior performance. In addition, the results about the quality of the reference dataset highlight improvements that could be made in these products (Albergel et al., 2012a; Balsamo et al., 2015).

The combination scheme can be used with more than two parent products, so that it can reflect varying strengths resulting from the different techniques or retrieval algorithms. In this case, products from the same sensor but different frequencies or algorithms are preferred. This is because the products will then have the same swath pattern, leading to a good number of paired observations and the same scan time, so that the weather conditions are constant. A possible approach to combine multiple satellite-derived products from various sensors, which have different swath patterns and scanning times, is applying the methodology to datasets at coarser temporal resolutions (*e.g.*, weekly) by averaging the original datasets at a finer temporal resolution. Lastly, it should be emphasized that the presented combination scheme is applicable to any spatiotemporal dataset where a reference dataset is available. From the results in this study, the following general guidelines are suggested for other applications of the dynamic linear combination. A minimum window size is recommended based on statistical significance to calculate the optimal weights. Secondly, the quality of a reference should be supported through a validation procedure to decide whether to accept or not the dynamically-combined product.

5.7 Appendix

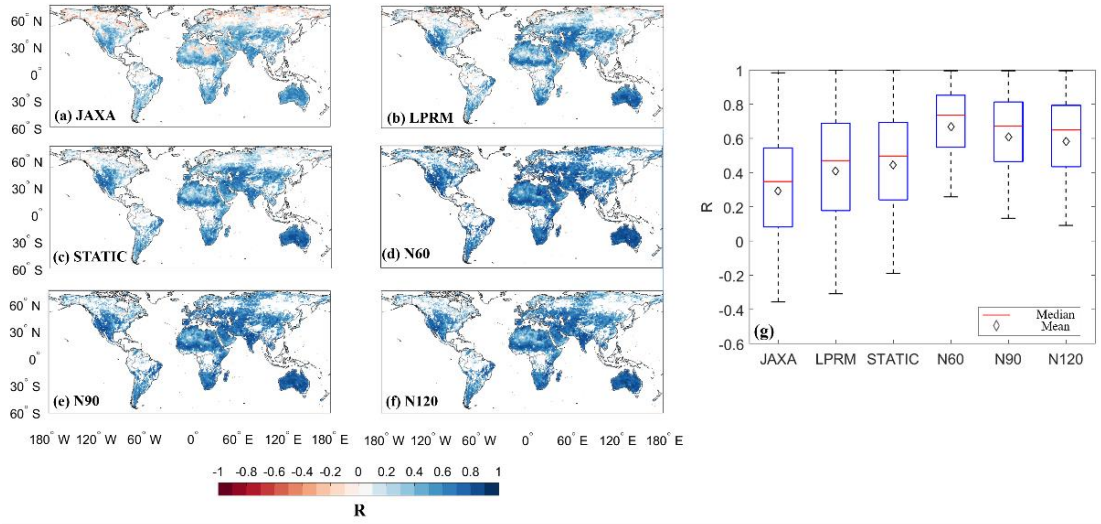


Figure A5.1. Results from experiments that uses MERRA-Land as the reference for various window sizes (N60, N90 and N120). Each panel shows the R between the reference and (a) JAXA, (b) LPRM, (c) static, (d) N60, (e) N90 and (f) N120. The more bluish colours in the maps indicate higher R against the reference, the overall performance for the various scenarios is summarized in a boxplot (g).

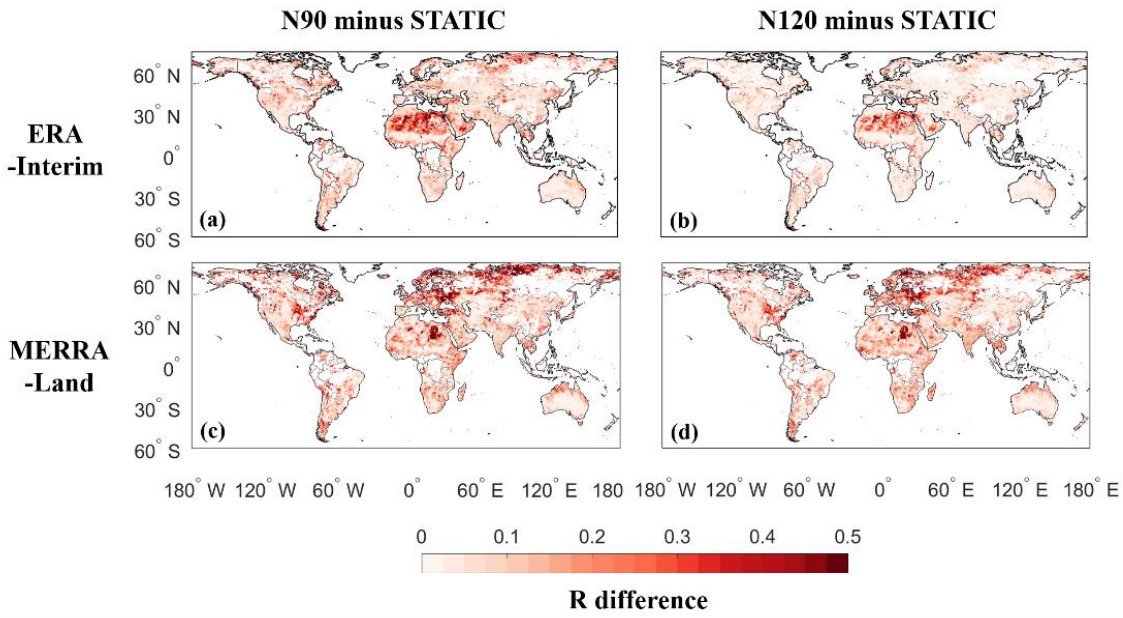


Figure A5.2. Differences in R between the static and dynamic products (N90 and N120). For ERA-Interim as the reference, (a) R of N90 minus R of static and (b) R of N120 minus R of static. (c) and (d) show corresponding results with (a) and (b) when using MERRA-Land as the reference. With relation to Figure 4a in the main manuscript, it is shown that the differences are more contrasted with shorter N sizes.

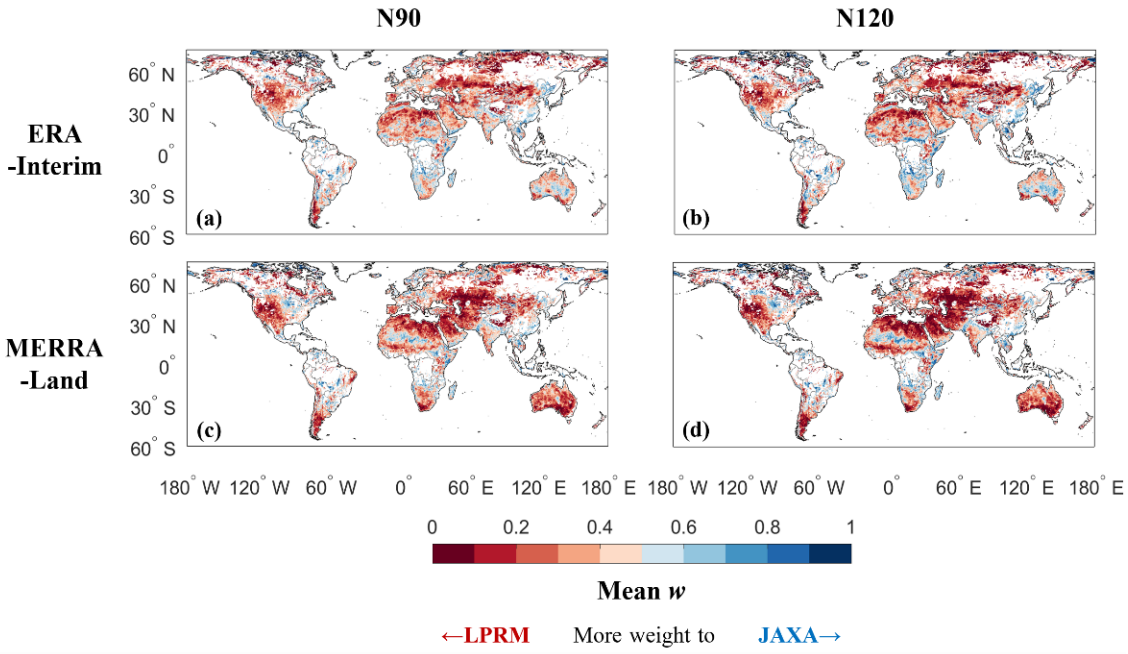


Figure A5.3. Mean weights used for dynamically combined soil moisture products. For ERA-Interim as the reference, (a) presents mean weights from N90 over the 2-year study period, and (b) from N120. (c) and (d) show corresponding results with (a) and (b) when using MERRA-Land as the reference.

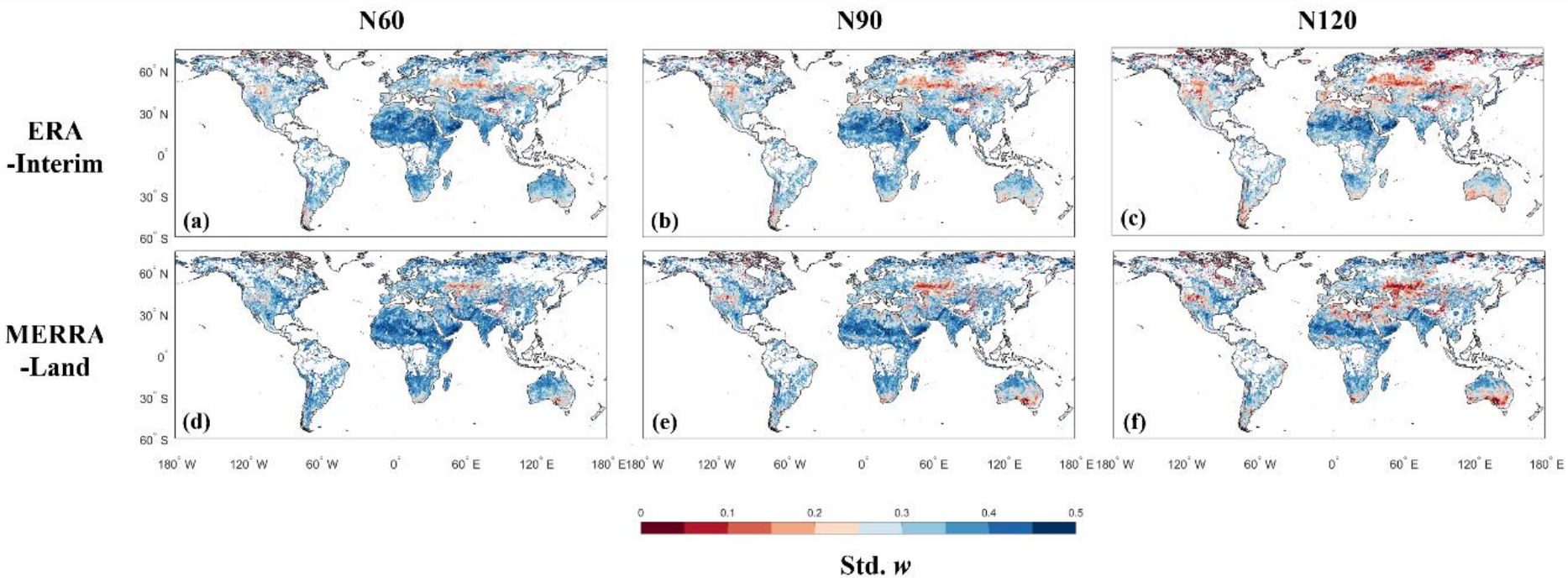


Figure A5.4. Standard deviations of optimal weights used for dynamically combined soil moisture products. For ERA-Interim as the reference, (a) presents standard deviations from N60 over the 2-year study period, (b) from N90, and (c) and N120. (c), (e) and (f) show corresponding results with (a), (b) and (c) when using MERRA-Land as the reference.

Chapter 6 Spatial Disaggregation of Coarse Soil Moisture Data by Using High Resolution Remotely Sensed Vegetation Products

A novel approach is presented to spatially disaggregate coarse soil moisture (SM) by only using remotely sensed vegetation index. The approach is based on the conditional relationship of vegetation with time-aggregated SM, allowing the coarse scale SM (0.25°) to be disaggregated to the spatial resolution of the vegetation product (0.05°). The method was applied to satellite-derived SM over Jan. 2010–Dec. 2011, using the high-resolution normalized difference vegetation index (NDVI). The results were evaluated against ground measurements during the 2-year period over the contiguous United States and Spain, and also compared with an existing independent disaggregation method that also requires land surface temperature observations. It is shown that the proposed approach can provide fine resolution SM with reasonable spatial variability.

This chapter is an edited version of: Kim, *S.*; Balakrishnan, B.; Liu, Y.Y.; Johnson, F.M.; Sharma, A. (2017). Spatial Disaggregation of Coarse Soil Moisture Data by Using High Resolution Remotely Sensed Vegetation Products. *IEEE Geoscience and Remote Sensing Letters*

6.1. Introduction

The direct use of satellite soil moisture (SM) products has been limited by their coarse spatial resolution (>100 km^2) and uncertainties resulting from the complex SM retrieval procedures (Dorigo et al., 2010; Kim et al., 2015a). Current research is focusing on both these problems; to reduce the uncertainties (Kim et al., 2015b, 2016b; Kornelsen & Coulibaly, 2015; Su et al., 2013a) and to downscale the satellite microwave measurements to a finer spatial resolution appropriate for regional SM assessments (Kim & Barros, 2002; Merlin et al., 2012; Pellenq et al., 2003; Piles et al., 2011; Shin & Mohanty, 2013).

Compared to microwave sensors, optical/thermal sensors can provide observations at a relatively high spatial resolution observation. The tradeoff is that the optical/thermal sensors are affected by atmosphere conditions such as clouds (Vermote et al., 2002). This suggests that data collected by optical or thermal sensors may be useful in spatially disaggregating the coarse microwave SM products. Disaggregation develops relationships between SM estimated from the microwave sensors and land surface temperature (LST) and/or appropriately selected vegetation indices (VI) which are available at finer spatial resolutions. For example, Fang et al. (2013) used look-up tables to relate the datasets at the coarse spatial scale to the daily temperature difference at a finer spatial scale to disaggregate the daily average SM. Peng et al. (2015) recently proposed a SM disaggregation method by using the vegetation temperature condition index (VTCI) as a soil moisture proxy. The method spatially distributes a coarse SM value (0.25°) to fine values (0.05°) which are linearly proportional to the VTCI at a 0.05° resolution.

However, these methods require the vegetation and LST data at both the coarse and fine scales to be available at the same time as the SM data. This presents a problem as there are many missing values in LST datasets from optical/thermal infrared sensors due to cloud masks and thus the disaggregated SM products tend to be discontinuous. For example, it is just around 45% annual mean percentage of the contiguous United States (CONUS) domain for which satellite land surface data are typically available (Crosson et al., 2012). In this chapter, the temporal interaction of SM with higher resolution vegetation proxies is considered as an alternative to the LST-derived information, and a simple SM disaggregation model is developed to provide a continuous time series of SM with a persistence structure closer to what is observed.

Two case studies were used to validate the proposed method. First, using the new method, coarse SM at a spatial resolution of 0.25° (approximately 25 km, hereafter referred to as ‘coarse’) is disaggregated to fine SM at 0.05° (approximately 5 km, hereafter referred to as ‘fine’) over the CONUS for a 2-year study period from January 2010 to December 2011. The results were evaluated against in situ SM measurements from ground stations distributed over the CONUS.

The method was also applied to the REMEDHUS network in Spain for a direct comparison with (Peng et al., 2015) by applying the same SM data, study period and area. Importantly, this analysis over the dense network is to evaluate the disaggregated SM in terms of its ability to reproduce the spatial variation at the fine spatial scale.

6.2. Data and Processing

6.2.1. Satellite Soil Moisture

The European Space Agency Climate Change Initiative (ESA CCI) SM has been generated using four passive and two active microwave space-borne instruments covering 36 years from 1978 to 2014 (Liu et al., 2012; Liu et al., 2011b). It consists of three products: active, passive and active-passive merged data which provide daily surface SM at a spatial resolution of 0.25°. Among them, the merged data (version 02.2) was used in this study which has been comprehensively validated with promising performances at the global scale (Dorigo et al., 2014), and also used in Peng et al. (2015).

6.2.2. Ground Soil Moisture Networks over Study Areas

The primary study area was selected as the CONUS (24.25°N-49.50°N, 66.75°W- 125°W). It covers a wide range of climate and land cover types which allows the general applicability of the proposed method to be assessed. The four dominant climate zones are cold in the northern areas, arid in the western regions, temperate in the central and middle Atlantic coast regions, and tropical in the southern Florida (Homer et al., 2015). There is also an extensive network of *in situ* SM stations for evaluating disaggregation results.

Spatially dense networks are necessary to validate spatial variability of disaggregated SM at a sub pixel scale. The secondary study area over the REMEDHUS network covers a flat spatial domain (41.00°N-41.75°N, 5.00°W-5.75°W, approximately 1300 km^2). The land cover mainly consists of croplands and shrub lands and the climate is classified as semi-arid continental

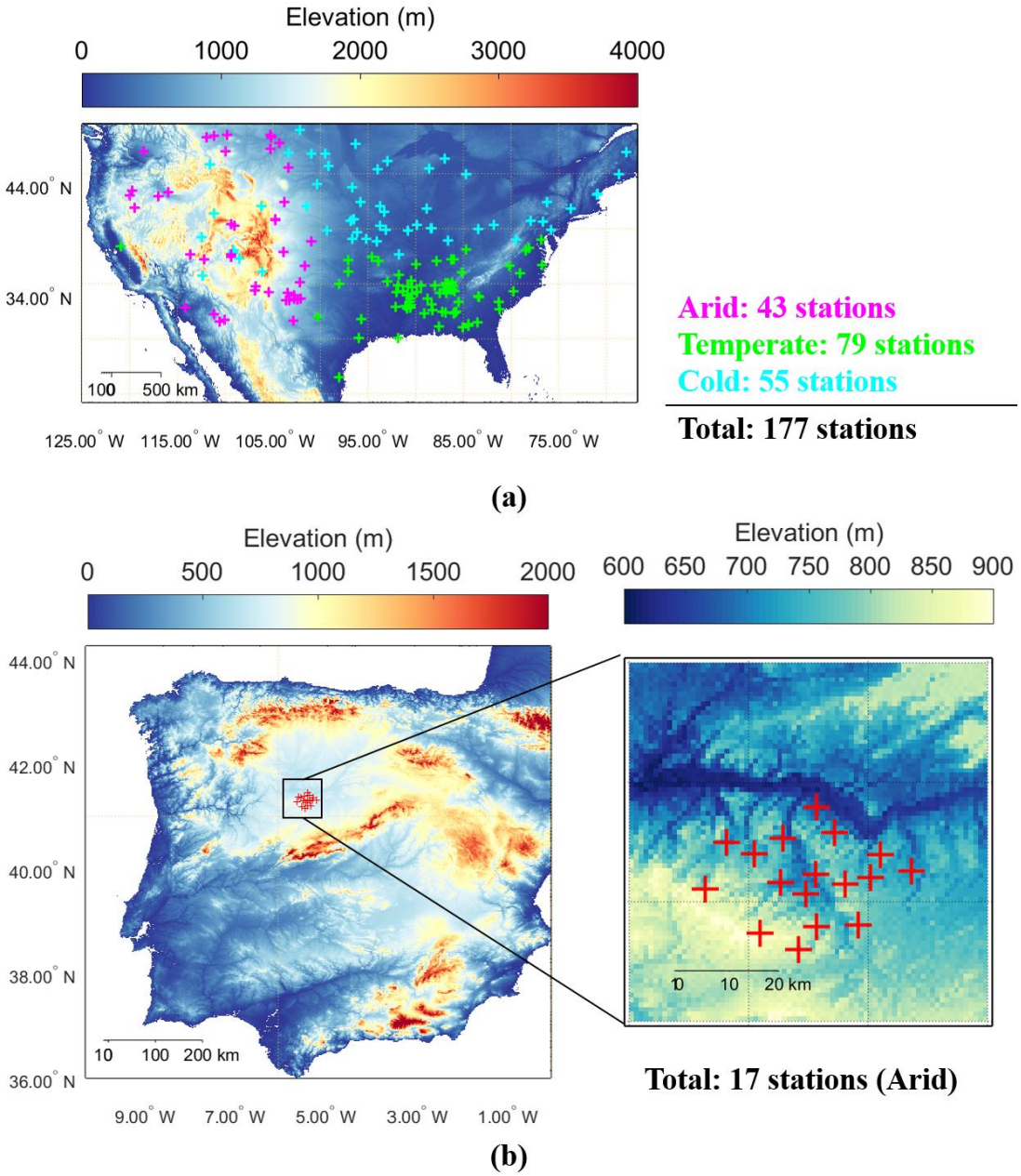


Figure 6.1. Locations of ground stations (red crosses) used for validation. (a) 177 stations over the CONUS. (b) 17 stations over the REMEDHUS network in Spain. Each location is classified into the climate zones by the updated world map of the Koppen-Geiger climate classification

Mediterranean and the network consists of more than 20 stations within the area and has been frequently used for validating disaggregated remotely sensed SM data (Piles et al., 2010; Sánchez et al., 2012) including (Peng et al., 2015).

The ground SM measurements were obtained from the ISMN measurements to which the data preprocessing steps in 4.2.2 were also applied. The timestamp of the CCI SM data in coordinated universal time was used to select the closest hourly ground soil moisture measurements. The daily observation timestamp was applied to select ground soil moisture measurements (hourly) that are closest to the timestamp. If there is more than one station in the fine grid cell, the measurements were averaged on daily basis. Only stations with at least 100 observations coincident with the coarse SM data and with positive temporal correlations were used. The distribution of the selected stations for each study area (177 for the CONUS and 17 for the REMEDHUS) is shown in Figure 6.1.

6.2.3. MODIS Vegetation Index

The MODIS-derived 16-day composite NDVI (MOD13Q1) was used for this study. As it consists of 10°×10° tiles in the sinusoidal projection, they were reprojected and resampled to both coarse (0.25°) and fine scales (0.05°) using the re-projection tool (Dwyer & Schmidt, 2006). It was used only quality assured (QA) data of which pixel-level QA code (Didan et al., 2015) is at least 'lower quality'. The pixel-level QA code for the MODIS NDVI product is a 16-bit binary code flagging the land surface/atmosphere conditions and representing quality levels. The 'lower quality' is the second level after the 'highest quality' among eleven grades in total. Note that, for the fine scale product, due to the large data size involved, only regions over coarse grid cells where an *in situ* station exists and 8 coarse grid cells surround it (*i.e.* 3×3 coarse grid cells where 15×15 fine grid cells coexist for each group of stations) were processed.

One of major issues for the MODIS products is the large gaps in time and space due to cloud masking. These gaps negatively affect the disaggregation process and hence a three-dimensional (3-D) gap filling method (Garcia, 2010) was simply adopted in a practical manner. This is a Penalized Least Square regression based on 3-D Discrete Cosine Transform (DCT-PLS) which solely relies on a smoothing parameter s (the higher the s, the smoother the result). In this study, the 16-day composite NDVI products over the study period were simultaneously filled by the DCP-PLS method using a s of 10^{-6} which was fairly applied for filling gaps in a SM product (Wang

et al., 2012). Then, the gap-filled NDVI maps were regarded as a series of snapshots at the middle of the composite periods and then linearly interpolated to daily time scale along the CCI SM timestamp at each grid cell. It should be noted that these simple processes for the NDVI data could introduce more uncertainty which may lead to poorer SM estimates, but this topic is beyond this study and could be addressed separately through various approaches. For example, Fang et al. (2013) used a sinusoidal fitting method for estimating daily values from the MODIS NDVI product. Weiss et al. (2014) developed an algorithm for filling gaps in MODIS EVI and LST datasets by utilizing ratios from nearby non-gap pixels derived at two points in time.

6.3. Methodology

6.3.1. Proposed Disaggregation Method

The rationale behind the proposed disaggregation method is that the disaggregated data should smoothly transition from one-time step to the next due to the SM memory interacting with regional vegetation conditions. To start, a simple linear relationship was formed to characterize coarse NDVI and temporally aggregated coarse SM within a past *n*-day window at each location as

$$\overline{SM}_{I,J,t} = L_{I,J} \cdot NDVI_{I,J,t} + C_{I,J} \qquad \text{Eq. 6.1}$$

where $L_{I,J}$ is a slope estimated by linear regression at a coarse grid cell of which spatial indices are *I* (row) and *J* (column) respectively. $C_{I,J}$ is an intercept and $\overline{SM}_{I,J,t}$ is an exponentially weighted temporal average of SM over the past *n* days prior to time *k* = *t* as

$$\overline{SM}_{I,J,t} = \frac{\sum_{k=t-n+1}^{t} \alpha_{I,J}^{k-t} \cdot SM_{I,J,k}}{\sum_{k=t-n+1}^{t} \alpha_{I,J}^{k-t}} \qquad \text{Eq. 6.2}$$

where $\alpha_{I,J}$(0-1) is a decay coefficient which reflects local properties together with temporal window size *n* (days, integer), and was adopted for more flexible fitting. Namely, Eq. 6.1 matches two variables by linearly scaling NDVI with the $L_{I,J}$ and $C_{I,J}$, and temporally aggregating/smoothing SM with the *n*-day window and $\alpha_{I,J}$. Optimal coefficients can be obtained by maximizing the coefficient of determination (R^2) of Eq. 6.1, and results are

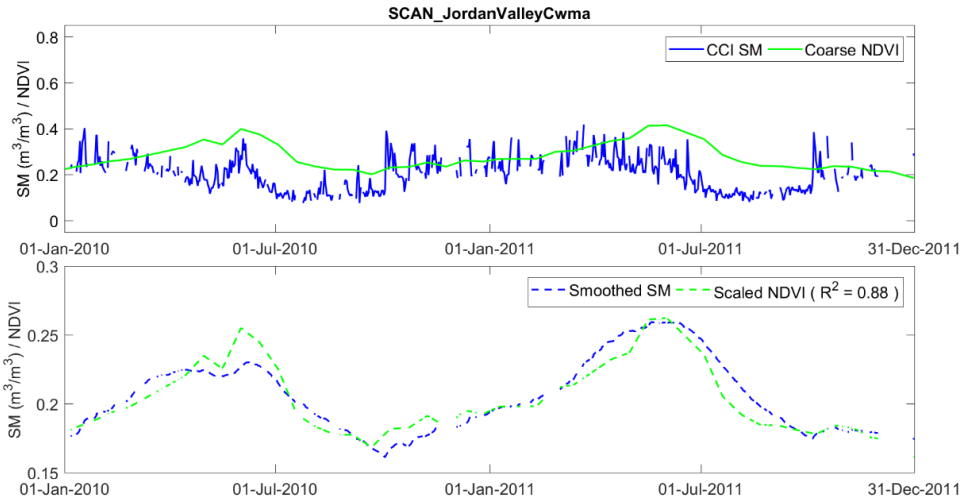


Figure 6.2. Relationship between scaled NDVI and temporally aggregating/smoothing SM at 'Jordan Valley Cwma' station

presented in the next section. Namely, given NDVI and coarse SM at a coarse grid cell, the four parameters can be considered as the best possible measures for statistically matching the two datasets.

Figure 6.2 provides an example of the relationship represented by Eq. 6.1 at a coarse grid cell (0.25°) located in the northwest of CONUS (42.95° N, 117.02°W). The climate zone is 'arid' and one of in situ stations from the Soil Climate Analysis Network (SCAN), 'Jordan Valley Cwma', is located within the coarse grid cell. The top panel of Figure 6.2 presents the NDVI and CCI SM datasets over the coarse grid cell. The NDVI timeseries was scaled by *L* and *C* in Eq. 6.1, and then the CCI SM was smoothed and aggregated by *n* and α in Eq. 6.2. The four parameters were optimized by maximizing R^2 between the scaled NDVI and smoothed coarse SM. As result, the maximized R^2 was 0.88, showing a good match between the two datasets.

Then, a finite difference between changes in NDVI and $\overline{SM}$ is adopted over two days, where $\Delta NDVI_{I,J,t}$ is the temporal difference in NDVI between $k = t$ and $k = t - 1$. Note that, the intercept $C_{I,J}$ is deleted in this case.

$$\overline{SM}_{I,J,t} = \overline{SM}_{I,J,t-1} + L_{I,J} \cdot \Delta NDVI_{I,J,t} \quad \text{Eq. 6.3}$$

The proposed disaggregation method links Eq. 6.2, Eq. 6.3, $L_{I,J}$ and $\alpha_{I,J}$ to the fine scale spatial resolution. Therefore, Eq. 6.2 and Eq. 6.3 are rearranged in terms of SM and NDVI at the fine spatial resolution as

$$\frac{\sum_{k=t-n+1}^{t} \alpha_{I,J}^{k-t} \cdot SM_{i,j,k}}{\sum_{k=t-n+1}^{t} \alpha_{I,J}^{k-t}} = \overline{SM}_{i,j,t-1} + L_{I,J} \cdot \Delta NDVI_{i,j,t} \quad \text{Eq. 6.4}$$

$$\frac{\alpha_{I,J}^{0} \cdot SM_{i,j,\mathrm{t}} + \sum_{k=t-n+1}^{t-1} \alpha_{I,J}^{k-t} \cdot SM_{i,j,k}}{\sum_{k=t-n+1}^{t} \alpha_{I,J}^{k-t}} = \overline{SM}_{i,j,t-1} + L_{I,J} \cdot \Delta NDVI_{i,j,t} \quad \text{Eq. 6.5}$$

$$SM_{i,j,t} = \left\{\overline{SM}_{i,j,t-1} + L_{I,J} \cdot \Delta NDVI_{i,j,t}\right\} \cdot \sum_{k=t-n+1}^{t} \alpha_{I,J}^{k-t} - \sum_{k=t-n+1}^{t-1} \alpha_{I,J}^{k-t} \cdot SM_{i,j,k} \quad \text{Eq. 6.6}$$

where $SM_{i,j,t}$ $NDVI_{i,j,t}$ are disaggregated SM and NDVI at a sub grid cell (i, j) and time *t*. When aggregated fine scale SM also needs to match the coarser scale value, so a re-scaling step is adopted to ensure consistency across both scales (Fang et al., 2013), and therefore quick responses in the daily coarse SM after rainfall can be directly propagated into the all sub-grids in proportion to the initially estimated SM values at all sub-grids by Eq. 6.6.

$$SM'_{i,j,t} = \left(SM_{i,j,t} - \frac{1}{p}\sum_{i,j} SM_{i,j,t}\right) + SM_{\boldsymbol{IJ},t} \quad \text{Eq. 6.7}$$

where $SM'_{i,j,t}$ is the corrected fine-scale SM and *p* is the total number of sub grid cells within a coarse grid cell, which is 25 (5×5) in the case of 0.25° to 0.05°.

Because the disaggregation by Eq. 6.6 requires a window of length *n* days for the calculations, there is a warm-up period prior to the study period until the first *n*-day data is available. For

these initial days, the coarse SM is disaggregated to the fine scale using the ratio of the fine scale and coarse scale NDVI on that day. The warm-up period was not considered in validation.

There are sometimes breaks in the timeseries of the CCI SM due to freeze-up and swath patterns of the instruments. These missing values within the *n*-day window produce large variability in the estimated temporal mean SM. To test this, 95% confidence limits on the mean SM are calculated using a two-sided *t*-test considering the number of available data in the *n*-day window. A small number of observations or a large standard deviation of SM values within the window increases width of the confidence limits. When the width is wider than ±0.2, the memory is regarded as unreliable and the coarse SM is disaggregated to the fine scale using the same method as the warm-up period disaggregation. The value of 0.2 is based on the threshold used for masking unreliable SM values under dense vegetation (De Jeu et al., 2008).

6.3.2. Summary of VTCI Based Disaggregation Approach

As discussed in the section 6.1, Peng et al. (2015) developed a SM disaggregation method which uses VTCI as a soil moisture proxy. The method assumes that fine scale LST and vegetation data can be combined using the VTCI to disaggregate the coarse soil moisture data. The method was chosen for validation purposes through a direct comparison with the results from this study, in terms of performance metrics and continuity in timeseries. A brief summary of this method and VTCI is presented here with full details in Peng et al. (2015).

The VTCI attempts to represent the joint relationship between LST, vegetation and soil moisture. Various VI could be used, but here only NDVI was considered. When pairs of LST (*x*-axis) and NDVI data (*y*-axis) at the coarse spatial scale are plotted, it generally forms a triangular or trapezoidal shape (Figure 2.1b) if it uses data from a large enough area to represent the entire range of SM and NDVI (Sandholt et al., 2002). This is called as the LST/NDVI feature space where a VTCI (0-1) is calculated by Eq. 6.8 with a pair of LST and NDVI at a sub grid cell (i, j, t).

$$VTCI_{i,j,t} = \frac{LST_{max} - LST_{i,j,t}}{LST_{max} - LST_{min}} \quad \text{Eq. 6.8}$$

where, LST_{max} and LST_{min} are the maximum and minimum LST conditioned by $NDVI_{i,j,t}$, which are linear regression lines as the dry and wet edges of the LST/NDVI feature space respectively. After calculating all VTCI across all i and j at time *t* by Eq. 6.8, a coarse $SM_{I,J,t}$ is simply disaggregated by linear proportion to the VTCI at the fine spatial resolution as Eq. 6.9 where $SM_{i,j,t}$ and $VTCI_{i,j,t}$ are disaggregated soil moisture and VTCI at the fine scale at the location (i, j, t). $SM_{I,J,t}$ is the corresponding coarse soil moisture at the location $(I,\ J)$ and the same time.

$$SM_{i,j,t} = p \times \frac{VTCI_{i,j,t}}{\sum_{i,j} VTCI_{i,j,t}} \times SM_{I,J,t} \quad \text{Eq. 6.9}$$

In the procedure, a level of spatial coverage for the LST data is to be assured for effectively implementing this method. Peng et al. (2015) used only LST data which covers at least 75% of the study area. Actually, it was shown the temporal coverage of disaggregated SM by the VTCI method remarkably decreases compared to the coarse SM as shown in the timeseries of Peng et al. (2015) (Figure A6.1), and it is this problem of missing LST data that the proposed method in this study addresses.

6.3.3. Evaluation Strategy

For both study areas, all SM estimates were evaluated against *in situ* measurements using temporal correlation (R), root mean square error (RMSE), bias and unbiased RMSE (ubRMSE) (Entekhabi et al., 2010b).

For the REMEDHUS area, the disaggregation results were directly compared with Peng et al. (2015) in terms of the four metrics, and spatial variability of disaggregated SM was evaluated as well. Correlations and separation distances were calculated for all possible pairs of station locations. The relationship of correlations with distance could then be checked for each product at the fine scale.

6.4. Results and Discussion

6.4.1. Parameter Optimization Results

For optimizing the parameters in Eq. 6.6, 1-year data of SM and NDVI before the study period was additionally used for considering the warm-up period. Table 6.1 shows the optimization results for each climate zone. All parameters (α, *n* and *L*) over the REMEDHUS area are reasonably stable and have high R^2. However, there is a much larger range of parameters and R^2 in the climate zones over the CONUS. It was found that R^2 is higher in the arid region than temperate and cold, and parameters vary as per the climate zones over the CONUS. Long windows were generally necessary for sufficiently smoothing the highly variable CCI SM to match the slowly-varying NDVI. In detail, the longest *n* values are required for the arid regions followed by cold regions and finally temperate regions. The arid regions were found to have generally larger *L* values than the cold and temperate regions.

Table 6.1. Statistics of optimal parameters for each study area.

Study Area	CC	#Stan.	Par. [a]	Med.	25th	75th
CONUS (177 stations)	Arid	43	α	1.00	1.00	1.00
			n	295	240	327
			L	0.12	0.06	0.20
			R^2	0.40	0.22	0.56
	Tem.	79	α	1.00	0.97	1.00
			n	194	157	244
			L	0.07	0.04	0.11
			R^2	0.20	0.07	0.44
	Cold	55	α	1.00	0.93	1.00
			n	254	202	326
			L	0.04	0.02	0.09
			R^2	0.25	0.11	0.47
REMEDHUS (17 stations)	Arid	17	α	0.98	0.98	0.98
			n	152	149	159
			L	0.50	0.38	0.52
			R^2	0.69	0.66	0.71

[a] Constraints for parameters: $0 \le \alpha \le 1$, $2 \le n \le 365$ (integer), and $0 \le L \le 1$ / CC = climate classification, Tem. = temperate, Par. = parameter, Med. = median, OOth = OO% percentile, #Stan. = number of stations

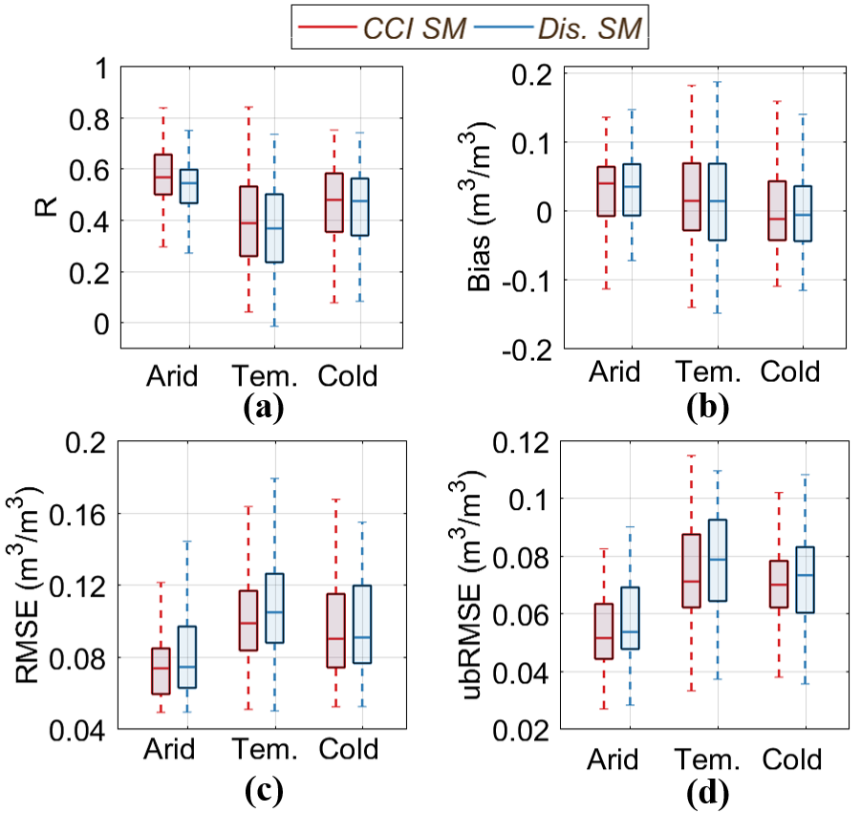


Figure 6.3. Box plots by three climate zones showing distributions of (a) R, (b) bias, (c) RMSE, and (d) ubRMSE of CCI SM (0.25°) and disaggregated SM (0.05°) against in situ SM from 177 ground stations over the CONUS (43 stations for arid region, 79 for temperate and 55 for cold)

6.4.2. Results of the CONUS

As shown in the box plots in Figure 6.3, the disaggregated SM are compared to the CCI SM in terms of the four metrics against all ground stations over the CONUS by the three climate zones (*i*.e. arid, temperate and cold). The CCI SM shows the best performance in the arid regions, the cold regions come second and the temperate, third. This sequence in performance by the climate zones directly propagates into the disaggregated SM with a slight performance degradation compare to the CCI SM. As the proposed SM disaggregation method only uses the NDVI product, the performance degradation is closely related with the quality and applied pre-processing of the NDVI product. Regarding this, to obtain better disaggregation results, it is necessary to conduct further investigations on the use of other VI products and associated pre-

processing steps such as gap-filling, temporal interpolation and quality assurance (Fang et al., 2013; Weiss et al., 2014).

Table 6.2. Disaggregation Results for REMEDHUS Area

Metric	R	Bias (m^3/m^3)	RMSE (m^3/m^3)	ubRMSE (m^3/m^3)
CCI SM (0.25°)	0.58±0.13	0.06±0.08	0.11±0.05	0.06±0.01
Dis. SM (0.05°)	0.49±0.14	0.06±0.09	0.11±0.04	0.05±0.02

*The results in the metrics are slightly different from Peng et al. (2015) due to the QC flags to the ground measurements and the temporal coverage of the disaggregated SM between the two approaches

6.4.3. Results of the REMEDHUS

The proposed disaggregation approach provides similar performances in terms of the four metrics (Table 2). This is in line with the results of the CONUS and means that the proposed SM-NDVI relation can replace LST information used in the latter approach. In addition, the temporal behavior of disaggregated SM is explicitly modelled in the proposed approach without breaks as shown in Figure 6.4a and b which are clearly compared to the results in Peng et al. (2015) (see Figure A6.1).

Figure A6.2 shows spatial patterns of disaggregated soil moisture by Peng et al. (2015) and the method proposed by this study on 22 May 2010. However, the spatial domain of this study only covers 3x3 coarse grid cells (*i*.e. 0.25°) in which the 17 REMEDHUS stations are distributed. Comparisons for similarities or differences between them can be made with a further implementation of the spatial disaggregation over a wider spatial domain. Regarding this, as shown in Figure 6.4c, it is presented spatial variability of in situ, CCI and disaggregated SM using the correlations of all possible pairs of station locations (*i*.e. 136 in this case) (*y*-axis) with distance (*x*-axis, km). Whereas the CCI SM has almost identical R over the pairs, the disaggregated SM properly captures the spatial variability of the in situ SM.

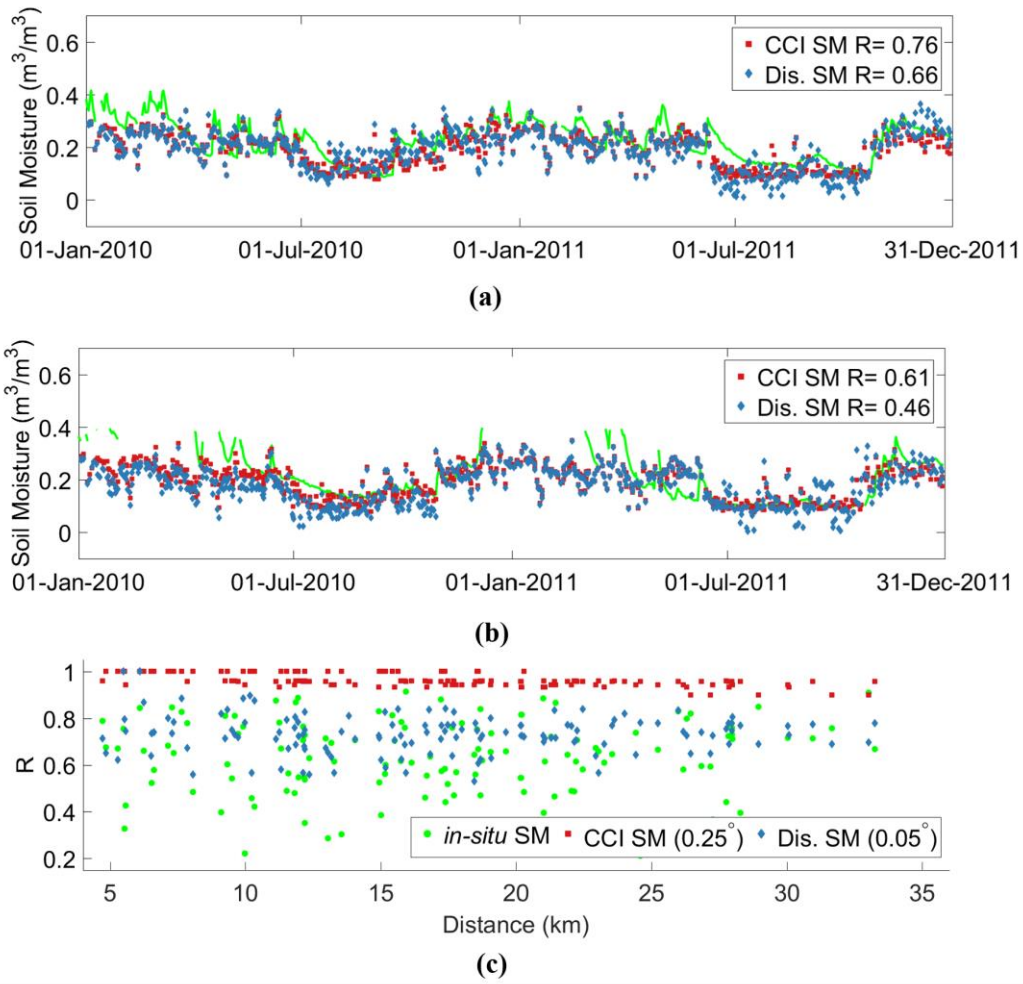


Figure 6.4. Timeseries of in *in-situ* (green), CCI (red), and disaggregated SM (blue) at (a) Canizal and (b) Guarena station in the REMEDHUS network. (c) Representation of spatial variability in *in-situ* (green diamonds), CCI, and disaggregated SM over the REMEDHUS network

6.5. Conclusions

In this study, the coarse data-derived SM-NDVI relation is proposed for spatially disaggregates coarse SM. The SM-NDVI relation is based on the optimal statistical matching of the scaled NDVI and smoothed CCI SM. The disaggregation results support the hypothesis that the SM-NDVI relation can replace the LST information in SM disaggregation. Accordingly, when LST data is unavailable or of poor quality, the proposed method provides a good alternative to existing approaches using LST data. Importantly, the proposed approach can provide continuous time

series of SM at the fine spatial resolution without losing the temporal variability of coarse SM. It also can correctly represent spatial variability in the disaggregated SM.

However, the work presented here is a proof-of-concept study to demonstrate the feasibility of the proposed method and there are a few future opportunities that could lead to better results. The uncertainty in the disaggregated product could be assessed using improved SM products (Kim et al., 2016b) and different vegetation indices such as Enhanced Vegetation Index and Leaf Area Index. Second, a simple gap-filling method (Garcia, 2010) and linear interpolation were used in the NDVI datasets. The method can lead to poor or physically unrealistic results with when gaps are large (Wang et al., 2012). It would be worth examining whether other gap-filling and interpolation methods (Fang et al., 2013; Weiss et al., 2014) could lead to better disaggregation performances.

6.6. Appendix

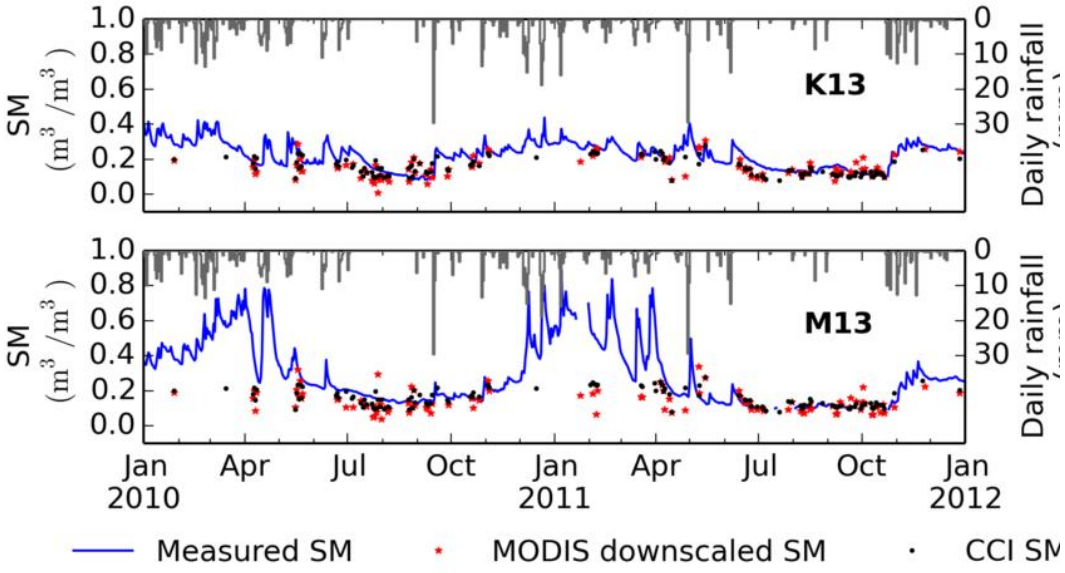


Figure A6.1. Disaggregation results from Peng et al. (2015). Timeseries of in *in-situ* (blue line), CCI (black dot), and disaggregated SM (red star) at (a) Canizal (K13) and (b) Guarena (M13) station in the REMEDHUS network.

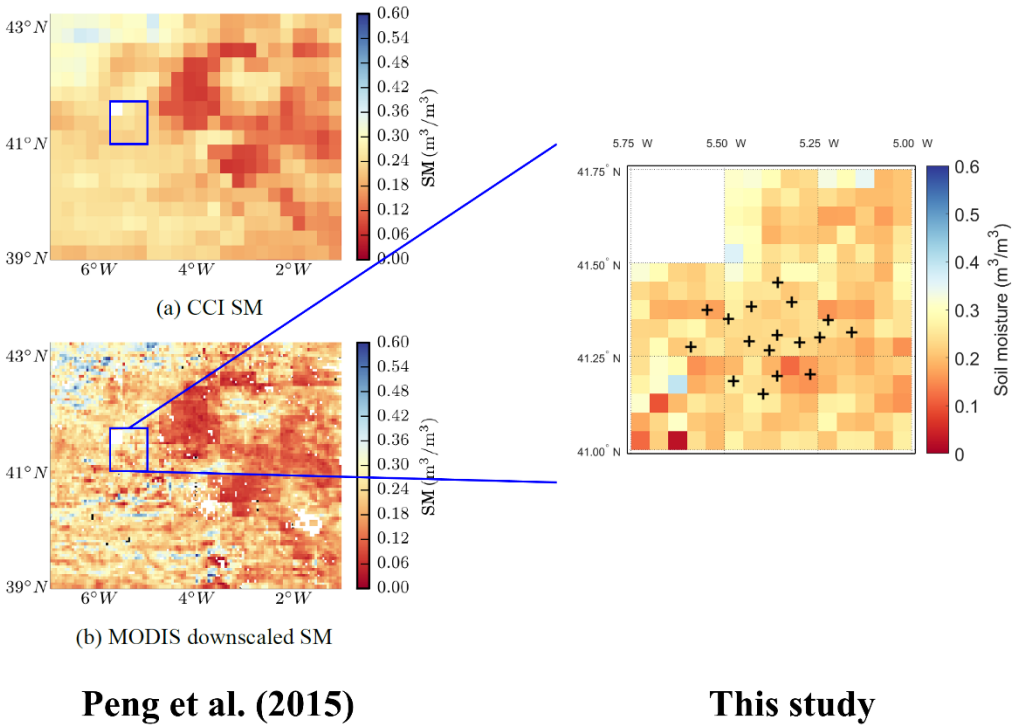


Figure A6.2. Comparison between spatial patterns of disaggregated soil moisture by Peng et al. (2015) (left-bottom) and the method proposed by this study (right) on 22 May 2010.

Chapter 7 Flood Warning in a Watershed Using Readily Available Soil Moisture and Rainfall Data

Soil moisture plays a key role in determining the antecedent condition of a watershed, while topographic attributes define how and where soil moisture and rainfall interact to create floods. Based on this, I present a method to identify flood risk at a location in a watershed by using only freely available remotely sensed or open access information such as soil moisture, rainfall, soil properties and topography. The method consists of three hydrologic modules, generation, transfer and accumulation of direct runoff, which are specified using only three calibration parameters for each catchment of interest. As the flood risk is achieved by providing a relative flood warning rather than simulating actual flowrates, the three parameters were optimized by maximizing the Pearson correlation coefficient between the simulated and observed flowrate. Two types of remotely sensed soil moisture data were separately applied for assessing how they affect model outcomes. The first is surface soil moisture available in near real-time for flood warning but less suited for determining antecedent conditions, and the second is the root zone soil moisture that is more suitable but takes longer time to be processed. As a proof-of-concept study, the proposed model was verified using contingency table-derived skill measures from the modelled and observed data series from a 2-year study period over 24 hydrologic reference stations in the Murray-Darling Basin in Australia. Although the case using root zone soil moisture data generally presented better performance in validation, using surface soil moisture in near real-time also showed promise as a means of providing quick warnings. Additionally, it was shown that the three calibration parameters are dependent on the watershed size and the soil moisture data used, allowing the shortcomings of using surface soil moisture for representing the antecedent condition to be mitigated where optimized parameters were available. More importantly, it was shown that the mean values of parameters estimated from neighbouring watersheds can be alternatively used for flood warning, allowing warnings to be issued in ungauged basins.

This chapter was submitted for publication: Kim, *S*.; Johnson, F.M.; Liu, Y.Y.; Sharma, A., Flood Warning in A Watershed Using Readily Available Soil Moisture and Rainfall Data.

7.1 Introduction

The Pakistan flood in 2010 left roughly 20% of the country's area under water, caused severe economic damage, and led to the loss of over 2000 lives and displacement of 20 million people (Syvitski & Brakenridge, 2013). As presented in Figure 7.1a-d, the satellite-derived SM clearly shows the flood evolution even at the coarse spatial resolution of 0.25° (approximately 25 km). One can observe changes in the SM images at pre-flood conditions (10 July), peak flood conditions (28 July – 4 August) and inundation conditions (13 August). As the event progressed, SM levels steadily increased, leading to more precipitation becoming runoff and ultimately producing extensive inundation (Figure 7.1e-h and i). This is consistent with the considerable literature on this topic, which states that antecedent SM affects runoff more significantly that often considered in simplistic modelling alternatives (Castillo et al., 2003; Pathiraja et al., 2012; Ruggenthaler et al., 2015).

For these reasons, antecedent SM and rainfall are critical elements to induce flooding in a watershed, and flood estimation using traditional hydrologic models requires information on rainfall and SM in space and time (Crow & Ryu, 2009). The combined effect of the spatiotemporal dynamics of SM and rainfall can explain most variability in runoff prediction and needs careful representation for accurate modelling outcomes (Durán-Barroso et al., 2016).

The case from the Pakistan flood suggests that using information on SM conditions could provide evidence of flood likelihood. Indeed, the use of remotely sensed and/or open access data for hydrologic prediction over remote and ungauged regions has been investigated in many studies across the world. For example, satellite-derived SM data has been used for determining reliable antecedent conditions of watersheds in rainfall-runoff modelling (Alvarez-Garreton et al., 2014; Brocca et al., 2010; Chen et al., 2011; Houser et al., 1998; Komma et al., 2008). Those

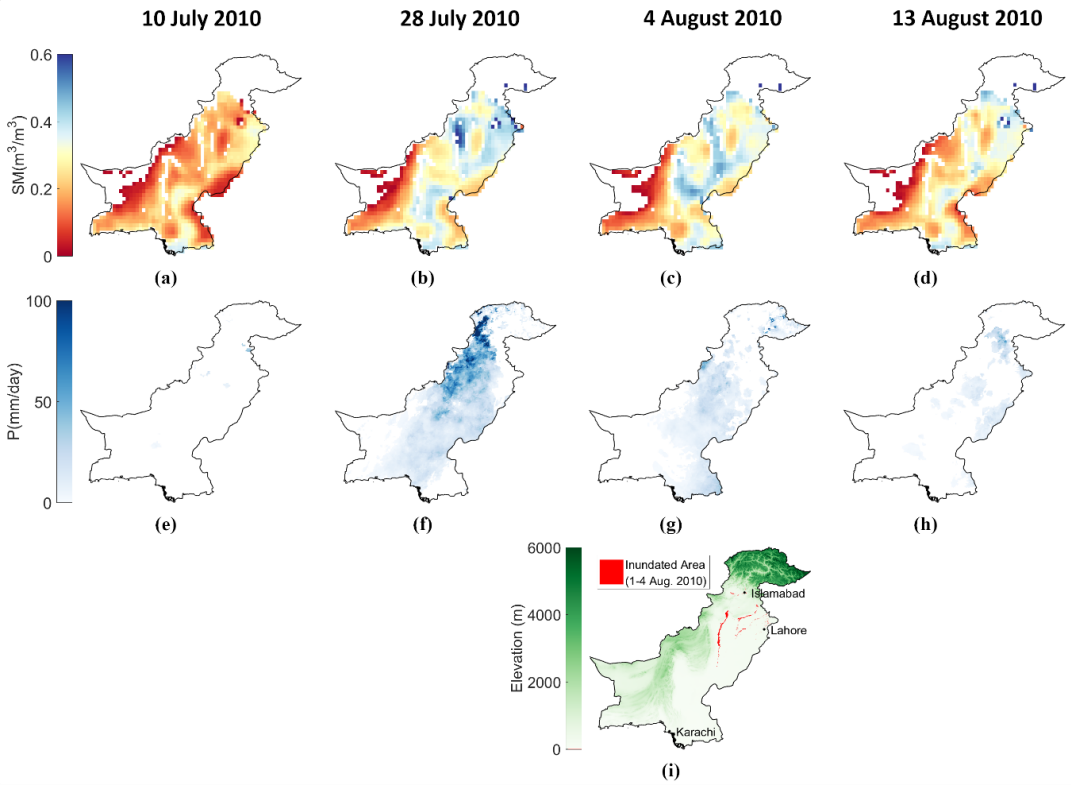


Figure 7.1. Spatiotemporal distributions of (a-d) soil moisture derived from the Advanced Microwave Scanning Radiometer - Earth Observing System at 0.25°, (e-h) rainfall from the CHIRPS final product at 0.05°, and (i) inundated area recorded from Dartmouth Flood Observatory during the Pakistan flood in July - August 2010

SM data were also assimilated into hydrological models for correcting biases in precipitation, initial condition or both (Crow & Ryu, 2009; Massari et al., 2014; Wanders et al., 2014). However, many previous approaches were based on lumped hydrologic models and have been applied for small watersheds, assuming spatiotemporally constant rainfall events over the area. While distributed models can be used for such applications, they normally require large amounts of data, information and parameters for accurately representing the processes involved (Francés et al., 2007). Such data are not fully available for the sparsely gauged or ungauged watersheds where flood warnings are often needed.

A novel approach is proposed in this chapter to address these limitations. A simple flood warning indicator is developed that provides a qualitative assessment of near real-time flood risk across a watershed. The flood warning indicator combines readily available satellite-derived

SM and rainfall products with open-access topographic and soil data. The flood warning indicator considers where and how much rainfall occurs, along with the transfer and accumulation of the excess rainfall after infiltration (Lagadec et al., 2016). To represent these process, three layers have been considered; 1) SM as the antecedent condition preceding the rainfall event, 2) rainfall as external forcing, and 3) topographic attributes defining how and where the two factors interact.

The objective of this study is to develop a model which generates flood warnings based on spatiotemporally variable remotely sensed rainfall and SM at a coarse spatial resolution over the catchment. This work is a proof-of-concept study focusing on multiple study areas but a short time-period. For this, the flood warning model was calibrated and tested over 24 hydrologic reference stations (HRS) in Murray-Darling basin in Australia (Turner et al., 2012) for a 2-year study period (each year for calibration and validation) to ascertain whether effective warnings are indeed possible or not.

7.2 Study Area and Data

7.2.1 Study Area

The Murray-Darling basin (MDB), consisting of three large rivers, the Darling River (2740 km), the Murray River (2520 km) and the Murrumbidgee River (1575 km), is a large semi-arid basin covering 14% of Australia's area (1.06×10^6 km^2). The basin is characterized by high natural hydroclimatic variability (Leblanc et al., 2012), and is regarded as Australia's food bowl, supplying one third of the national food supply with agricultural area covering more than 80% of the basin. As the MDB has been highly regulated, the proposed flood warning method was tested at 24 Hydrologic Reference Stations (HRS) in the MDB (Turner et al., 2012). The HRS have been selected as they have relatively long records of high quality flow data and have minimal regulation, land use change and other factors, with areas ranging from 130 to 22,885 km^2 and elevation, 284 to 1351 m (above sea level). Locations of the 24 HRS catchments are shown in Figure 7.2 with further details provided in Table A7.1.

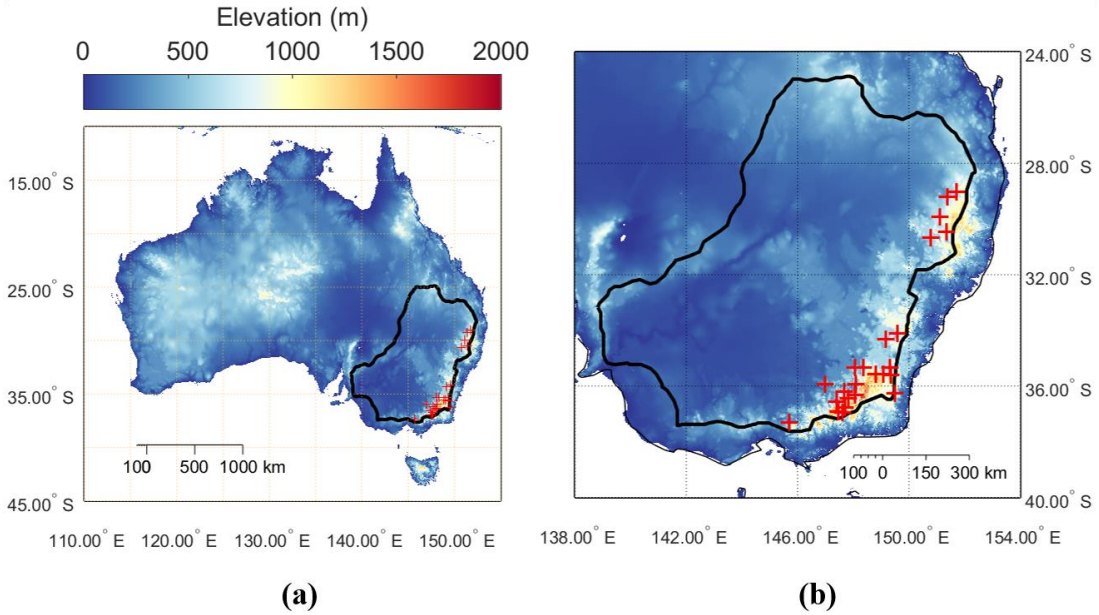


Figure 7.2. Locations of the 24 HRS stations in the MDB used for this study at (a) continental scale and (b) regional scale

7.2.2 Data and Processing

The required datasets covering the 2-year study period from 1 April 2015 to 31 March 2017 are summarized in Table 7.1. The first year of data was used for model calibration and the second year for model validation. Two types of SM data were investigated in the modelling; 1) real time or near real time data which for each data type is the product available that has the shortest latency. 2) data with higher quality due to further processing but not available as quickly as the other (longer latency). The aim of having these two SM data types is to investigate the quality of predictions that can be made in real time as well as the quality of the predictions that could be made using more accurate data. The latter case can be used to assess the model performance if improvements are realized in real time remotely sensed products.

A spatial resolution of 0.05° (approximately 5 km) was used as the common spatial resolution in this study. All data were resampled to this resolution using bilinear interpolation unless otherwise mentioned. The flowrate data for the 24 HRS stations were obtained from the Australian Bureau of Meteorology (http://www.bom.gov.au/waterdata/) for model calibration and validation which represent highly contrasting attributes in terms of flowrate magnitudes

and temporal variability. Mean and standard deviation over the 24 HRS stations for the calibration period are almost three times of those of the validation period.

Table 7.1. Summary of data used in this study

Data		Source	Name	Latency	Resolution (Temp./Spa.)	Units
Soil Moisture		AMSR2 LPRM	Level 3 Surface Zone SM	1 day	Daily/0.25°	m^3/m^3
		SMAP	Level 4 Root Zone SM	7 days	Daily/9km	m^3/m^3
Rainfall		AWAP	Gridded Daily Rainfall	1 day	Daily/0.05°	mm
Topographic attributes	Soil Property	HWSD	Topsoil proportion		-/0.0083°	%
	DEM	SRTM	DEM V4	-	-/3”	m (above sea level)
	Land Cover	MODIS	Land Cover (MCD12C1)	-	Yearly/0.05°	-

7.2.2.1 Soil Moisture Data

To specify antecedent conditions, the AMSR2-LPRM surface zone SM (SZSM) product at 0.25° retrieved from the C-band (6.9 GHz) microwave observations at the descending overpass (*i*.e. 1:30 AM) was used because it is expected that the soil moisture retrieved from the C-band will be more accurate than the X-band (10.7 GHz) (Parinussa et al., 2011c). Gaps in the SM product are inevitable due to the swath pattern of the instrument and the vegetation mask. Therefore the gaps in time and space over the study period were simultaneously filled by the 3-D gap filling method (Garcia, 2010) as per Chapter 6 .

For comparison with the AMSR2-LPRM SZSM, this work also considered a root zone SM (RZSM) product from the Soil Moisture Active Passive (SMAP) (Entekhabi et al., 2010a). It is known that assimilation of RZSM into a hydrologic model for defining the antecedent condition provides improvement in the model performance than using RZSM alone (Brocca et al., 2012). The SMAP

Level 4 9 km Equal-Area Scalable Earth (EASE) -Grid Surface and Root Zone Soil Moisture Geophysical Data, Version 2 (SPL4SMGP) (Reichle et al., 2016), for which the spatial resolution is 9km×9km was used. This product is generated every 3 hours (*i*.e. 8 SM images in a day) using an ensemble Kalman filter merging SMAP observations with SM estimates from the NASA Catchment land surface model (Reichle & Koster, 2005). The product contains time-average RZSM along with spatial coordinate information. The 8 images in a day were arithmetically averaged on a daily basis. Note that the SMAP data provides spatially fully covered RZSM (m^3/m^3), but the typical latency of the product is 7 days that is comparable with the AMSR2-LPRM SZSM data.

7.2.2.2 Rainfall Data

The 0.05° gridded rainfall data product from the Australian water availability project (AWAP) (Raupach et al., 2009) was used as the main rainfall forcing data in this study. The AWAP rainfall data is available from 1900 onwards as operational daily maps over Australia with a day of latency in typical, and has been used in several studies (Delworth & Zeng, 2014; Gergis et al., 2012; Ji et al., 2014). The AWAP product was developed using a thin plate smoothing spline for interpolating monthly rainfall climatology, and a successive correction method to interpolate anomalies of daily rainfall (Beesley et al., 2009).

7.2.2.3 Digital Elevation Models and Soil Data

The 90m DEM from the Shuttle Radar Topographic Mission (SRTM) (Jarvis et al., 2008) was used for defining the topographic attributes after resampling to 0.05° by nearest-neighbour interpolation. Based on the resampled DEM, the method proposed by Wang and Liu (2006) was first applied for filling surface depressions which can lead to disconnected stream-flow patterns. The topo-toolbox (Schwanghart & Kuhn, 2010; Schwanghart & Scherler, 2014) was used for delineating watersheds through terrain analyses including calculation of gradient, flow direction detection by the steepest descent method (O'Callaghan & Mark, 1984) and calculation of flow accumulation numbers at each grid.

As discussed in the next section, direct runoff was estimated using the National Resources Conservation Service - Curve Number (NRCS-CN) method. This study used the CN derived potential retention (*S*) which requires land cover and soil type data (Hong & Adler, 2008). The most recent MODIS land cover (LC) product at 0.05° spatial resolution (MCD12C1) (data from 2012) (NASA-LP-DAAC, 2012) was used. Soil attribute maps of silt, clay and sand at 90m spatial resolution were obtained from the Harmonized World Soil Database (HWSD) data (v1.2) (Nachtergaele & Batjes, 2012; Wieder et al., 2014). A CN map was prepared by classifying each 0.05° grid cell into 12 texture classes of the United States Department of Agriculture (USDA) soil classification system (Cronshey, 1986) by using the soil attributes maps. Then the USDA 12 texture classes were simply categorized into four hydrologic soil groups (HSG) A, B, C and D, and the CN of each grid cell was selected from the standard lookup tables (NRCS, 2010) according to the HSG classification and MODIS LC data. The estimated CN map was assumed to represent the normal antecedent moisture condition (AMC II).

7.3 Method

As shown in Figure 7.3, the proposed approach consists of three major modules for 1) runoff generation, 2) transfer and 3) accumulation. The first module generates direct runoff using the Soil Moisture Accounting (SMA) based NRCS-CN method (Michel et al., 2005) from the rainfall data at each grid with the antecedent conditions determined by the SM data. The spatiotemporally varying isochrones are generated by the second module using a modified version of the NRCS lag time equation (NRCS, 2010). The third module accumulates the direct runoff on a daily basis based on the isochrones which was adopted from Saghafian et al. (2002). Thus, the proposed flood warning system depends on accumulated direct runoff which is based on SM and rainfall during last a few days. Details of the model calibration and evaluation strategy are now presented. Note that the calculated direct runoff is expressed as a proxy ($\tilde{Q}$) because its variability and timing have been only considered for assessing the flood risk.

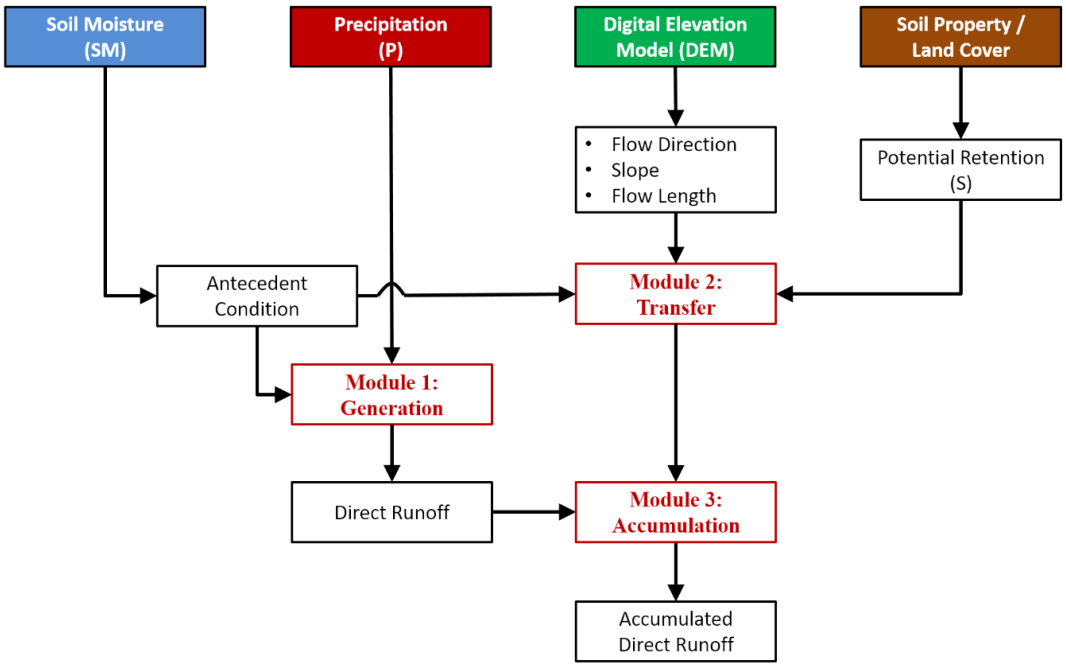


Figure 7.3. Flow chart of proposed method consisting of three modules

7.3.1 Module 1: Generation

The original NRCS-CN is a rainfall-runoff model developed by the Soil Conservation Service of the US Department of Agriculture (NRCS, 2004). In the model, the direct runoff (Q) is simply calculated by Eq. 7.1 as an explicit function of rainfall (P) and the potential maximum retention (S). The potential retention is transformed from the dimensionless curve number (CN) by Eq. 7.2. The initial abstraction (I_a) includes surface storage, interception and infiltration before runoff and is assumed to be a proportion of S by a coefficient λ, which was originally suggested equal to 0.2 (*i*.e. $I_a = 0.2S$). Units for all the parameters mentioned above (*i*.e. Q, P, S and I_a) are millimetres.

$$Q = \frac{(P - \mathrm{I_a})^2}{S + P - \mathrm{I_a}} \qquad \text{Eq. 7.1}$$

for $P > I_a$, otherwise $Q = 0$

$$S = \frac{25{,}400}{\mathrm{CN}} - 254 \qquad \text{Eq. 7.2}$$

The simplicity of this approach has resulted in widespread applications since it was first developed in 1949 (NRCS, 2004). As it is an event-based rainfall-runoff modelling process there have been difficulties in determining the antecedent SM condition of the watershed in relation to the maximum potential retention *S* (Michel et al., 2005). In the original NRCS-CN method, the CN-derived *S* represents the average Antecedent Moisture Condition (AMC II), and it varies from AMC I (drier) to AMC III (wetter) depending on the antecedent condition. Although the 5-day Antecedent Precipitation Index (API_5) had been commonly used to determine the AMC level, the need for alternatives has been suggested due to poor results derived from API_5 (Beck et al., 2009; Brocca et al., 2008; Durán-Barroso et al., 2016; Massari et al., 2014; Mishra et al., 2005).

As an alternative to the original method, Michel et al. (2005) proposed a simple method incorporating SMA into the NRCS-CN formulation. In this method, the initial SM condition is represented as SM store level (V_0) within the total soil storage which equals the sum of the maximum potential retention (*S*) and threshold for runoff generation (S_a). The amount of direct runoff (*Q*) generated depends on the relative differences between V_0, S_a and *P*, which are classified into three typical cases. The first case shows a condition in which sum of V_0 and *P* is lower than S_a, and then *P* is completely retained within the S_a column. In the second case where V_0 is lower than S_a but $P + V_0$ is greater than S_a, the initial amount of *P* is first captured in S_a and then the remaining part (*i*.e. $P + V_0 - S_a$) contributes to direct runoff (*Q*). In the third case when V_0 is higher than S_a, *Q* depends on relative differences in fluxes of *P* and *Q* into and out of the total soil column (*i*.e. $S + S_a$). Further details for these three cases can be found in Michel et al. (2005).

The main issue for the SMA model is to estimate V_0 which was not provided with an explicit form in the original work. In this study, as suggested by Durán-Barroso et al. (2016), V_0 was assessed by the degree of saturation (θ_e) as

$$V_0 = \theta_e \cdot S \qquad \text{Eq. 7.3}$$

Where θ_e is a scaled value (0 to 1) of SM (θ) with respect to the maximum (θ_{max}) and minimum observed SM (θ_{min}) at each grid cell (Brocca et al., 2008).

Michel et al. (2005) considered S_a as a constant portion of S (*i*.e. $S_a = 0.33S$) and Durán-Barroso et al. (2016) calibrated it as an independent parameter from S. In this study, S_a was regarded as a dependent parameter to S such as $S_a = \alpha S$, and the ratio (α) was optimized to improve the model calibration as presented in Figure 7.4. Once V_0 and P are given, direct runoff at each grid cell can be generated depending on the given condition of S_a.

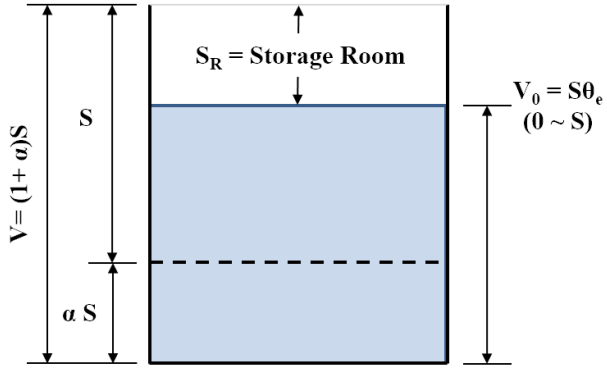


Figure 7.4. Schematic diagram of SM storage used in this study.

7.3.2 Module 2: Transfer

The NRCS time lag method was originally developed to calculate the time of concentration by considering flow length, slope and water retention by S over a watershed (NRCS, 2010). In this study, transfer of the generated direct runoff is based on grid to grid temporal lag (T_L) which represents grid to grid delay of flood effects. Adopting the approach to estimate the time of concentration, T_L (hour) was calculated as

$$T_L = \frac{l^{0.8} {S_R}^{0.7}}{\beta Y^{0.5}} \quad \text{Eq. 7.4}$$

where, l = flow length (km), S_R = soil storage room (mm), Y = average watershed land slope (%), β = time coefficient.

Instead of using S in the original equation, the soil water storage room (S_R), defined as $S_R = S + S_a - V_0$ in Figure 7.4, was alternatively used to calculate T_L by the SMA procedure incorporated in Module 1. Therefore, T_L is varying in space and time because it is calculated by the spatiotemporally varying S_R. Here, the time coefficient β is the second parameter to be optimized in this study.

7.3.3 Module 3: Accumulation

The calculated direct runoff for the last p days needs to be accumulated at a watershed outlet on a given day (t). For demonstration purposes, an example with 3-day of p is shown in Figure 7.5, with a detailed explanation for Module 3 following.

The storage room ($S_R(i, j, k)$) and direct runoff ($Q(i, j, k)$) matrices for last 3 days based on Module 1 are prepared in advance using rainfall and SM maps. The size of the matrices is $m \times n \times p$ of which spatial coordinates i are for rows (1 to m) and j for columns (1 to n), and k for time steps (t-3 to t). One can generate isochrones with the resulting S_R matrix and results from the terrain analyses (*i*.e. flow direction, slope and flow length) obtained through Module 2. In the condition at $k=t-3$, T_L is first calculated at the grid to grid basis by using temporal average $S_R(:, :, t\text{-}3{:}t)$ considering that the direct runoff generated at $k=t-3$ has been transferred to the outlet. Then total T_L from a grid cell to the outlet is calculated by accumulating each grid to grid T_L along the flow direction until it reaches the outlet. Once calculations over all grid cells in the watershed are completed, isochrones are formed by which the watershed area is divided into sections with discrete 24-hour units because the temporal resolution of rainfall and SM data is a day.

At $k=t$-3, the section of interest is the area between 72 and 96-hour isochrones which represent the total travel time to $k=t$ (3 days) and $k=t+1$ (4 days) respectively. Then the area overlapping with the rainfall field is regarded as the actual contributing area (at $k=t-3$) where

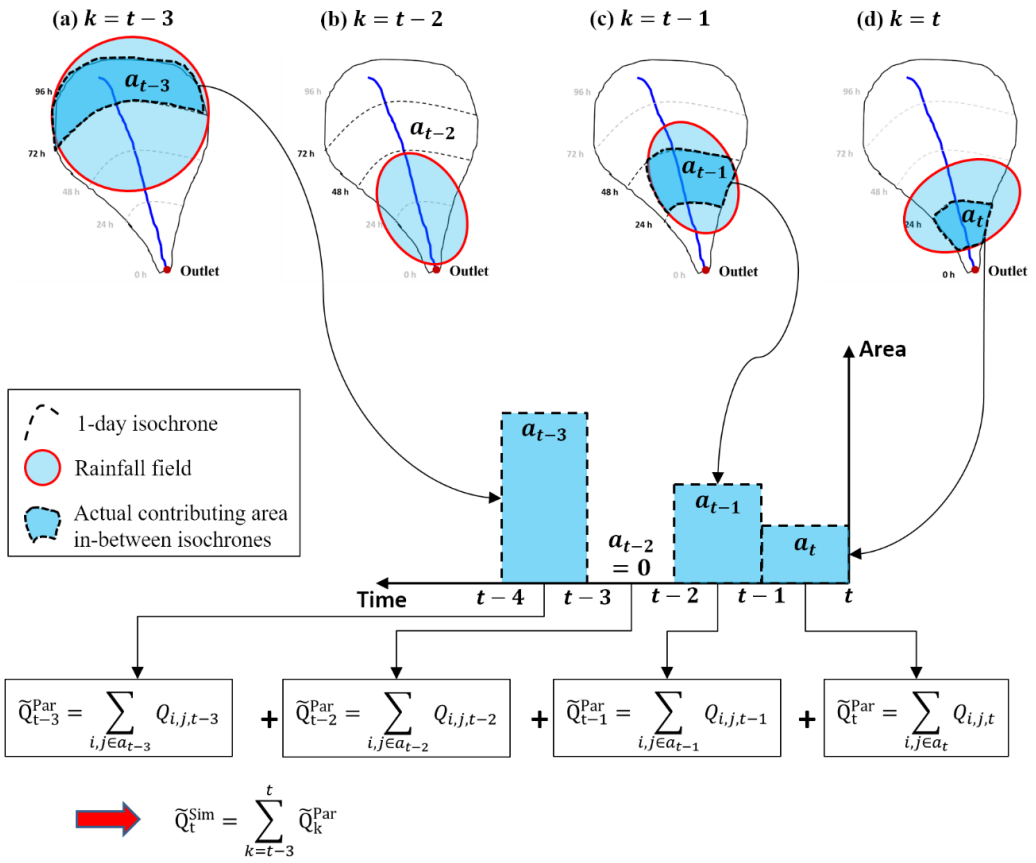


Figure 7.5. Spatiotemporally varying time-area calculation of Module 3

direct runoff is generated and integrated to $\tilde{Q}^{\mathrm{Par}}_{t-3}$ partially contributing to the total flowrate at the outlet at *k*=*t*. Here, $\tilde{Q}^{\mathrm{Par}}_{t-3}$ is generally expressed as Eq. 7.5 and defined as total summation of $\tilde{Q}_{i,j,k}$ that is direct runoff at a grid cell (*i*, *j*) within the actual contributing area at time *k* (a_k).

$$\tilde{Q}^{\mathrm{Par}}_{k} = \sum_{i,j \in a_k} \tilde{Q}_{i,j,k} \qquad \text{Eq. 7.5}$$

In the same manner, $\tilde{Q}^{\mathrm{Par}}_{t-2}$, $\tilde{Q}^{\mathrm{Par}}_{t-1}$ and $\tilde{Q}^{\mathrm{Par}}_{t}$ are sequentially calculated and then summed to $\tilde{Q}^{\mathrm{Sim}}_{t}$ as in Eq. 7.6. Note that $\tilde{Q}^{\mathrm{Par}}_{t-2}$ is in this example because a_{t-2} is zero even if it rains on the day. In this way, Module 3 is implemented for each day during the study period.

$$\tilde{Q}_t^{\mathrm{Sim}} = \sum_{k=t-p}^{t} \tilde{Q}_k^{\mathrm{Par}} \qquad \text{Eq. 7.6}$$

An important consideration in Module 3 is *p* for calculating $\tilde{Q}_t^{\mathrm{Sim}}$, which is the last (third) parameter to be optimized in this study.

7.3.4 Optimization of Model Parameters

In this study, there are three parameters to be optimized; 1) α for determining the threshold for runoff generation (S_a= αS), 2) β in Eq. 7.4 calculating T_L and 3) p in Eq. 7.6 calculating the accumulated direct runoff to the outlet.

As mentioned in the introduction, this study aims to provide a flood indicator rather than predicting flowrate at a location. For this reason, the three parameters were optimized to maximize the Pearson correlation coefficient (R) between timeseries of simulations ($\tilde{Q}_t^{\mathrm{Sim}}$) and observations ($Q_t^{\mathrm{Obs}}$) at each station (Legates & McCabe, 1999).

$$\text{Maximize } \mathrm{R} = \mathrm{f}(\alpha, \beta, p) = \frac{\mathrm{E}\left[(\tilde{Q}_t^{\mathrm{Sim}} - \mu^{\mathrm{Sim}})(Q_t^{\mathrm{Obs}} - \mu^{\mathrm{Obs}})\right]}{\sigma^{\mathrm{Sim}}\sigma^{\mathrm{Obs}}} \qquad \text{Eq. 7.7}$$

$$\text{Subject to } 0 \le \alpha \le 2;\ 1 \le \beta \le 30;\ 1 \le p \le 2p^0 \text{ (integer)}$$

where μ^{Sim} and μ^{Obs} are the mean values and σ^{Sim} and σ^{Obs} are the standard deviations of $\tilde{Q}_t^{\mathrm{Sim}}$ and Q_t^{Obs} respectively. The upper bounds of the constraints were set as double the published recommended values (Durán-Barroso et al., 2016; Michel et al., 2005; NRCS, 2010). For example, the maximum S_a is twice S, β is limited to 30 which is nearly twice the value of 16.88 by only the unit conversion of the original NRCS equation for calculating the time of concentration, and finally p^0 is the maximum number of days calculated by the original NRCS equation using the CN-derived S.

7.3.5 Model Evaluation Strategy

The novelty in the method is how the accumulated direct runoff is used to indicate flood risk. Rather than focusing on actual flood magnitudes, the method provides a relative flood warning

by considering the exceedance probability of a particular flow accumulation relative to all flow accumulations over time at that location.

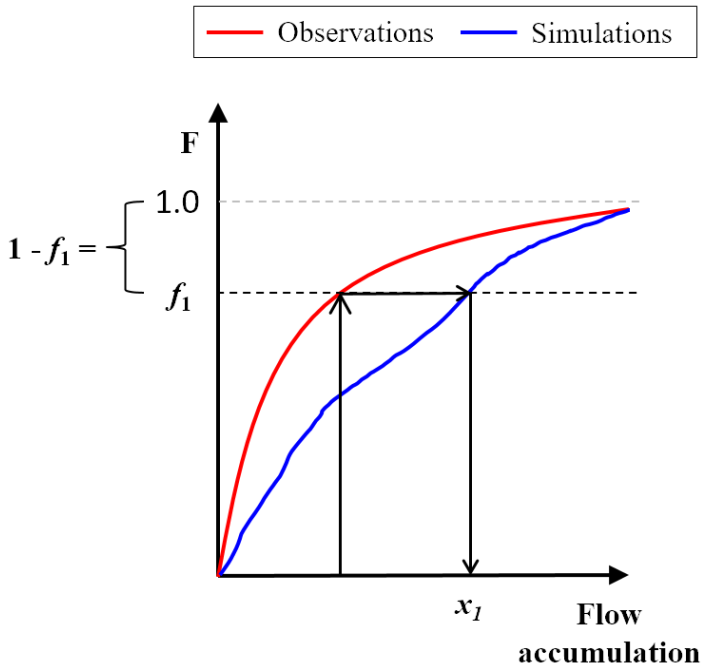


Figure 7.6. Schematic for flood warning in a watershed based on matching the observed probability of a flood event with the simulated flow accumulation CDF

For example, as presented in Figure 7.6, a flood warning threshold is set by mapping the historical floods in the catchment onto a cumulative distribution function (CDF) (red line) and adopting an exceedance probability (1-f_1) at which a flood occurs. This is then transferred to the model flow accumulation CDF (blue line) and the equivalent simulation flow accumulation threshold, x_1, is ascertained. By considering the thresholds at gauged locations within different catchments, a common threshold for regionalizing the flood trigger can be determined. The advantage of this approach is that the flood trigger indicator can then be applied at any point in the watershed because the proposed model can be spatially implemented.

As the time-period considered is 2 years for this proof-of-concept study, the evaluation strategy shown in Figure 7.7a and b was adopted by alternatively focusing on high flowrates during the period. For this, top 10% of simulated daily direct runoff ($\tilde{Q}_f^{\mathrm{Sim}}$) were first selected at

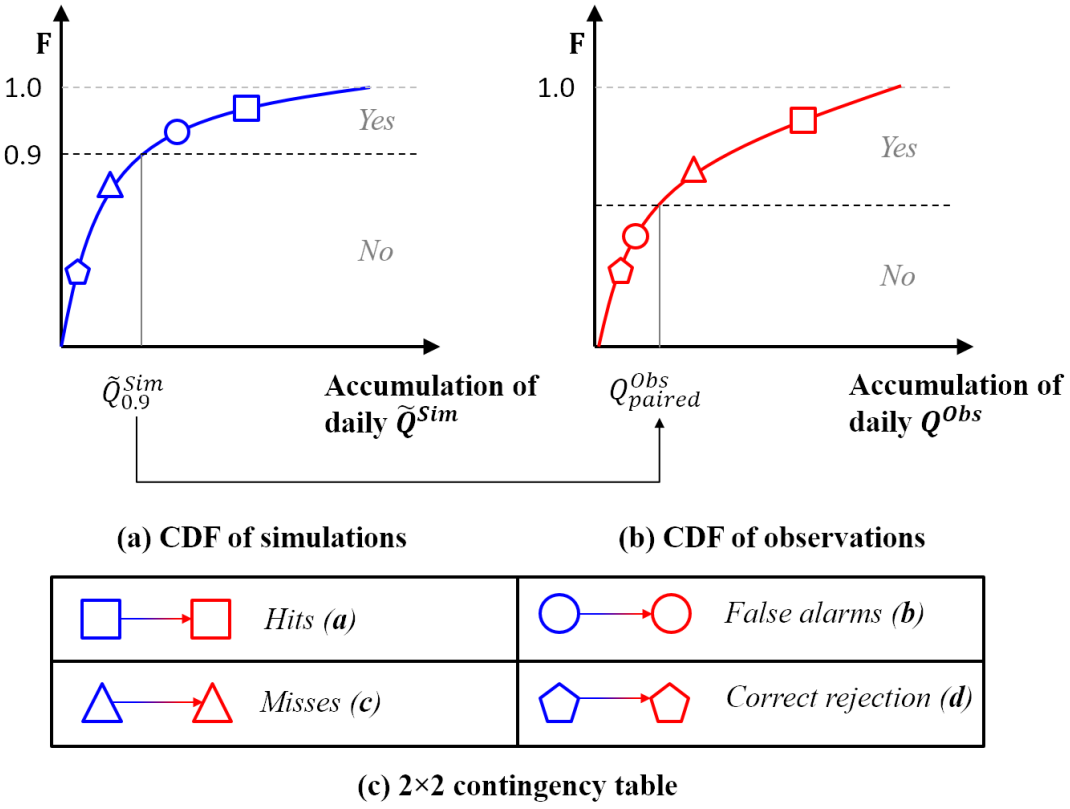


Figure 7.7. Cumulative distribution functions (CDFs) of daily (a) simulations and (b) observations, showing (c) 2x2 contingency table based forecast evaluations in terms of hits (a, squares), false alarms (b, circles), misses (c, triangles) and correct rejection (d, pentagons) respectively.

each station, and then the paired observation ($Q_{\text{paired}}^{\text{Obs}}$) regarded as a threshold for determining whether the flood occurred or not. Based on this, a 2×2 contingency table was established at each station for the model evaluation as shown in Figure 7.7c, consisting of hits (*a*), false alarms (*b*), misses (*c*) and correct rejections (*d*). If the simulation is in perfect linear correlation with the observation, both *b* and *c* are zeros in the table.

The contingency tables were used to assess the model performance using a number of performance attributes (Wilks, 2011). The proportion correct (PC) represents the model accuracy and is calculated by

$$\mathrm{PC} = \frac{\mathrm{a+d}}{n} \quad \text{range [0, 1], best: 1} \qquad \text{Eq. 7.8}$$

In cases where the number of events forecast is substantially less than non-occurrence flood events, the threat score (TS) is particularly useful. This is computed by

$$\mathrm{TS} = \frac{\mathrm{a}}{\mathrm{a+b+c}} \quad \text{range [0, 1], best: 1} \qquad \text{Eq. 7.9}$$

Together with PC and TS, other metrics considered include the frequency of hits (FOH) and probability of detection (POD) as shown from Eq. 7.10 and Eq. 7.11 (Bartholmes et al., 2009) for evaluating the proposed flood warning method.

$$\mathrm{FOH} = \frac{\mathrm{a}}{\mathrm{a+b}} \quad \text{range [0, 1], best: 1} \qquad \text{Eq. 7.10}$$

$$\mathrm{POD} = \frac{\mathrm{a}}{\mathrm{a+c}} \quad \text{range [0, 1], best: 1} \qquad \text{Eq. 7.11}$$

A total of three scenarios, S1 to S3, as shown in Table 7.2 according to the two SM data types were tested. S1, S2 and S3 commonly use the flowrate data for optimizing the three parameters. Among them, S1 and S2 are based on the SMAP RZSM and AMSR2-LPRM SZSM data respectively, which is for evaluating how the different SM datasets affect the parameter estimation and model capacity. S3 applies mean values of the optimal α and β parameters derived from S2 for

Table 7.2. Three scenarios applied in this study

Scenario	**Parameter** (α, β, *p*)	**Soil Moisture**	**Rainfall**
1	Optima for each station using given data	SMAP Root Zone SM	AWAP
2	Optima for each station using given data	AMSR2-LPRM Surface Zone SM	AWAP
3	Means of optima from Scenario 2	AMSR2-LPRM Surface Zone SM	AWAP

a purpose of parameter regionalization. However, *p* was used as it was obtained from S2 because *p* is related watershed sizes. Through the parameter regionalization, calibrated parameters from gauged basins can be transferred to ungauged ones based on geographical and functional similarity among the basins (Kim et al., 2016a).

7.4 Results and Discussion

7.4.1 Parameter Optimization Results

During the 1-year calibration period (1 April 2015 - 31 March 2016), the parameter optimization was implemented over the 24 HRS stations. Figure 7.8 presents the results for the maximised R between $\tilde{Q}_t^{\mathrm{Sim}}$ and flowrate for the calibration and validation periods, and the three parameters (*i*.e. α, β and *p*).

The maximised R of S1 is generally higher than S2, supporting that the RZSM provides better hydrologic responses compared to the SZSM which is in line with the results from Brocca et al. (2012). It is noteworthy that S3, using mean values of the optimal α and β parameters derived

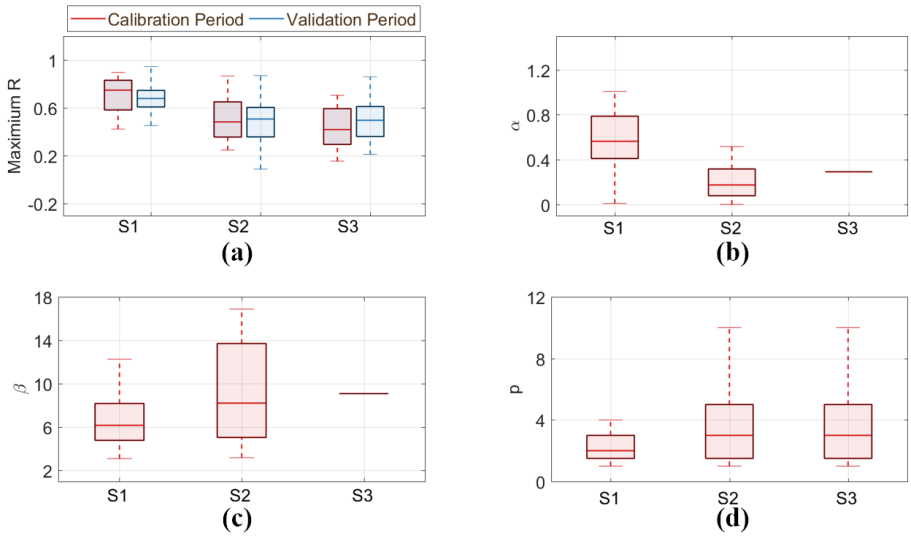


Figure 7.8. Parameter optimization results for S1, S2 and S3 using flowrate data as observations. (a) Maximised temporal correlation (R) between simulations and observations for the calibration (red) and validation period (blue), (b) α for determining the threshold for runoff generation (S_a= αS), (c) β for calculating the grid to grid temporal lag (T_L) and (d) necessary last *p* days for calculating $\tilde{Q}_t^{\mathrm{Sim}}$ to the outlet. The horizontal bars in (b) and (c) for S3 indicate mean values of each parameter from S2.

from S2, shows a similar range of R with S2 in both calibration and validation periods. This shows potential applicability of the near real-time SM data for flood warning and that alternative parameters derived from neighbours can be used in ungauged basins through the parameter regionalization.

The α and β parameters vary depending on which SM data has been used (Figure 7.8b and c). In S1, a higher α induces lower direct runoff, and lower β increases T_L compared to S2. On the other hand, the lower α and higher β in S2 works to increase water flux at an outlet. The reason for this is less persistence in the SZSM in S2 than the RZSM in S1. Quickly dried SZSM values underestimate V_0 over the region, and accordingly induce lower direct runoff and higher T_L. However, actual V_0 is not lowered as much as assessed due to the persistence of SM at lower levels resulting in higher actual direct runoff than the expectation. Accordingly, S2 makes up for the less persistence in the SZSM by increasing the water flux with the lower α and higher β parameters. With this, shortcomings of the SZSM for determining the antecedent condition can be mitigated. In case of p (Figure 7.8d), moderate to strong monotonic relations with the watershed sizes represented by Spearman's rank correlation coefficients (ρ), which are 0.63 and 0.81 for S1 and S2 respectively, were found.

7.4.2 Model Evaluation Results

Evaluation results for the 10% threshold for flood warnings are presented as box plots in Figure 7.9a-d in terms of the 2×2 contingency table-based skill scores (*i*.e. PC, TS, FOH and POD). The box plots show the four skill scores over the 24 HRS stations by S1, S2 and S3, and the calibration and validation periods.

S1 using the SMAP RZSM data generally shows better results than S2 and S3 at both calibration and validation periods. In addition to this, S1 provides similar or slightly better results at both periods resulting from the similar R of S1 during the calibration period (Figure 7.9a). The score similarity of S1 is mainly due to similar ratios of hits, false alarms, misses and correct rejections over both periods. However, S2 presents decreases in misses and increases in correct rejections while keeping similar rates of hits and false alarms from the calibration to validation

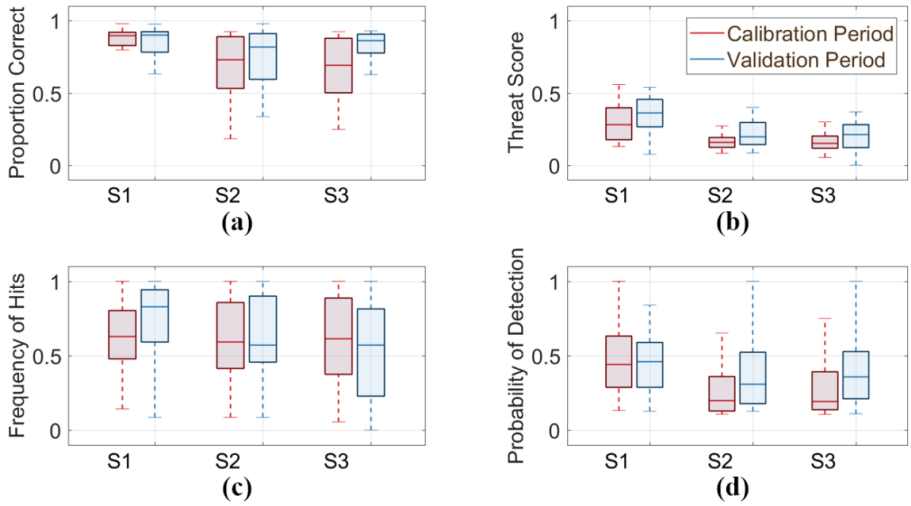


Figure 7.9. Estimated 2×2 contingency table based four skill scores for S1, S2 and S3 using flowrate data as observations with 10% of threshold by the calibration (red) and validation period (blue). (a) Proportion correct (PC), (b) threat score (TS), (c) frequency of hits (FOH), and (d) probability of detection (POD)

period. As the higher flood magnitudes during the calibration period form higher flood warning thresholds in the CDF of observations, more observations were classified as non-flooded cases. Accordingly, many misses were converted to correct rejections, leading to better or similar performance compared to the calibration period as shown in Figure 7.9. This tendency is stronger for S3 than S2, and results in better PC, TS and POD in the validation period, which are calculated based on misses and correct rejections. More importantly, the result that S3 presents similar performance as S2, which supports the argument that regionalized parameters can be potentially applicable over ungauged basins.

It should be noted that this entire study relies on the use of a short (2-year) dataset to assess performance, which is a limiting factor given the statistical nature on the approach. It can therefore be expected that with availability of longer and better-quality data, the results presented here can be improved further.

7.5 Discussion

This proof-of-concept study focused on assessing feasibility of the proposed method using the 24 HRS study areas over the 2-year calibration and validation period. From the above model

evaluation results, the following suggestions are made. First, the data-dependency of α and β parameters suggests that estimation and use of parameters should be considered separately for each SM and rainfall data combination to be used for later application. For example, in case the AMSR2-LPRM SZSM and AWAP rainfall data are used in a watershed, it is more suitable to derive parameters using these data, rather than using other SM or rainfall products that may be available. Second, the physical limit of the SZSM for representing the antecedent condition can be controlled through the parameter optimization with lower α and higher β values. Third, the moderate to strong monotonic relation between *p* and the watershed size suggests that the *p* parameter can be regionalised as a function of watershed size. Then, as supported by the results from S2 and S3 showing that mean values of parameters estimated from neighbouring watersheds can be alternatively used for flood warning, in case of an ungauged basin having nearby gauged basins, alternative parameters regionalized from the neighbours can be usefully applied for issuing meaningful flood warnings.

Although the maximised R following parameter optimization for all scenarios were generally moderate/strong, weak R (< 0.3) were observed over a few stations especially for S2 and S3, which resulted in relatively low skill scores. For this, improvements in either or both data (*i*.e. SM and rainfall) can result in an improved flood warning system, with reduced improvement if only one of the two are changed to better products. For example, the daily AMSR2-LPRM SZSM product at 0.25° is a near real-time SM data used in this study after simply filling gaps (Garcia, 2010) and then resampling to 0.05° with bilinear interpolation, which offers a large room for improvement in terms of quality (Enenkel et al., 2016; Kim et al., 2015b, 2016b; Liu et al., 2011b).

It should be noted that, even though the AWAP rainfall data used here has a high spatial resolution of 0.05° and is available in near real-time, it only covers the Australia domain. The 0.05° gridded daily precipitation data from the Climate Hazards group Infrared Precipitation with Stations (CHIRPS) (Funk et al., 2015) can be used for applications over other regions. The CHIRPS precipitation data is available from 1981 onwards with the quasi-global coverage of 50°N-50°*S*, and the latest version of data, Version 2.0, was released in February 2015

(http://chg.geog.ucsb.edu/data/chirps/). However, the final product is produced with an approximate 3-week latency using all available station data but the preliminary product is produced with a 2-day latency using sparse gauge data only. Some rainfall products are available in near real-time but the coarse spatial resolution and low accuracy limit their application (Duan et al., 2016). Improvements in characterizing rainfall extremes using the Global Precipitation Mission (GPM) (Andrea et al., 2016) offer hope that better rainfall products with low latency will be made available in not too distant a future.

7.6 Conclusions

In this study, a method was presented for flood warning in a watershed using readily available SM and rainfall products, using open access data derived soil property and topographic information at coarse spatial scales. This flood warning method consisted of three modules for generation, transfer and accumulation of direct runoff. The novelty of the method was the use of an accumulated direct runoff for indicating flood risk at a location. This provides a flood warning based on the assessment of an exceedance probability, rather than predicting actual flood magnitudes that require greater model complexity and input accuracy. To assess the performance of this proposed approach, the sensitivity of the response under given SM and rainfall has been evaluated. For the model development, three parameters in the model were optimized by maximizing the temporal correlation between simulated values and observed flowrate from 24 hydrologic reference stations in the MDB over the 1-year calibration period. The outcomes from the proposed warning system show promise, with possibilities of extension to completely ungauged locations where no past flow observations exist.

The proposed method, however, suffers from a number of limitations which can be addressed through further development. First, extended analyses at decadal-scales are required for applying the exceedance probability based flood warning system and estimating more representative parameters. Second, further investigations are necessary by applying various combinations of SM and rainfall products over watersheds having different attributes to those in MDB chosen in this study. This will shed greater light on the need for more sophisticated

regionalisation alternatives than the simplistic nearest catchment approach used here. Third, improved SM and rainfall data needs to be incorporated into the proposed flood warning method, as and when new products are released. Forth, satellite-derived flood signals can be alternatively used for optimizing the parameters to be applied for ungauged basins in which direct measurements of flowrates are not available. For example, Khan et al. (2012) used microwave radiometer-detected water surface signal (Kugler & De Groeve, 2007) which are highly correlated with observed runoff for calibrating a hydrological model. Similarly, Revilla-Romero et al. (2014) attempted to convert the water surface signal to river discharge, which could again form the basis for application to remote and ungauged catchments. Lastly, model refinements are needed when extending to regions having markedly different flood-causing mechanisms (such as snowmelt for instance).

7.7 Appendix

Table A7.1. List of 24 the HRS stations within the MDB used in this study

No.	Station ID	Name	[a] Lat. (°)	[b] Lon. (°)	[c] Ele. (*m*)	Area (km^2)
1	218001	Tuross River @ Tuross Vale	-36.26	149.51	942.92	90.00
2	401009	Maragle Creek @ Maragle	-35.93	148.10	594.62	216.00
3	401012B	Murray River @ Biggara - Backup	-36.32	148.05	519.21	1257.00
4	401015	Bowna Creek @ Yambla	-35.92	146.98	217.49	290.00
5	401203	Mitta Mitta River @ Hinnomunjie	-36.95	147.61	706.87	1519.00
6	401208	Cudgewa Creek @ Berringama	-36.21	147.68	519.59	351.00
7	401210	Snowy Creek @ Below Granite Flat	-36.57	147.41	575.53	416.00
8	401212	Nariel Creek @ Upper Nariel	-36.45	147.83	795.49	252.00
9	401216	Big River @ Jokers Creek	-36.93	147.47	902.06	357.00
10	401217	Gibbo River @ Gibbo Park	-36.76	147.71	779.65	390.00
11	405209	Acheron River @ Taggerty	-37.32	145.71	326.26	629.00
12	410057	Goobarragandra River @ Lacmalac	-35.33	148.35	596.06	668.00
13	410061	Adelong Creek @ B@low Road	-35.33	148.07	320.12	148.00
14	410705	Molonglo River @ Burbong	-35.34	149.31	761.92	509.00
15	410730	Cotter River @ Gingera	-35.59	148.82	1221.46	130.00
16	410731	Gudgenby River @ Tennent	-35.57	149.07	717.11	672.00
17	410734	Queanbeyan River @ Tinderry	-35.61	149.35	925.59	516.00
18	412050	Crookwell River @ Narrawa North	-34.31	149.17	584.69	762.00
19	412066	Abercrombie River @ Hadley No.2	-34.11	149.60	749.97	1630.00
20	416003	Tenterfield Creek @ Clifton	-29.03	151.72	797.93	559.00
21	416008	Beardy River @ Haystack	-29.22	151.38	412.59	908.00
22	418005	Copes Creek @ Kimberley	-29.92	151.11	632.94	249.00
23	418014	Gwydir River @ Yarrowyck	-30.47	151.36	879.99	835.00
24	419005	Namoi River @ North Cuerindi	-30.68	150.78	397.10	2532.00

* a: latitude, b: longitude, c: elevation above sea level

Chapter 8 Conclusion

This chapter presents a summary of the main achievements from this thesis for the research objectives presented in Chapter 1. Then an outline of the research currently ongoing and planned in the future is presented, along with a discussion of the limitations associated with this research.

8.1 Achievements and Summary

This thesis was initially motivated by how to assess flood risk over sparsely gauged or ungauged basins. This required a methodology for forecasting and key information resources (rainfall and soil moisture) that could be made available without delay (and expenditure) for use in remote locations around the world. Accordingly, the research first focused on near-real-time open-access soil moisture products that had global coverage. The spatiotemporal availability of passive microwave derived soil moisture products in near real-time at the global scale was identified as a viable alternative to ground measurements. However, significant improvement was needed in order to make an impact on the quality of the flood forecasts that could be issued using these measurements. This thesis therefore represents a series of efforts to use satellite soil moisture products for assessing flood risk by redressing their drawbacks in terms of accuracy and spatial resolution. The main achievements in this thesis are summarized as follows as well as in Table 8.1.

First, as reported in Chapter 3 , it was found the differences in performance of two alternate soil moisture products, JAXA and LPRM, which are derived from the same passive microwave observations of the AMSR2 radiometer, were a result of key uncertainties in each of the retrieval algorithms used. Notwithstanding the identical raw observations in time and space, and the simple radiative transfer model which is the common basis for both algorithms, these differences were found to be significant. The chapter explained why this is so merely documenting their performance in terms of error statistics against ground measurements. It was found that the reasons lie in the formulation of the different retrieval algorithms. These

Table 8.1. Summary of achievements in this thesis

Objective			Chapter	Main Achievements	Publication
1	Validation		3	· Algorithm based investigation on differences in soil moisture products · Identification of complementarity · Basis for improvement	· Kim et al. (2015a)
2	Improvement	Accuracy	4, 5	· Presentation of simple linear combination based methods for improving R of soil moisture in static and dynamic manner	· Kim et al. (2015b) · Kim et al. (2016b)
		Spatial Resolution	6	· Presentation of soil moisture memory based disaggregation method providing continuous timeseries and spatial variability by only using vegetation index	· Kim et al. (2017)
3	Application		7	· Presentation of flood warning method using operational soil moisture and rainfall data, and open access data derived information	· Kim et al. (submitted)

differences are a result of variations in parameterizing vegetation, surface roughness and physical surface temperature, and then linking a dielectric constant to soil moisture, with the two products showing complementarity in their performance with temperature, surface roughness, vegetation and ground wetness. The main contribution of Chapter 3 was that the complementarity in the performance of the soil moisture products was identified practically and accordingly provided the basis for the combination based improvement in soil moisture reported in Chapter 4 and Chapter 5 .

Based on this complementarity, a static linear combination was presented in Chapter 4 as a simple way to take the strengths of each soil moisture product into a combined dataset. Two soil moisture products were linearly combined into a new product by applying a weight (0-1) to each original product such that the summed weight equals one. The weights are optimal to

maximize correlation between the combined product and a reference, and represents the relative strengths (and weaknesses) of each soil moisture product against the other. The global mean temporal correlation coefficient of the combined product with ERA-Interim as the reference was 0.52 which outperformed the individual JAXA (0.35) and LPRM (0.45) products. Further to this, the complementarity between the two original products (JAXA and LPRM) varied not only in space but also in time because it changed with regional conditions such as vegetation, temperature and soil wetness.

Chapter 5 presented an extended method to the static combination approach in Chapter 4 to reflect the time-varying complementarity in the soil moisture products. This was achieved by using a temporally dynamic data segment as the basis for combination instead of the entire data. The dynamic data segment is presumed to have more relevant information than the entire dataset assuming model structure inadequacies associated with JAXA and LPRM vary over time, and was used for calculating the time-varying weights at a spatiotemporal point of interest. The mean R of the dynamically-combined products were 0.57 and 0.62 for the ERA-Interim and MERRA-Land reanalysis products as the reference respectively; higher than those of the statically-combined products (0.55 and 0.54).

In Chapter 6 , an approach was presented to spatially disaggregate coarse soil moisture by only using a remotely sensed vegetation index product. The method provides a continuous timeseries of disaggregated soil moisture with a persistence structure closer to what is observed. This is necessary because existing disaggregation approaches based on optical sensor-derived LST have many missing values. The large number of missing values is a result of cloud masks in the LST data and they induce a discontinuity in the timeseries of disaggregated soil moisture. It was found that the soil moisture memory (Koster & Suarez, 2001) interacting with regional vegetation could be an alternative to the LST. This idea was validated using ground measurements over the contiguous United States and REMEDHUS network in Spain. The disaggregated soil moisture represented better spatial variability compared to the coarse soil moisture.

Chapter 7 presented a method for flood warning within a watershed by incorporating readily available soil moisture and rainfall data, using open access data derived information into a series of flood warning calculations. It consisted of three modules for generation, transfer and accumulation of direct runoff. Three parameters in the model were optimized by maximizing temporal correlation between simulated and observed flowrate. Using the 24 Hydrologic Reference Stations in the Murray-Darling Basin, it was demonstrated that there are clear opportunities for the proposed method to be of use for the purpose of flood warning in ungauged watersheds.

8.2 Limitations and Ongoing/Future Research

First of all, it is necessary a consistent selection of data, study area and period which has not been achieved throughout the chapters due to the data availability and/or validation purpose at the time of each study. On the premise of this, Figure 8.1 shows the big picture that is eventually pursued based on the achievements from this thesis. First, for the improvement of the soil moisture products in Chapter 4 , 5 and 6, it is necessary for all proposed methods in this thesis to be integrated for producing an improved soil moisture product in terms of accuracy and spatial resolution. Further research is required to incorporate these into a single unified product which can be practically used in many hydrological applications. Then, the flood warning method proposed in Chapter 7 also needs to evolve to be applicable in ungauged basins. The following further investigations are necessary to realize the grand aim of this thesis, which is to provide real-time flood warnings to rural and remote communities around the world.

Ideally, the work reported in Chapter 3 could be extended using multiple-year data instead of the single year data that was used. This would ensure a more comprehensive validation and examination of inter-annual variations, and provide better information for using and improving the soil moisture products in the future. The use of multiple-year data would also further the validation of the combined soil moisture products reported in Chapter 4 and 5. The 2-year data only allows limited validation using ground stations in which the AMSR2-LPRM product is found

better. More robust results would be obtained by using the temporally extended AMSR2 soil moisture data for which temporal coverage is now almost 5-years at present (February 2017).

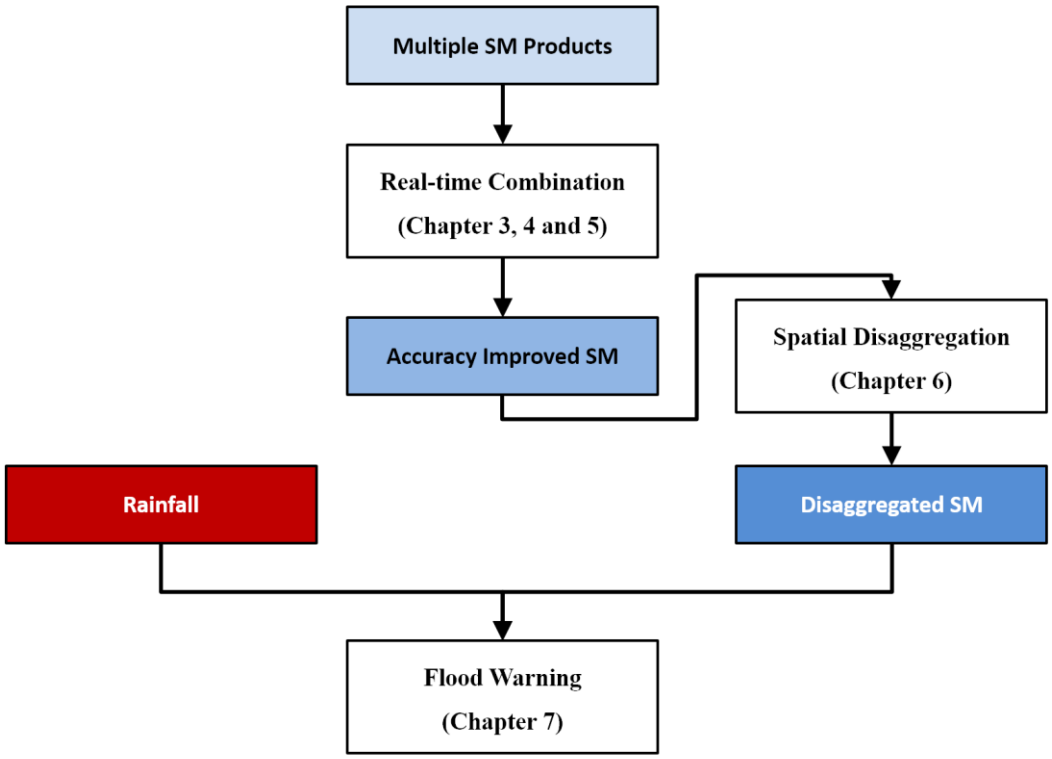


Figure 8.1. The bigger picture of the proposed flood warning system that uses improved soil moisture inputs in terms of both accuracy and spatial resolution

In addition to this, possible improvement exists for the combination based approaches in Chapter 4 and 5. First, three or more soil moisture products could be considered for the linear combination. This would add the strengths of each product, resulting from different algorithms and observational instruments. More importantly, using soil moisture data to estimate optimal weights for the combination in close to real-time would provide greater accuracy in case of operational flood forecasting. The reference datasets used for the static/dynamic linear combination are reanalysis products that are not available at the time of soil moisture retrievals as a few months are usually needed to prepare the reanalysis products. Therefore, a method is required for estimating optimal weights without such reference data, which could be an indirect way to use other applicable information such as regional vegetation, temperature and wetness condition.

For the spatial disaggregation of coarse soil moisture, the soil moisture memory has been considered as an alternative to the optical sensor derived LST data where large gaps exist consequently causing frequent discontinuity in timeseries of disaggregated soil moisture. This problematic situation can be also resolved by filling the gaps in the LST data that can be used in the existing disaggregation methods using LST data such as Fang et al. (2013) and Peng et al. (2015). An applicable approach for this gap-filling is to use passive microwave derived LST (Holmes et al., 2009). Optical sensing and passive microwave derived LST showed a good agreement in Parinussa et al. (2016), and the cloud penetrating capability of microwave and high spatial resolution of optical sensing can be combined in future work.

For the proposed flood warning method to be generally applicable, it should be supported by further investigations with other operational soil moisture and rainfall products under different geographical and climate settings. In addition to this, it would be useful for calibrating parameters to use satellite-derived signals of water level and flow, and/or another objective function so the proposed method can be made applicable to ungauged watersheds. Additionally, methods for reducing the latency of the data products that led to improved forecasts are required. This would allow the proposed method to be adopted for operational flood warnings in ungauged catchments where it is most needed.

8.3 Closing Remarks

It is important to have adequate spatiotemporal information on soil moisture to improve hydrological forecasting capacities. Even though (active or passive) microwave based soil moisture retrievals are a unique resource for providing surface soil moisture in near real-time at the global scale, improvements are still required. This thesis has attempted to answer this need through a series of investigations. Together with the thesis outcomes, ongoing and future research will form an integrated pathway for producing an improved soil moisture product available at finer spatial resolution, which can be used for various regional applications. Lastly, the proposed flood warning method has potential to be an applicable tool for ungauged

watersheds with further investigations and refinement, and it is also possible to be cast into a hydrological model for simulating actual flowrates.

References

Al-Yaari, A., Wigneron, J. P., Ducharne, A., Kerr, Y., de Rosnay, P., de Jeu, R., . . . Mialon, A. (2014). Global-scale evaluation of two satellite-based passive microwave soil moisture datasets (SMOS and AMSR-E) with respect to Land Data Assimilation System estimates. *Remote Sensing of Environment, 149*(0), 181-195. doi: http://dx.doi.org/10.1016/j.rse.2014.04.006

Albergel, C., de Rosnay, P., Balsamo, G., Isaksen, L., & Muñoz-Sabater, J. (2012a). Soil Moisture Analyses at ECMWF: Evaluation Using Global Ground-Based In Situ Observations. *Journal of Hydrometeorology, 13*(5), 1442-1460. doi: http://dx.doi.org/10.1175/JHM-D-11-0107.1

Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., . . . Wagner, W. (2012b). Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. *Remote Sensing of Environment, 118*, 215-226. doi: http://dx.doi.org/10.1016/j.rse.2011.11.017

Albergel, C., Dorigo, W., Balsamo, G., Muñoz-Sabater, J., de Rosnay, P., Isaksen, L., . . . Wagner, W. (2013). Monitoring multi-decadal satellite earth observation of soil moisture products through land surface reanalyses. *Remote Sensing of Environment, 138*(0), 77-89. doi: http://dx.doi.org/10.1016/j.rse.2013.07.009

Alvarez-Garreton, C., Ryu, D., Western, A. W., Crow, W. T., & Robertson, D. E. (2014). The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment. *Journal of Hydrology, 519*(0), 2763-2774. doi: http://dx.doi.org/10.1016/j.jhydrol.2014.07.041

Alvarez-Garreton, C., Ryu, D., Western, A. W., Su, C. H., Crow, W. T., Robertson, D. E., & Leahy, C. (2015). Improving operational flood ensemble prediction by the assimilation of

satellite soil moisture: comparison between lumped and semi-distributed schemes. *Hydrol. Earth Syst. Sci., 19*(4), 1659-1676. doi: http://dx.doi.org/10.5194/hess-19-1659-2015

Andrea, L., Ashish, S., Venkat, L., & Pierluigi, C. (2016). A global assessment of the timing of extreme rainfall from TRMM and GPM for improving hydrologic design. *Environmental Research Letters, 11*(5), 054003.

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., . . . Vitart, F. (2015). ERA-Interim/Land: a global land surface reanalysis data set. *Hydrol. Earth Syst. Sci., 19*(1), 389-407. doi: http://dx.doi.org/10.5194/hess-19-389-2015

Bartholmes, J., Thielen, J., Ramos, M., & Gentilini, S. (2009). The european flood alert system EFAS–Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts. *Hydrology and Earth System Sciences, 13*(2), 141-153.

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Operational Research Quarterly, 20*(4), 451-468. doi: http://dx.doi.org/10.1057/jors.1969.103

Beck, H. E., Jeu, R. A. M. d., Schellekens, J., Dijk, A. I. J. M. v., & 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. doi: http://dx.doi.org/10.1109/JSTARS.2009.2031227

Becker, F., & Choudhury, B. J. (1988). Relative sensitivity of normalized difference vegetation index (NDVI) and microwave polarization difference index (MPDI) for vegetation and desertification monitoring. *Remote Sensing of Environment, 24*(2), 297-311. doi: https://doi.org/10.1016/0034-4257(88)90031-4

Beesley, C., Frost, A., & Zajaczkowski, J. (2009). *A comparison of the BAWAP and SILO spatially interpolated daily rainfall datasets.* Paper presented at the 18th World IMACS/MODSIM Congress, Cairns, Australia.

Bogena, H. R., Huisman, J. A., Baatz, R., Hendricks Franssen, H. J., & Vereecken, H. (2013). Accuracy of the cosmic-ray soil water content probe in humid forest ecosystems: The worst case scenario. *Water Resources Research, 49*(9), 5778-5791. doi: http://dx.doi.org/10.1002/wrcr.20463

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., . . . Llorens, P. (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: http://dx.doi.org/10.1016/j.rse.2011.08.003

Brocca, L., Melone, F., & Moramarco, T. (2008). On the estimation of antecedent wetness conditions in rainfall–runoff modelling. *Hydrological Processes, 22*(5), 629-642. doi: http://dx.doi.org/10.1002/hyp.6629

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., & Hasenauer, S. (2010). Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrology and Earth System Sciences, 14*(10), 1881-1893. doi: http://dx.doi.org/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 Modeling. *IEEE transactions on Geoscience and remote sensing, 50*(7), 2542-2555. doi: http://dx.doi.org/10.1109/TGRS.2011.2177468

Carsell, K., Pingel, N., & Ford, D. (2004). Quantifying the Benefit of a Flood Warning System. *Natural Hazards Review, 5*(3), 131-140. doi: http://dx.doi.org/10.1061/(ASCE)1527-6988(2004)5:3(131)#sthash.AHBFrTUr.dpuf

Castillo, V., Gomez-Plaza, A., & Martınez-Mena, M. (2003). The role of antecedent soil water content in the runoff response of semiarid catchments: a simulation approach. *Journal of Hydrology, 284*(1), 114-130. doi: http://dx.doi.org/10.1016/S0022-1694(03)00264-6

Chen, F., Crow, W. T., Starks, P. J., & Moriasi, D. N. (2011). Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture. *Advances in Water Resources, 34*(4), 526-536. doi: http://dx.doi.org/10.1016/j.advwatres.2011.01.011

Choudhury, B. J., Schmugge, T. J., Chang, A., & Newton, R. W. (1979). Effect of surface roughness on the microwave emission from soils. *Journal of Geophysical Research: Oceans, 84*(C9), 5699-5706. doi: http://dx.doi.org/10.1029/JC084iC09p05699

Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). *Applied hydrology*.

Chowdhury, S., & Sharma, A. (2009). Multisite seasonal forecast of arid river flows using a dynamic model combination approach. *Water Resources Research, 45*(10), W10428. doi: http://dx.doi.org/10.1029/2008WR007510

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*(4), 559-583. doi: http://dx.doi.org/10.1016/0169-2070(89)90012-5

Cronshey, R. (1986). Urban hydrology for small watersheds: US Dept. of Agriculture, Soil Conservation Service, Engineering Division.

Crosson, W. L., Al-Hamdan, M. Z., Hemmings, S. N. J., & Wade, G. M. (2012). A daily merged MODIS Aqua–Terra land surface temperature data set for the conterminous United

States. *Remote Sensing of Environment, 119*, 315-324. doi: http://dx.doi.org/10.1016/j.rse.2011.12.019

Crow, W. T. (2007). A Novel Method for Quantifying Value in Spaceborne Soil Moisture Retrievals. *Journal of Hydrometeorology, 8*(1), 56-67. doi: 10.1175/JHM553.1

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., . . . Walker, J. P. (2012). Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Reviews of Geophysics, 50*(2), RG2002. doi: http://dx.doi.org/10.1029/2011RG000372

Crow, W. T., Miralles, D. G., & Cosh, M. H. (2010). A Quasi-Global Evaluation System for Satellite-Based Surface Soil Moisture Retrievals. *Geoscience and Remote Sensing, IEEE Transactions on, 48*(6), 2516-2527. doi: http://dx.doi.org/10.1109/TGRS.2010.2040481

Crow, W. T., & Ryu, D. (2009). A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals. *Hydrology and Earth System Sciences, 13*(1), 1-16. doi: http://dx.doi.org/10.5194/hess-13-1-2009

D'Odorico, P., Caylor, K., Okin, G. S., & Scanlon, T. M. (2007). On soil moisture–vegetation feedbacks and their possible effects on the dynamics of dryland ecosystems. *Journal of Geophysical Research: Biogeosciences, 112*(G4). doi: http://dx.doi.org/10.1029/2006JG000379

De Jeu, R. A. M., & Owe, M. (2003). Further validation of a new methodology for surface moisture and vegetation optical depth retrieval. *International Journal of Remote Sensing, 24*(22), 4559-4578. doi: http://dx.doi.org/10.1080/0143116031000095934

De 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. doi: http://dx.doi.org/10.1007/s10712-008-9044-0

De Lannoy, G. J. M., & Reichle, R. H. (2015). Global Assimilation of Multi-Angle and Multi-Polarization SMOS Brightness Temperature Observations into the GEOS-5 Catchment Land Surface Model for Soil Moisture Estimation. *Journal of Hydrometeorology*. doi: http://dx.doi.org/10.1175/JHM-D-15-0037.1

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., . . . Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society, 137*(656), 553-597. doi: http://dx.doi.org/10.1002/qj.828

Delworth, T. L., & Zeng, F. (2014). Regional rainfall decline in Australia attributed to anthropogenic greenhouse gases and ozone levels. *Nature Geoscience, 7*(8), 583-587. doi: http://dx.doi.org/10.1038/ngeo2201

Diamond, H. J., Karl, T. R., Palecki, M. A., Baker, C. B., Bell, J. E., Leeper, R. D., . . . Thorne, P. W. (2013). U.S. Climate Reference Network after One Decade of Operations: Status and Assessment. *Bulletin of the American Meteorological Society, 94*(4), 485-498. doi: http://dx.doi.org/10.1175/BAMS-D-12-00170.1

Didan, K., Barreto-Munoz, A., Solano, R., & Huete, A. (2015). MODIS vegetation index user's guide (MOD13 series) *Version 3.00, June 2015 (Collection 6)* (pp. 1-32): Vegetation Index and Phenology Lab, The University of Arizona.

Dobson, M. C., Ulaby, F. T., Hallikainen, M. T., & El-Rayes, M. A. (1985). Microwave Dielectric Behavior of Wet Soil-Part II: Dielectric Mixing Models. *IEEE transactions on Geoscience and remote sensing, GE-23*(1), 12.

Dong, J., Ni-Meister, W., & Houser, P. R. (2007). Impacts of vegetation and cold season processes on soil moisture and climate relationships over Eurasia. *Journal of Geophysical Research: Atmospheres, 112*(D9). doi: http://dx.doi.org/10.1029/2006JD007774

Dorigo, W. A., Gruber, A., De Jeu, R. A. M., Wagner, W., Stacke, T., Loew, A., . . . Kidd, R. (2014). Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sensing of Environment*(0). doi: http://dx.doi.org/10.1016/j.rse.2014.07.023

Dorigo, W. A., 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. *Hydrol. Earth Syst. Sci., 14*(12), 2605-2616. doi: http://dx.doi.org/10.5194/hess-14-2605-2010

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., . . . Jackson, T. (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: http://dx.doi.org/10.5194/hess-15-1675-2011

Dorigo, W. A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A. D., . . . Drusch, M. (2013). Global automated quality control of in situ soil moisture data from the International Soil Moisture Network. *Vadose Zone Journal, 12*(3). doi: http://dx.doi.org/10.2136/vzj2012.0097

Draper, C. S., Reichle, R. H., M. De Lannoy, G. J., & Liu, Q. (2012). Assimilation of passive and active microwave soil moisture retrievals. *Geophysical Research Letters, 39*(4). doi: http://dx.doi.org/10.1029/2011GL050655

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. doi: http://dx.doi.org/10.1016/j.rse.2008.11.011

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: Atmospheres, 112*(D3), D03102. doi: 10.1029/2006JD007478

Du, J. (2012). A method to improve satellite soil moisture retrievals based on Fourier analysis. *Geophysical Research Letters, 39*(15). doi: http://dx.doi.org/10.1029/2012GL052435

Duan, Z., Liu, J., Tuo, Y., Chiogna, G., & Disse, M. (2016). Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales. *Science of The Total Environment, 573*, 1536-1553. doi: http://dx.doi.org/10.1016/j.scitotenv.2016.08.213

Durán-Barroso, P., González, J., & Valdés, J. B. (2016). Improvement of the integration of Soil Moisture Accounting into the NRCS-CN model. *Journal of Hydrology, 542*, 809-819. doi: http://dx.doi.org/10.1016/j.jhydrol.2016.09.053

Dwyer, M. J., & Schmidt, G. (2006). The MODIS reprojection tool. In J. J. Qu, W. Gao, M. Kafatos, R. E. Murphy & V. V. Salomonson (Eds.), *Earth science satellite remote sensing Vol. 2: Data, Computational Processing, and Tools* (pp. 162-177): Tsinghua University Press, Beijing and Springer-Verlag GmbH Berlin Heidelberg.

Enenkel, M., Reimer, C., Dorigo, W., Wagner, W., Pfeil, I., Parinussa, R., & De Jeu, R. (2016). Combining satellite observations to develop a global soil moisture product for near-real-time applications. *Hydrol. Earth Syst. Sci., 20*(10), 4191-4208. doi: 10.5194/hess-20-4191-2016

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., . . . Johnson, J. (2010a). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE, 98*(5), 704-716. doi: https://doi.org/10.1109/JPROC.2010.2043918

Entekhabi, D., Reichle, R. H., Koster, R. D., & Crow, W. T. (2010b). Performance Metrics for Soil Moisture Retrievals and Application Requirements. *Journal of Hydrometeorology, 11*(3). doi: http://dx.doi.org/10.1175/2010JHM1223.1

Fang, B., & Lakshmi, V. (2014). Soil moisture at watershed scale: Remote sensing techniques. *Journal of Hydrology, 516*, 258-272. doi: http://dx.doi.org/10.1016/j.jhydrol.2013.12.008

Fang, B., Lakshmi, V., Bindlish, R., Jackson, T. J., Cosh, M., & Basara, J. (2013). Passive Microwave Soil Moisture Downscaling Using Vegetation Index and Skin Surface Temperature. *Vadose Zone Journal, 12*(3). doi: http://dx.doi.org/10.2136/vzj2013.05.0089

Francés, F., Vélez, J. I., & Vélez, J. J. (2007). Split-parameter structure for the automatic calibration of distributed hydrological models. *Journal of Hydrology, 332*(1–2), 226-240. doi: http://dx.doi.org/10.1016/j.jhydrol.2006.06.032

Fujii, H., Koike, T., & K., I. (2009). Improvement of the AMSR-E Algorithm for Soil Moisture Estimation by Introducing a Fractional Vegetation Coverage Dataset Derived from MODIS Data. *Journal of The Remote Sensing Society of Japan, 29*(1), 11. doi: http://doi.org/10.11440/rssj.29.282

Fujii, H., Koike, T., Ohta, T., Ishidaira, H., Jackson, T. J., & Heathman, G. (2000). *Soil moisture observation under different vegetation conditions by GBMR.* Paper presented at the Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International, Honolulu, HI.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., . . . Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data, 2*, 150066. doi: http://dx.doi.org/10.1038/sdata.2015.66

Gadani, D. H., Rana, V. A., Vyas, A. D., & Bhatnagar, S. P. (2011). Effect of saline water on emissivity of soil. *Indian Journal of Radio & Space Physics (IJRSP), 40*(4), 7.

Garcia, D. (2010). Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics & Data Analysis, 54*(4), 1167-1178. doi: http://dx.doi.org/10.1016/j.csda.2009.09.020

Gergis, J., Gallant, A. J. E., Braganza, K., Karoly, D. J., Allen, K., Cullen, L., . . . McGregor, S. (2012). On the long-term context of the 1997–2009 'Big Dry' in South-Eastern Australia: insights from a 206-year multi-proxy rainfall reconstruction. *Climatic Change, 111*(3), 923-944. doi: http://dx.doi.org/10.1007/s10584-011-0263-x

GLOBE-Task-Team, & Others. (1999). *The Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Version 1.0*. Retrieved from: http://www.ngdc.noaa.gov/mgg/topo/globe.html

Granger, C. W. J., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting, 3*(2), 197-204. doi: http://dx.doi.org/10.1002/for.3980030207

Gruber, A., Dorigo, W., Zwieback, S., Xaver, A., & Wagner, W. (2013). Characterizing coarse-scale representativeness of in situ soil moisture measurements from the International Soil Moisture Network. *Vadose Zone Journal, 12*(2).

Gruber, A., Su, C. H., Zwieback, S., Crow, W., Dorigo, W., & Wagner, W. (2016). Recent advances in (soil moisture) triple collocation analysis. *International Journal of Applied*

Earth Observation and Geoinformation, 45, Part B, 200-211. doi: http://dx.doi.org/10.1016/j.jag.2015.09.002

Gruhier, C., de Rosnay, P., Hasenauer, S., Holmes, T., de Jeu, R., Kerr, Y., . . . Zribi, M. (2010). Soil moisture active and passive microwave products: intercomparison and evaluation over a Sahelian site. *Hydrology & Earth System Sciences, 14*(1), 141-156. doi: http://dx.doi.org/10.5194/hess-14-141-2010

Hain, C. R., Crow, W. T., Mecikalski, J. R., Anderson, M. C., & Holmes, T. (2011). An intercomparison of available soil moisture estimates from thermal infrared and passive microwave remote sensing and land surface modeling. *Journal of Geophysical Research: Atmospheres, 116*(D15), D15107. doi: http://dx.doi.org/10.1029/2011JD015633

Holmes, T. R. H., De Jeu, R. A. M., Owe, M., & Dolman, A. J. (2009). Land surface temperature from Ka band (37 GHz) passive microwave observations. *Journal of Geophysical Research: Atmospheres, 114*(D4), n/a-n/a. doi: http://dx.doi.org/10.1029/2008JD010257

Homer, C. G., Dewitz, J. A., Yang, L., Jin, S., Danielson, P., Xian, G., . . . Megown, K. (2015). Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogramm. Eng. Remote Sens, 81*(5), 345-354.

Hong, Y., & Adler, R. (2008). Estimation of global SCS curve numbers using satellite remote sensing and geospatial data. *International Journal of Remote Sensing, 29*(2), 471-477. doi: http://dx.doi.org/10.1080/01431160701264292

Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., & Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation.

Huffman, G. J., & Bolvin, D. T. (2014). TRMM and other data precipitation data set documentation. from ftp://precip.gsfc.nasa.gov/pub/trmmdocs/3B42_3B43_doc.pdf

Imaoka, K., Kachi, M., Kasahara, M., Ito, N., Nakagawa, K., & Oki, T. (2010). Instrument performance and calibration of AMSR-E and AMSR2. *International Archives of the Photogrammetry, Remote Sensing and Special Information Science, 38*(Part 8).

Jackson, T. J. (1993). III. Measuring surface soil moisture using passive microwave remote sensing. *Hydrological Processes, 7*(2), 139-152. doi: http://dx.doi.org/10.1002/hyp.3360070205

Jackson, T. J., & Schmugge, T. J. (1991). Vegetation effects on the microwave emission of soils. *Remote Sensing of Environment, 36*(3), 203-212. doi: https://doi.org/10.1016/0034-4257(91)90057-D

Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT).

Ji, F., Ekström, M., Evans, J. P., & Teng, J. (2014). Evaluating rainfall patterns using physics scheme ensembles from a regional atmospheric model. *Theoretical and Applied Climatology, 115*(1), 297-304. doi: http://dx.doi.org/10.1007/s00704-013-0904-2

Jin, Y. Q. (1999). A flooding index and its regional threshold value for monitoring floods in China from SSM/I data. *International Journal of Remote Sensing, 20*(5), 1025-1030. doi: http://dx.doi.org/10.1080/014311699213064

Justice, C. O., Vermote, E., Townshend, J. R. G., Defries, R., Roy, D. P., Hall, D. K., . . . Barnsley, M. J. (1998). The Moderate Resolution Imaging Spectroradiometer (MODIS): land remote sensing for global change research. *IEEE transactions on Geoscience and remote sensing, 36*(4), 1228-1249. doi: https://doi.org/10.1109/36.701075

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J., Font, J., & Berger, M. (2001). Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE transactions on Geoscience and remote sensing, 39*(8), 1729-1735. doi: https://doi.org/10.1109/36.942551

Khan, M. Z. K., Mehrotra, R., Sharma, A., & Sankarasubramanian, A. (2014). Global Sea Surface Temperature Forecasts Using an Improved Multimodel Approach. *Journal of Climate, 27*(10), 3505-3515. doi: http://dx.doi.org/10.1175/JCLI-D-13-00486.1

Khan, S. I., Hong, Y., Vergara, H. J., Gourley, J. J., Brakenridge, G. R., Groeve, T. D., . . . Yong, B. (2012). Microwave Satellite Data for Hydrologic Modeling in Ungauged Basins. *IEEE Geoscience and Remote Sensing Letters, 9*(4), 663-667. doi: http://dx.doi.org/10.1109/LGRS.2011.2177807

Kim, D., Jung, I., & Chun, J. A. (2016a). A comparison between parameter regionalization and model calibration with flow duration curves for prediction in ungauged catchments. *Hydrol. Earth Syst. Sci. Discuss., 2016*, 1-29. doi: 10.5194/hess-2016-487

Kim, G., & Barros, A. P. (2002). Downscaling of remotely sensed soil moisture with a modified fractal interpolation method using contraction mapping and ancillary data. *Remote Sensing of Environment, 83*(3), 400-413. doi: http://dx.doi.org/10.1016/S0034-4257(02)00044-5

Kim, S., Liu, Y. Y., Johnson, F. M., Parinussa, R. M., & Sharma, A. (2015a). A global comparison of alternate AMSR2 soil moisture products: Why do they differ? *Remote Sensing of Environment, 161*(0), 43-62. doi: http://dx.doi.org/10.1016/j.rse.2015.02.002

Kim, S., Parinussa, R. M., Liu, Y. Y., Johnson, F. M., & Sharma, A. (2015b). A framework for combining multiple soil moisture retrievals based on maximizing temporal correlation. *Geophysical Research Letters, 42*(16), 6662-6670. doi: 10.1002/2015gl064981

Kim, S., Parinussa, R. M., Liu, Y. Y., Johnson, F. M., & Sharma, A. (2016b). Merging Alternate Remotely-Sensed Soil Moisture Retrievals Using a Non-Static Model Combination Approach. *Remote Sensing, 8*(6), 518. doi: ARTN 518

10.3390/rs8060518

Koike, T. (2013). Soil Moisture Algorithm Descriptions of GCOM-W1 AMSR2 (Rev. A) (pp. 8-1~8-13): Earth Observation Research Center, Japan Aerospace Exploration Agency Retrieved from http://suzaku.eorc.jaxa.jp/GCOM_W/data/doc/NDX-120015A.pdf.

Koike, T., Nakamura, Y., Kaihotsu, I., Davva, G., Matsuura, N., Tamagawa, K., & Fujii, H. (2004). Development of an advanced microwave scanning radiometer (AMSR-E) algorithm of soil moisture and vegetation water content. *Annual Journal of Hydraulic Engineering, JSCE, 48*(2), 6. doi: http://doi.org/10.2208/prohe.48.217

Koike, T., Tsukamoto, T., Kumakura, T., & Lu, M. (1996). *Spatial and Seasonal Distribution of Surface Wetness Derived from Satellite Data.* Paper presented at the Proceeding of the International Workshop on Macro-Scale Hydrological Modeling.

Komma, J., Blöschl, G., & Reszler, C. (2008). Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting. *Journal of Hydrology, 357*(3–4), 228-242. doi: http://dx.doi.org/10.1016/j.jhydrol.2008.05.020

Kornelsen, K. C., & Coulibaly, P. (2015). Reducing multiplicative bias of satellite soil moisture retrievals. *Remote Sensing of Environment, 165*(0), 109-122. doi: http://dx.doi.org/10.1016/j.rse.2015.04.031

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., . . . Lawrence, D. (2004). Regions of strong coupling between soil moisture and precipitation. *Science, 305*(5687), 1138-1140. doi: http://dx.doi.org/10.1126/science.1100217

Koster, R. D., & Suarez, M. J. (2001). Soil Moisture Memory in Climate Models. *Journal of Hydrometeorology, 2*(6), 558-570. doi: http://dx.doi.org/10.1175/1525-7541(2001)002<0558:SMMICM>2.0.CO;2

Kugler, Z., & De Groeve, T. (2007). The Global Flood Detection System *JRC Scientific and Technical Reports*. Luxembourg: Office for Official Publications of the European Communities: European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen.

Lacava, T., Cuomo, V., Di Leo, E. V., Pergola, N., Romano, F., & Tramutoli, V. (2005). Improving soil wetness variations monitoring from passive microwave satellite data: The case of April 2000 Hungary flood. *Remote Sensing of Environment, 96*(2), 135-148. doi: http://dx.doi.org/10.1016/j.rse.2005.01.015

Lagadec, L.-R., Patrice, P., Braud, I., Chazelle, B., Moulin, L., Dehotin, J., . . . Breil, P. (2016). Description and evaluation of a surface runoff susceptibility mapping method. *Journal of Hydrology, 541, Part A*, 495-509. doi: http://dx.doi.org/10.1016/j.jhydrol.2016.05.049

Leblanc, M., Tweed, S., Van Dijk, A., & Timbal, B. (2012). A review of historic and future hydrological changes in the Murray-Darling Basin. *Global and Planetary Change, 80–81*, 226-246. doi: http://dx.doi.org/10.1016/j.gloplacha.2011.10.012

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. *Water Resources Research, 35*(1), 233-241. doi: 10.1029/1998WR900018

Liu, G. (1998). A fast and accurate model for microwave radiance calculations. *Journal of the Meteorological Society of Japan, 76*(2), 335-343. doi: http://doi.org/10.2151/jmsj1965.76.2_335

Liu, Y. Y., de Jeu, R. A., McCabe, M. F., Evans, J. P., & van Dijk, A. I. (2011a). Global long-term passive microwave satellite-based retrievals of vegetation optical depth. *Geophysical Research Letters, 38*(18). doi: http://dx.doi.org/10.1029/2011GL048684

Liu, Y. Y., Dorigo, W. A., Parinussa, R. M., de Jeu, R. A. M., Wagner, W., McCabe, M. F., . . . van Dijk, A. I. J. M. (2012). Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote Sensing of Environment, 123*(0), 280-297. doi: http://dx.doi.org/10.1016/j.rse.2012.03.014

Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A. M., Wagner, W., van Dijk, A. I. J. M., . . . Evans, J. P. (2011b). Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrology and Earth System Sciences, 15*(2), 425-436. doi: http://dx.doi.org/10.5194/hess-15-425-2011

Liu, Y. Y., van Dijk, A. I. J. M., McCabe, M. F., Evans, J. P., & de Jeu, R. A. M. (2013). Global vegetation biomass change (1988–2008) and attribution to environmental and human drivers. *Global Ecology and Biogeography, 22*(6), 692-705. doi: http://dx.doi.org/10.1111/geb.12024

Loew, A., & Mauser, W. (2008). On the disaggregation of passive microwave soil moisture data using A Priori knowledge of temporally persistent soil moisture fields. *IEEE*

Transactions on Geoscience and Remote Sensing, 46(3), 819-834. doi: 10.1109/TGRS.2007.914800

Massari, C., Brocca, L., Moramarco, T., Tramblay, Y., & Didon Lescot, J.-F. (2014). Potential of soil moisture observations in flood modelling: Estimating initial conditions and correcting rainfall. *Advances in Water Resources, 74*, 44-53. doi: http://dx.doi.org/10.1016/j.advwatres.2014.08.004

Meesters, A. G. C. A., De Jeu, R. A. M., & Owe, M. (2005). Analytical derivation of the vegetation optical depth from the microwave polarization difference index. *IEEE Geoscience and Remote Sensing Letters, 2*(2), 121-123. doi: https://doi.org/10.1109/LGRS.2005.843983

Merlin, O., Rudiger, C., Al Bitar, A., Richaume, P., Walker, J. P., & Kerr, Y. H. (2012). Disaggregation of SMOS Soil Moisture in Southeastern Australia. *Geoscience and Remote Sensing, IEEE Transactions on, 50*(5), 1556-1571. doi: http://dx.doi.org/10.1109/TGRS.2011.2175000

Merlin, O., Walker, J. P., Chehbouni, A., & Kerr, Y. (2008). Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency. *Remote Sensing of Environment, 112*(10), 3935-3946. doi: http://dx.doi.org/10.1016/j.rse.2008.06.012

Michel, C., Andréassian, V., & Perrin, C. (2005). Soil Conservation Service Curve Number method: How to mend a wrong soil moisture accounting procedure? *Water Resources Research, 41*(2), W02011. doi: http://dx.doi.org/10.1029/2004WR003191

Mishra, S. K., Jain, M. K., Pandey, R. P., & Singh, V. P. (2005). Catchment area-based evaluation of the AMC-dependent SCS-CN-based rainfall–runoff models. *Hydrological Processes, 19*(14), 2701-2718. doi: http://dx.doi.org/10.1002/hyp.5736

Mo, T., Choudhury, B. J., Schmugge, T. J., Wang, J. R., & Jackson, T. J. (1982). A model for microwave emission from vegetation-covered fields. *Journal of Geophysical Research: Oceans, 87*(C13), 11229-11237. doi: http://dx.doi.org/10.1029/JC087iC13p11229

Mohanty, B. P., Cosh, M. H., Lakshmi, V., & Montzka, C. (2017). Soil moisture remote sensing: State-of-the-science. *Vadose Zone Journal, 16*(1).

Nachtergaele, F., & Batjes, N. (2012). *Harmonized world soil database*: FAO.

Narayan, U., Lakshmi, V., & Njoku, E. G. (2004, 20-24 Sept. 2004). *A simple algorithm for spatial disaggregation of radiometer derived soil moisture using higher resolution radar observations.* Paper presented at the IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium.

NASA-LP-DAAC. (2012). *Land Cover Type Yearly L3 Global 0.05Deg CMG*.

Neelam, M., & Mohanty, B. P. (2015). Global sensitivity analysis of the radiative transfer model. *Water Resources Research, 51*(4), 2428-2443.

Njoku, E. G., Ashcroft, P., Chan, T. K., & Li, L. (2005). Global survey and statistics of radio-frequency interference in AMSR-E land observations. *IEEE transactions on Geoscience and remote sensing, 43*(5), 938-947. doi: https://doi.org/10.1109/TGRS.2004.837507

Njoku, E. G., & Entekhabi, D. (1996). Passive microwave remote sensing of soil moisture. *Journal of Hydrology, 184*(1-2), 101-129. doi: http://dx.doi.org/10.1016/0022-1694(95)02970-2

Njoku, E. G., Jackson, T. J., Lakshmi, V., Chan, T. K., & Nghiem, S. V. (2003). Soil moisture retrieval from AMSR-E. *Geoscience and Remote Sensing, IEEE Transactions on, 41*(2), 215-229. doi: http://dx.doi.org/10.1109/TGRS.2002.808243

Njoku, E. G., & Kong, J.-A. (1977). Theory for passive microwave remote sensing of near-surface soil moisture. *Journal of Geophysical Research, 82*(20), 3108-3118. doi: http://dx.doi.org/10.1029/JB082i020p03108

Njoku, E. G., & Li, L. (1999). Retrieval of land surface parameters using passive microwave measurements at 6-18 GHz. *IEEE transactions on Geoscience and remote sensing, 37*(1), 79-93. doi: https://doi.org/10.1109/36.739125

NRCS. (2004). Chapter 10 Estimation of Direct Runoff from Storm Rainfall *Part 630 Hydrology National Engineering Handbook*: Natural Resources Conservation Service, United States Department of Agriculture.

NRCS. (2010). Chapter 15 Time of Concentration *Part 630 Hydrology National Engineering Handbook*: Natural Resources Conservation Service, United States Department of Agriculture.

O'Callaghan, J. F., & Mark, D. M. (1984). The extraction of drainage networks from digital elevation data. *Computer vision, graphics, and image processing, 28*(3), 323-344.

Owe, M., De Jeu, R. A. M., & Holmes, T. (2008). Multisensor historical climatology of satellite-derived global land surface moisture. *Journal of Geophysical Research: Earth Surface, 113*(F1), F01002. doi: http://dx.doi.org/10.1029/2007JF000769

Owe, M., De Jeu, R. A. M., & Walker, J. (2001). A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *Geoscience and Remote Sensing, IEEE Transactions on, 39*(8), 1643-1654. doi: http://dx.doi.org/10.1109/36.942542

Paloscia, S., Macelloni, G., & Santi, E. (2006). Soil Moisture Estimates From AMSR-E Brightness Temperatures by Using a Dual-Frequency Algorithm. *Geoscience and Remote Sensing,*

IEEE Transactions on, 44(11), 3135-3144. doi: http://dx.doi.org/10.1109/TGRS.2006.881714

Paloscia, S., & Pampaloni, P. (1988). Microwave polarization index for monitoring vegetation growth. *Geoscience and Remote Sensing, IEEE Transactions on, 26*(5), 617-621. doi: http://dx.doi.org/10.1109/36.7687

Parinussa, R. M., Holmes, T. R. H., & Crow, W. T. (2011a). The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations. *Hydrology & Earth System Sciences Discussions, 8*(4). doi: http://dx.doi.org/10.5194/hess-15-3135-2011

Parinussa, R. M., Holmes, T. R. H., Wanders, N., Dorigo, W. A., & de Jeu, R. A. M. (2014). A Preliminary Study toward Consistent Soil Moisture from AMSR2. *Journal of Hydrometeorology, 16*(2), 932-947. doi: http://dx.doi.org/10.1175/JHM-D-13-0200.1

Parinussa, R. M., Holmes, T. R. H., Yilmaz, M. T., & Crow, W. T. (2011b). The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations. *Hydrol. Earth Syst. Sci., 15*(10), 3135-3151. doi: 10.5194/hess-15-3135-2011

Parinussa, R. M., Lakshmi, V., Johnson, F., & Sharma, A. (2016). Comparing and combining remotely sensed land surface temperature products for improved hydrological applications. *Remote Sensing, 8*(2), 162. doi: http://dx.doi.org/10.3390/rs8020162

Parinussa, R. M., Meesters, A. G. C. A., Liu, Y., Dorigo, W., Wagner, W., & De Jeu, R. A. M. (2011c). Error estimates for near-real-time satellite soil moisture as derived from the land parameter retrieval model. *Geoscience and Remote Sensing Letters, IEEE, 8*(4), 779-783. doi: https://doi.org/10.1109/LGRS.2011.2114872

Pathiraja, S., Westra, S., & Sharma, A. (2012). Why continuous simulation? The role of antecedent moisture in design flood estimation. *Water Resources Research, 48*(6), W06534. doi: http://dx.doi.org/10.1029/2011WR010997

Pellenq, J., Kalma, J., Boulet, G., Saulnier, G. M., Wooldridge, S., Kerr, Y., & Chehbouni, A. (2003). A disaggregation scheme for soil moisture based on topography and soil depth. *Journal of Hydrology, 276*(1–4), 112-127. doi: http://dx.doi.org/10.1016/S0022-1694(03)00066-0

Peng, J., Niesel, J., & Loew, A. (2015). Evaluation of soil moisture downscaling using a simple thermal-based proxy-the REMEDHUS network (Spain) example. *Hydrology and Earth System Sciences, 19*(12), 4765. doi: http://dx.doi.org/10.5194/hess-19-4765-2015

Petropoulos, G. P., Ireland, G., & Barrett, B. (2015). Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Physics and Chemistry of the Earth, Parts A/B/C, 83-84*(0), 36-56. doi: 10.1016/j.pce.2015.02.009

Piles, M., Camps, A., Vall-Llossera, M., Corbella, I., Panciera, R., Rudiger, C., . . . Walker, J. (2011). Downscaling SMOS-derived soil moisture using MODIS visible/infrared data. *IEEE transactions on Geoscience and remote sensing, 49*(9), 3156-3166. doi: http://dx.doi.org/10.1109/tgrs.2011.2120615

Piles, M., Camps, A., Vall-Llossera, M., Sánchez, N., Martínez-Fernández, J., Monerris, A., . . . Acevo, R. (2010). *Soil moisture downscaling activities at the REMEDHUS Cal/Val site and its application to SMOS.* Paper presented at the Microwave Radiometry and Remote Sensing of the Environment (MicroRad), 2010 11th Specialist Meeting on.

Rao, Y. S., & Sharma, S. (2006). *Monitoring of flood over Gujarat region using AQUA AMSR-E derived surface soil moisture*.

Raupach, M., Briggs, P., Haverd, V., King, E., Paget, M., & Trudinger, C. (2009). Australian water availability project (AWAP): CSIRO marine and atmospheric research component: final report for phase 3. *Centre for Australian weather and climate research (bureau of meteorology and CSIRO). Melbourne, Australia, 67*.

Rautiainen, K., Lemmetyinen, J., Schwank, M., Kontu, A., Ménard, C. B., Mätzler, C., . . . Pulliainen, J. (2014). Detection of soil freezing from L-band passive microwave observations. *Remote Sensing of Environment, 147*(0), 206-218. doi: http://dx.doi.org/10.1016/j.rse.2014.03.007

Reichle, R., Lannoy, G. D., Koster, R. D., Crow, W. T., & Kimball, J. S. (2016). *SMAP L4 9 km EASE-Grid Surface and Root Zone Soil Moisture Geophysical Data, Version 2 (SPL4SMGP)*.

Reichle, R. H., & Koster, R. D. (2004). Bias reduction in short records of satellite soil moisture. *Geophysical Research Letters, 31*(19), L19501. doi: http://dx.doi.org/10.1029/2004GL020938

Reichle, R. H., & Koster, R. D. (2005). Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model. *Geophysical Research Letters, 32*(2), L02404. doi: http://dx.doi.org/10.1029/2004GL021700

Reichle, R. H., Koster, R. D., De Lannoy, G. J. M., Forman, B. A., Liu, Q., Mahanama, S. P. P., & Touré, A. (2011). Assessment and Enhancement of MERRA Land Surface Hydrology Estimates. *Journal of Climate, 24*(24), 6322-6338. doi: http://dx.doi.org/10.1175/JCLI-D-10-05033.1

Reichle, R. H., Koster, R. D., Dong, J., & Berg, A. A. (2004). Global Soil Moisture from Satellite Observations, Land Surface Models, and Ground Data: Implications for Data

Assimilation. *Journal of Hydrometeorology, 5*(3), 430-442. doi: 10.1175/1525-7541(2004)005<0430:GSMFSO>2.0.CO;2

Revilla-Romero, B., Thielen, J., Salamon, P., De Groeve, T., & Brakenridge, G. R. (2014). Evaluation of the satellite-based Global Flood Detection System for measuring river discharge: influence of local factors. *Hydrol. Earth Syst. Sci., 18*(11), 4467-4484. doi: http://dx.doi.org/10.5194/hess-18-4467-2014

Rosolem, R., Shuttleworth, W. J., Zreda, M., Franz, T. E., Zeng, X., & Kurc, S. A. (2013). The Effect of Atmospheric Water Vapor on Neutron Count in the Cosmic-Ray Soil Moisture Observing System. *Journal of Hydrometeorology, 14*(5), 1659-1671. doi: http://dx.doi.org/10.1175/JHM-D-12-0120.1

Ruggenthaler, R., Schöberl, F., Markart, G., Klebinder, K., Hammerle, A., & Leitinger, G. (2015). Quantification of soil moisture effects on runoff formation at the hillslope scale. *Journal of Irrigation and Drainage Engineering, 141*(9), 05015001. doi: http://dx.doi.org/10.1061/(ASCE)IR.1943-4774.0000880

Sánchez, N., Piles, M., Scaini, A., Martínez-Fernández, J., Camps, A., & Vall-Llossera, M. (2012). *Spatial patterns of SMOS downscaled soil moisture maps over the remedhus network (Spain).* Paper presented at the Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International.

Saghafian, B., Julien, P. Y., & Rajaie, H. (2002). Runoff hydrograph simulation based on time variable isochrone technique. *Journal of Hydrology, 261*(1–4), 193-203. doi: http://dx.doi.org/10.1016/S0022-1694(02)00007-0

Sandholt, I., Rasmussen, K., & Andersen, J. (2002). A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status.

Remote Sensing of Environment, 79(2), 213-224. doi: http://dx.doi.org/10.1016/S0034-4257(01)00274-7

Schmugge, T., Neill, P. E. O., & Wang, J. R. (1986). Passive Microwave Soil Moisture Research. *IEEE transactions on Geoscience and remote sensing, GE-24*(1), 12-22. doi: http://dx.doi.org/10.1109/TGRS.1986.289584

Schwanghart, W., & Kuhn, N. J. (2010). TopoToolbox: A set of Matlab functions for topographic analysis. *Environmental Modelling & Software, 25*(6), 770-781. doi: http://dx.doi.org/10.1016/j.envsoft.2009.12.002

Schwanghart, W., & Scherler, D. (2014). Short Communication: TopoToolbox 2-MATLAB-based software for topographic analysis and modeling in Earth surface sciences. *Earth Surface Dynamics, 2*(1), 1.

Scipal, K., Dorigo, W., & de Jeu, R. (2010, 25-30 July 2010). *Triple collocation - A new tool to determine the error structure of global soil moisture products.* Paper presented at the Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International.

Scipal, K., Holmes, T., de Jeu, R., Naeimi, V., & Wagner, W. (2008). A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophysical Research Letters, 35*(24), L24403. doi: 10.1029/2008GL035599

Sharma, A., & Mehrotra, R. (2014). An information theoretic alternative to model a natural system using observational information alone. *Water Resources Research, 50*(1), 650-660. doi: http://dx.doi.org/10.1002/2013WR013845

Shin, Y., & Mohanty, B. P. (2013). Development of a deterministic downscaling algorithm for remote sensing soil moisture footprint using soil and vegetation classifications. *Water Resources Research, 49*(10), 6208-6228.

Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research: Oceans (1978–2012), 103*(C4), 7755-7766.

Su, C.-H., Ryu, D., Western, A. W., & Wagner, W. (2013a). De-noising of passive and active microwave satellite soil moisture time series. *Geophysical Research Letters, 40*(14), 3624-3630. doi: http://dx.doi.org/10.1002/grl.50695

Su, C.-H., Ryu, D., Young, R. I., Western, A. W., & Wagner, W. (2013b). Inter-comparison of microwave satellite soil moisture retrievals over the Murrumbidgee Basin, southeast Australia. *Remote Sensing of Environment, 134*(0), 1-11. doi: http://dx.doi.org/10.1016/j.rse.2013.02.016

Syvitski, J. P., & Brakenridge, G. R. (2013). Causation and avoidance of catastrophic flooding along the Indus River, Pakistan. *GSA Today, 23*(1), 4-10.

Terui, N., & van Dijk, H. K. (2002). Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting, 18*(3), 421-438. doi: http://dx.doi.org/10.1016/S0169-2070(01)00120-0

Timmermann, A. (2006). Forecast combinations. *Handbook of economic forecasting, 1*, 135-196.

Tomer, S., Al Bitar, A., Sekhar, M., Zribi, M., Bandyopadhyay, S., & Kerr, Y. (2016). MAPSM: A Spatio-Temporal Algorithm for Merging Soil Moisture from Active and Passive Microwave Remote Sensing. *Remote Sensing, 8*(12), 990. doi: http://dx.doi.org/10.3390/rs8120990

Turner, M., Bari, M., Amirthanathan, G., & Ahmad, Z. (2012). *Australian network of hydrologic reference stations-advances in design, development and implementation.* Paper presented at the Hydrology and Water Resources Symposium 2012.

Ulaby, F. T., Moore, R. K., & Fung, A. K. (1982). Microwave remote sensing: Active and passive. Volume 2-Radar remote sensing and surface scattering and emission theory.

Vermote, E. F., El Saleous, N. Z., & Justice, C. O. (2002). Atmospheric correction of MODIS data in the visible to middle infrared: first results. *Remote Sensing of Environment, 83*(1), 97-111. doi: http://dx.doi.org/10.1016/S0034-4257(02)00089-5

Wagner, W., Dorigo, W., de Jeu, R., Fernandez, D., Benveniste, J., Haas, E., & Ertl, M. (2012). *Fusion of active and passive microwave observations to create an essential climate variable data record on soil moisture.* Paper presented at the XXII ISPRS Congress, Melbourne, Australia.

Wagner, W., Scipal, K., Pathe, C., Gerten, D., Lucht, W., & Rudolf, B. (2003). Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data. *Journal of Geophysical Research: Atmospheres, 108*(D19), 4611. doi: http://dx.doi.org/10.1029/2003JD003663

Walker, J., Panciera, R., & Monerris, A. (2013). A National Soil Moisture Monitoring Capability, Annex 12. Basis of an Australian Radar Soil Moisture Algorithm Theoretical Baseline Document. Sydney, Australia: Australian Centre for Space Engineering Research (ACSER), Faculty of Engineering, University of New South Wales.

Walker, J. P., & Houser, P. R. (2004). Requirements of a global near-surface soil moisture satellite mission: accuracy, repeat time, and spatial resolution. *Advances in Water Resources, 27*(8), 785-801.

Wan, Z., Wang, P., & Li, X. (2004). Using MODIS Land Surface Temperature and Normalized Difference Vegetation Index products for monitoring drought in the southern Great Plains, USA. *International Journal of Remote Sensing, 25*(1), 61-72. doi: 10.1080/0143116031000115328

Wanders, N., Karssenberg, D., de Roo, A., de Jong, S. M., & Bierkens, M. F. P. (2014). The suitability of remotely sensed soil moisture for improving operational flood forecasting. *Hydrol. Earth Syst. Sci., 18*(6), 2343-2357. doi: http://dx.doi.org/10.5194/hess-18-2343-2014

Wang, G., Garcia, D., Liu, Y., De Jeu, R., & Dolman, A. J. (2012). A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations. *Environmental Modelling & Software, 30*, 139-142. doi: http://dx.doi.org/10.1016/j.envsoft.2011.10.015

Wang, J. R., & Choudhury, B. J. (1981). Remote sensing of soil moisture content, over bare field at 1.4 GHz frequency. *Journal of Geophysical Research: Oceans, 86*(C6), 5277-5282. doi: 10.1029/JC086iC06p05277

Wang, J. R., & Schmugge, T. J. (1980). An Empirical Model for the Complex Dielectric Permittivity of Soils as a Function of Water Content. *Geoscience and Remote Sensing, IEEE Transactions on, GE-18*(4), 288-295. doi: http://dx.doi.org/10.1109/TGRS.1980.350304

Wang, L., & Liu, H. (2006). An efficient method for identifying and filling surface depressions in digital elevation models for hydrologic analysis and modelling. *International Journal of Geographical Information Science, 20*(2), 193-213. doi: http://dx.doi.org/10.1080/13658810500433453

Wasko, C., Sharma, A., & Rasmussen, P. (2013). Improved spatial prediction: A combinatorial approach. *Water Resources Research, 49*(7), 3927-3935. doi: http://dx.doi.org/10.1002/wrcr.20290

Weiss, D. J., Atkinson, P. M., Bhatt, S., Mappin, B., Hay, S. I., & Gething, P. W. (2014). An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS*

Journal of Photogrammetry and Remote Sensing, 98, 106-118. doi: http://dx.doi.org/10.1016/j.isprsjprs.2014.10.001

Wieder, W. R., Boehnert, J., Bonan, G. B., & Langseth, M. (2014). *Regridded Harmonized World Soil Database v1.2. Data set*. Retrieved from: http://daac.ornl.gov

Wigneron, J., Schmugge, T. J., Chanzy, A., Calvet, J., & Kerr, Y. (1998). Use of passive microwave remote sensing to monitor soil moisture. *Agronomie, 18*(1), 27-43.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. 100): Academic press.

Yee, M., Walker, J. P., Dumedah, G., Monerris, A., & Rüdiger, C. (2013). *Towards Land Surface Model Validation from Using Satellite Retrieved Soil Moisture*. Paper presented at the 20th International Congress on Modelling and Simulation, Adelaide, Australia.

Yee, M. S., Walker, J. P., Monerris, A., Rüdiger, C., & Jackson, T. J. (2016). On the identification of representative in situ soil moisture monitoring stations for the validation of SMAP soil moisture products in Australia. *Journal of Hydrology, 537*, 367-381. doi: http://dx.doi.org/10.1016/j.jhydrol.2016.03.060

Yilmaz, M. T., & Crow, W. T. (2013). The Optimality of Potential Rescaling Approaches in Land Data Assimilation. *Journal of Hydrometeorology, 14*(2). doi: http://dx.doi.org/10.1175/JHM-D-12-052.1

Zreda, M., Shuttleworth, W., Zeng, X., Zweck, C., Desilets, D., Franz, T., & Rosolem, R. (2012). COSMOS: the COsmic-ray Soil Moisture Observing System. *Hydrology & Earth System Sciences, 16*(11). doi: http://dx.doi.org/10.5194/hess-16-4079-2012