

UNIVERSITÀ DELLA CALABRIA



UNIVERSITA' DELLA CALABRIA

Dipartimento di Ingegneria Informatica, Modellistica, Elettronica e Sistemistica

**Dottorato di Ricerca in
Ingegneria Civile e Industriale**

CICLO

XXXI

**ASSIMILATION OF SATELLITE SOIL MOISTURE IN HYDROLOGICAL MODELING:
ASSESSMENT OF OBSERVATIONS PREPROCESSING
AND ERROR CHARACTERIZATION METHODS**

Settore Scientifico Disciplinare ICAR/02

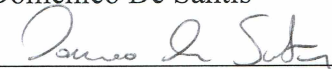
Coordinatore: Ch.mo Prof. Franco Furgiuele

Firma 

Supervisore/Tutor: Ch.mo Prof. Daniela Biondi

Firma 

Dottorando: Dott. Domenico De Santis

Firma 

ABSTRACT

Il contenuto d'acqua nel suolo svolge un ruolo fondamentale all'interno di numerosi processi che avvengono sulla superficie terrestre, ed in particolare di quelli che fanno parte del ciclo idrologico. In tal senso il contenuto d'acqua nel suolo rappresenta una variabile chiave anche nell'ambito della generazione dei deflussi nei bacini idrografici per effetto degli eventi pluviometrici, e la corretta caratterizzazione della sua evoluzione temporale risulta estremamente funzionale ad una efficace previsione degli eventi di piena. Dato il ruolo di estremo interesse occupato nell'evoluzione dei processi non solo idrologici ma anche ad esempio climatici e agricoli, crescente attenzione è stata dedicata alla modellazione del contenuto d'acqua nel suolo ai diversi fini applicativi, nonché al monitoraggio strumentale della grandezza, che avviene sia in situ, a scala sostanzialmente puntuale con sensori caratterizzati da elevata accuratezza e risoluzione temporale, che da remoto. Con riferimento al secondo caso, il monitoraggio da satellite ha avuto notevoli sviluppi negli ultimi anni, arrivando a fornire informazioni su scala globale che si distinguono per risoluzioni spaziali e temporali sempre più spinte, anche se riferite ai soli primi centimetri di suolo. Queste tre opzioni per la descrizione dell'andamento del contenuto d'acqua nel suolo devono essere viste come complementari, in virtù delle loro diverse peculiarità, nonché delle limitazioni e degli errori che le caratterizzano. In tal senso, un'interessante opportunità è costituita dalle tecniche di data assimilation sviluppate per integrare in maniera ottimale, sulla base delle relative incertezze, le osservazioni con le previsioni da modello.

Una potenziale applicazione è l'assimilazione delle osservazioni da satellite all'interno dei modelli afflussi-deflussi, al fine di migliorare le stime delle variabili di stato che rappresentano il contenuto d'acqua nel suolo, e da queste la simulazione delle portate fluviali. Numerosi studi sono stati svolti sul tema, con risultati spesso contrastanti, evidenziando un grande potenziale per questo genere di applicazione, ma anche la necessità di approfondire le numerose scelte procedurali tipicamente richieste in un lavoro di data assimilation. Le tecniche di data assimilation comunemente usate forniscono soluzioni ottime per problemi con precise ipotesi di base (ad esempio l'assenza di errori sistematici), attraverso il confronto fra osservazioni e stime da modello (che devono essere eventualmente 'mappate' qualora rappresentino grandezze diverse, ad esempio contenuto d'acqua riferito a diversi volumi/spessori di suolo) basato

sulle relative varianze d'errore. Numerose soluzioni sono state proposte per affrontare i vari steps richiesti dal data assimilation, che si sono dimostrati avere un ruolo decisivo sui risultati finali. Le soluzioni proposte riguardano tanto i modelli, ad esempio attraverso una migliorata rappresentazione delle incertezze di stima o con modifiche alla struttura che siano funzionali all'assimilazione delle osservazioni satellitari, che le osservazioni.

Le operazioni condotte sulle osservazioni ai fini della successiva integrazione in modelli previsionali hanno costituito il tema principale di questo lavoro. Generalmente, nelle fasi che precedono l'assimilazione delle misure da satellite di contenuto d'acqua nel suolo sono analizzate le seguenti questioni: la verifica della qualità delle osservazioni satellitari, la differenza fra gli spessori di terreno indagato dal sensore e riprodotto nel modello, la correzione delle differenze sistematiche fra i dataset di osservazioni e simulazioni da modello, la caratterizzazione delle varianze degli errori random.

Procedure di quality check sono messe a punto per scartare osservazioni ritenute troppo poco attendibili; in tal senso sono fondamentali gli indicatori inclusi nei dataset satellitari, che, fornendo ad esempio informazioni sulle condizioni ambientali durante la misura o feedback dall'algorithm di stima, consentono una caratterizzazione della qualità del dato. Il setup delle procedure di quality check è funzione ovviamente dell'applicazione finale, tenendo conto degli effetti derivanti tanto dall'utilizzo di un dato poco accurato che dalla sua eliminazione.

Un altro aspetto di cui tenere conto riguarda la profondità di suolo in cui è rilevato il dato di contenuto d'acqua da satellite, limitata a pochi centimetri, laddove i volumi di controllo dei modelli sono generalmente maggiori. A tal fine, la struttura di alcuni modelli è stata modificata inserendo uno strato superficiale di spessore ridotto. Una soluzione di uso comune (talvolta anche nel caso di modelli multilayer) è la propagazione dell'informazione superficiale allo spessore di interesse attraverso un filtro esponenziale, che restituisce un indice indicato come SWI (soil water index). La semplicità di questo approccio, basato su un unico parametro, ne ha determinato un'ampia diffusione in vari ambiti applicativi, e dataset globali di contenuto d'acqua da satellite ottenuti con questo metodo sono attualmente in distribuzione.

L'eventuale presenza di differenze sistematiche fra il dato da satellite in corso di processamento e la stima da modello deve essere poi corretta, andando ad inficiare in caso contrario le prestazioni del generico sistema di data assimilation, finalizzato alla

sola riduzione degli errori random. Diversi approcci sono al riguardo disponibili; quelli di uso predominante risultano indirizzati al matching delle caratteristiche complessive dei due dataset (ad esempio in termini di varianza). Tuttavia, quando la correzione risulta preliminare al data assimilation, pare più appropriato l'uso di tecniche che cerchino di tenere conto della struttura di errore dei due dataset, in modo da effettuare il matching della sola parte informativa (anche nota come segnale), separando quindi i contributi legati all'errore.

Sull'osservazione così preprocessata si effettua, quindi, una stima della varianza degli errori random, che contribuirà a determinare il suo peso quando sarà combinata con la previsione 'a priori' del modello. Una inadeguata caratterizzazione in questa fase impedisce di giungere al valore di 'analisi' ottimale, caratterizzato cioè da varianza di errore minima, e può portare anche al peggioramento delle performance iniziali del modello. Anche per questo step sono stati suggeriti diversi approcci, fra cui quello di uso consolidato è denominato Triple Collocation (TC), e si basa sull'utilizzo di tre dataset indipendenti per i quali si assume la stazionarietà della varianza di errore. Un metodo alternativo, in grado di fornire una stima tempovariabile della grandezza qui indagata, è la propagazione analitica degli errori (EP, error propagation) associati agli input e ai parametri attraverso le equazioni del modello da cui deriva l'osservazione (la misura da satellite del contenuto d'acqua non è in alcun caso diretta ma prevede il processamento delle grandezze effettivamente misurate dai sensori di bordo). Questo secondo approccio tuttavia non garantisce stime in magnitudo plausibili come la TC, non tenendo conto del contributo degli errori dovuti alla struttura del modello.

L'analisi delle operazioni di preprocessing e caratterizzazione degli errori delle osservazioni da satellite di contenuto d'acqua nel suolo è stata principalmente svolta attraverso lo sviluppo di due applicazioni.

Nella prima applicazione sono trattati i temi del quality check delle osservazioni satellitari e, soprattutto, del trasferimento dell'osservazione superficiale di contenuto d'acqua da satellite a spessori di suolo di maggiore interesse applicativo, usando l'approccio del filtro esponenziale di largo uso in letteratura, in un contesto di verifica della capacità della stima derivata da dati satellitari di riprodurre l'andamento osservato in situ del contenuto d'acqua su strati di spessore maggiore. L'aspetto innovativo introdotto nel lavoro di tesi è costituito dalla messa a punto di uno schema di propagazione degli errori originale, finalizzato alla caratterizzazione per via analitica

dell'andamento temporale delle varianze degli errori random del SWI. Le equazioni di propagazione degli errori sono state ricavate e poste in una pratica forma ricorsiva, consentendo di tenere in conto fattori che notoriamente introducono inaccuratezze negli output del filtro esponenziale. Con l'approccio proposto diventa, infatti, possibile propagare le varianze d'errore tempovariabili disponibili in alcuni dataset satellitari di contenuto d'acqua superficiale, nonché valutare gli effetti sul SWI in termini di varianza di errore legati alla disponibilità temporale di misure in input e all'incertezza nel parametro del filtro. Una valutazione preliminare della procedura di propagazione degli errori proposta è stata effettuata verificando l'effettiva corrispondenza fra varianza d'errore del SWI stimata ed effettivi scostamenti rispetto a misure in situ di riferimento; contestualmente sono state anche testate diverse configurazioni della procedura di quality check usando gli indicatori disponibili per il prodotto satellitare usato.

Nella seconda applicazione sono trattati i temi della correzione delle differenze sistematiche fra i dataset di osservazioni e simulazioni da modello, e della caratterizzazione delle varianze degli errori random nelle osservazioni, ai fini della valutazione degli effetti dell'assimilazione di misure satellitari di contenuto d'acqua del suolo sulle performance di modelli afflussi-deflussi. Lo studio, svolto durante un soggiorno di ricerca presso il gruppo di Idrologia del CNR-IRPI di Perugia ed in particolare con i ricercatori Luca Brocca e Christian Massari, presenta diversi aspetti innovativi, il primo dei quali è costituito dall'elevato numero (diverse centinaia) di bacini di studio, distribuiti nel continente europeo e complessivamente rappresentativi di diverse condizioni climatiche e fisiografiche, laddove i lavori precedenti su queste tematiche coinvolgevano generalmente aree geografiche ridotte e/o un numero contenuto di bacini. Il dataset di partenza, inclusivo di valori di portata, precipitazione, temperatura e osservazioni satellitari di contenuto d'acqua nel suolo per quasi 900 bacini, è stato costruito dal gruppo di Idrologia del CNR-IRPI. Un secondo aspetto d'interesse riguarda l'aver considerato, oltre ad osservazioni da sensori di tipo sia attivo che passivo provenienti da diverse missioni spaziali, diverse scelte procedurali per le fasi di rimozione delle differenze sistematiche e di caratterizzazione degli errori delle osservazioni. Nel complesso, sebbene le metodologie utilizzate costituiscano delle pratiche riconosciute e usate in questi ambiti, l'utilizzo di procedure comuni per un così largo numero di bacini rappresenta un'applicazione raramente riscontrata in letteratura che ha come principale pregio quello di consentire di superare le soggettività introdotte

con la scelta di soluzioni sito-specifiche sovente fatte in precedenti studi su scala più ridotta e talvolta orientate all'ottimizzazione dei risultati finali della procedura di data assimilation. Un terzo tema analizzato, oggetto di attenzione recente nella letteratura del settore, è legato alla presenza di bias di tipo 'moltiplicativo' nelle serie temporali di contenuto d'acqua nel suolo da modello e derivate da satellite, ancora presenti in seguito alla fase di rimozione delle differenze sistematiche, e al suo effetto sugli output di portata ottenuti assimilando l'osservazione. Con riferimento all'obiettivo generale del miglioramento della previsione idrologica, in questa applicazione i benefici dell'assimilazione dei dati da satellite sono apparsi variabili, confermando in qualche modo i risultati contrastanti presenti in letteratura. Quale contributo a questo dibattito, lo studio fornisce indicazioni sulla bontà dell'assimilazione di diversi prodotti satellitari in diverse aree geografiche e sotto diverse condizioni preliminari (ad esempio differenti regimi climatici ma anche differenti accuratezze degli input pluviometrici disponibili), e sugli effetti dei diversi approcci metodologici usati per le operazioni preliminari all'assimilazione nel modello.

La tesi è strutturata come segue. Il capitolo 1 è costituito da una breve introduzione alle tematiche del lavoro, mentre il capitolo 2 ha per oggetto il contenuto d'acqua del suolo (definizioni, fattori e processi che ne determinano le dinamiche spaziali e temporali, cenni al ruolo nelle varie applicazioni incluse quelle idrologiche) e le caratteristiche dei vari approcci con cui ne viene descritta l'evoluzione (modellazione, misure in situ e da remoto). Il capitolo 3 è incentrato sul data assimilation, fornendo una panoramica dei diversi approcci, una sintesi di risultati ed evidenze relativi all'assimilazione delle misure di contenuto d'acqua nel suolo, e la formulazione matematica dei metodi più comunemente utilizzati per tale scopo. Nel capitolo 4 è fornito un inquadramento teorico su problematiche e metodologie relative alle operazioni di preprocessing e di caratterizzazione degli errori delle osservazioni. Nei capitoli 5 e 6 sono mostrate nel dettaglio le due applicazioni sopra descritte che costituiscono l'aspetto peculiare di questa tesi.

Alla fine di questa importante esperienza non sono poche le persone a cui è giusto dedicare un pensiero.

Un primo ringraziamento va alla mia supervisor, Daniela Biondi, di cui ho potuto costantemente apprezzare le qualità umane oltre che professionali, e che è stata una guida preziosa e fondamentale durante il percorso. Con lei ringrazio anche il Prof. Pasquale Versace, per il sostegno e la fiducia e le tante opportunità offertemi, e gli altri membri del suo gruppo di lavoro, con cui ho condiviso tanto negli ultimi anni.

Un grazie sentito va anche a tutti i componenti del gruppo di Idrologia del CNR-IRPI di Perugia, dove ho trascorso un troppo breve periodo, per la loro grande disponibilità e professionalità, a partire dai ricercatori Luca Brocca e Christian Massari che più hanno contribuito a questo mio lavoro, supportandomi con aiuto concreto e sempre preziosi materiali e indicazioni.

Durante questo percorso ho avuto la possibilità di conoscere e interfacciarmi con tante persone nella mia università e non solo. I colleghi del XXXI ciclo e i membri del collegio del Dottorato di Ricerca in Ingegneria Civile e Industriale, così come gli altri dottorandi e ricercatori frequentati durante le numerose esperienze esterne, hanno fornito aiuto, stimoli, e gradita compagnia, e per tutto questo sono inclusi nei miei ringraziamenti.

Last but not least, famiglia e amici che sono rimasti sempre vicini infondendo forza e coraggio, malgrado il poco tempo che sono riuscito a ritagliare per loro.

CONTENTS

1. INTRODUCTION.....	12
2. SOIL MOISTURE ESTIMATION	14
2.1 Introduction.....	14
2.2 Soil moisture definition and units.....	14
2.3 Soil moisture processes and variability	15
2.4 Soil moisture key applications.....	22
2.5 Soil moisture modelling	25
2.6 In situ soil moisture observations	31
2.7 Remote sensing soil moisture observations.....	41
2.7.1 Active microwave sensing of surface soil moisture	48
2.7.1.1 ASCAT	53
2.7.2 Passive microwave sensing of surface soil moisture.....	55
2.7.2.1 SMOS	64
2.7.2.2 SMAP	66
2.7.2.3 AMSR2.....	68
2.7.3 ESA-CCI soil moisture products	68
3. DATA ASSIMILATION METHODS	73
3.1 Introduction.....	73
3.2 An overview of data assimilation methods.....	73
3.3 Data assimilation of soil moisture	78
3.4 System formulation and analysis scheme.....	82
3.5 Sequential data assimilation.....	84
3.5.1 Linear dynamics: the Kalman Filter	84
3.5.2 Nonlinear dynamics	84
3.5.2.1 Ensemble Kalman Filter.....	86
4. OBSERVATION PREPROCESSING AND ERROR CHARACTERIZATION	88
4.1 Introduction.....	88
4.2 Overview of steps prior to data assimilation	89
4.3 Quality check	93
4.4 Surface soil moisture propagation to deeper layers.....	95
4.4.1 The exponential filter	98
4.4.2 The SMAR method	100
4.5 Bias correction methods	101
4.5.1 Linear methods.....	104
4.5.2 CDF matching	106
4.6 Error characterization methods for evaluation and data assimilation of satellite soil moisture.....	107
4.6.1 Standard statistical measurements	113
4.6.2 Triple collocation.....	115
4.6.3 Error propagation.....	118
5. ERROR PROPAGATION IN EXPONENTIAL FILTER APPROACH	120
5.1 Introduction.....	120
5.2 Factors involved in SWI errors.....	120
5.3 Soil water index error variance estimates.....	124
5.4 Case study	125

5.4.1 Study areas and in situ observations	125
5.4.2 Remote sensing observations and quality flags	128
5.4.3 Experimental setup	131
5.5 Application of EP equations neglecting the error in SWI parameter	131
5.5.1 Data masking on ASCAT product attributes	132
5.5.2 Data masking based on SSM noise and SWI noise	137
5.6 Application of EP equations considering the error in SWI parameter	142
5.7 Conclusions	148
6. EFFECTS OF DIFFERENT RESCALING AND ERROR CHARACTERIZATION	
SCHEMES IN AN EXTENSIVE DATA ASSIMILATION EXPERIMENT	150
6.1 Introduction.....	150
6.2 Study catchments and datasets	151
6.2.1 Study catchments and hydrological data	151
6.2.2 Climatic data	152
6.2.3 Remotely-sensed soil moisture data.....	153
6.3 Hydrological model and calibration results.....	154
6.4 Observation preprocessing and error characterization	157
6.5 Model error representation.....	161
6.6 EnKF implementation	163
6.7 Results and discussion	163
6.8 Conclusions	170
7. CONCLUSIONS	172
ANNEX 1	176
REFERENCES.....	193

1. INTRODUCTION

Soil moisture (SM) is a key state variable modulating energy and water exchanges at the land and atmosphere interface. Knowledge of the temporal and spatial variability of soil moisture is essential for a multitude of applications, for example those related to climate and agriculture. Soil moisture also plays an important role in the hydrological cycle, as it controls the partitioning of precipitation into runoff and infiltration, with effects in runoff modelling and flood forecasting. For this reason SM data are considered a valuable information to be integrated in flood early warning systems.

Currently, three different approaches are used for the monitoring of soil moisture from the point to the global scales: in situ observations, prediction models, and remote sensing. In particular, soil moisture data from remote sensing observations are attracting increasing interest, with global coarse-scale retrievals that are obtained from active and passive sensors on board several space missions, including dedicated ones; however, remote sensing measurements can be related to the surface soil moisture (SSM), with a contributing depth of just few centimetres. Prediction models and in situ and remote sensing observations have different characteristics and must be seen as complementary: their integrated use allows to improve our monitoring capabilities of soil moisture dynamics, as well as our knowledge about important natural processes.

In this sense, SM measurements could have a great potential in improving flood simulation through data assimilation (DA) techniques, that allow for updating the model soil moisture states in an optimal way, by considering the respective uncertainties.

In recent years, increasing availability of satellite observations with improved spatial and temporal resolutions has brought great interest into this kind of application; however, it still remains controversial if satellite SM assimilation introduces added-value in runoff modelling, and much more research is still needed. The contrasting results obtained in various studies revealed a number of scientific and practical issues that might have a significant impact on the performance of DA systems. These critical aspects are still under-research and need to be further addressed, although several approaches have been developed to overcome them.

Some of these issues concern the satellite observations, such as the low quality of soil moisture retrievals under certain surface conditions (e.g. dense vegetation, frozen soils,

open water), the mismatch between soil depth simulated in the model and observed by the remote sensor, the bias between satellite data and model states, the assessment of the magnitude and the structure of the observation errors.

In this sense, a preprocessing phase (generally including procedures for quality check, propagation of surface information to deeper layers, bias correction, and ultimately aimed to the mapping between observed and modelled variables) and error characterization of satellite-based soil moisture measurements are necessary to their subsequent assimilation in prediction models. These operations constituted the main theme of this work.

The analysis was mainly carried out through two applications. In the first application an analytical error propagation (EP) scheme was proposed for the exponential filter, that is a widespread approach to extend SSM information to deeper layers (like those in model simulations). The EP scheme is aimed to assess the time-variant error variance of the exponential filter outputs and allows to take in account some shortcomings of the method.

In the second application, a data assimilation experiment was made on a large number of catchments located across Europe, in order to assess the effects of integrating remotely sensed SM on a rainfall-runoff model performances. Both passive and active remote sensing products are used, two alternative approaches are considered for satellite data bias correction, as well as different configurations for the method (i.e. the Triple Collocation) which is selected for observation error characterization.

The two applications related to the assessment of satellite soil moisture preprocessing and error characterization practices are showed in chapters 5 and 6, respectively. Prior to the applications description, a theoretical background of the work is given in the following chapters. The main aspects related to soil moisture estimation, including the characteristics of the different monitoring approaches, are discussed in chapter 2; an overview of the data assimilation approaches is provided in chapter 3; finally, the main methods for observation preprocessing and error characterization are summarized in chapter 4.

2. SOIL MOISTURE ESTIMATION

2.1 Introduction

Accurate knowledge of soil moisture is widely recognized as essential for a multitude of applications, including hydrological studies and flood forecasting, numerical weather prediction and climate monitoring, agricultural applications, drought risk assessments, vegetative stress predictions as well as ecosystem preservation.

This chapter is about some important aspects related to the monitoring of soil moisture. Firstly, soil moisture definition and units are given, and processes and factors are showed that determine its spatial and temporal evolution at different scales. Some applications where soil moisture plays an important role are then described, with particular regard to its influence in runoff modelling.

Three different approaches are used for the monitoring of soil moisture, i.e. simulation models and in situ and remote sensing observations. These approaches have different characteristics, also in terms of spatial and temporal resolution, and are described highlighting advantages and limitations. They must be seen as complementary and their integrated use allows to improve our monitoring capabilities of soil moisture dynamics, as well as our knowledge about important natural processes.

2.2 Soil moisture definition and units

Soil moisture (SM) is the content of water in the soil, held in the spaces between soil particles, and is generally referred in terms of average conditions with regard to a given soil volume (or to the corresponding depth range). A part of the soil total volume is in fact composed by pores that can be filled with air or water. The porosity is introduced to represent the fraction of the volume of voids, V_v , over the total volume, V_t :

$$n = \frac{V_v}{V_t} \quad (2.1)$$

with n values that are typically less than 0.5.

Soil moisture is usually expressed as the volumetric fraction of water in a given soil volume (e.g. m^3 water per m^3 soil):

$$\theta_{vol} = \frac{V_w}{V_t} \quad (2.2)$$

and ranges from 0 (soil completely dry) to n (saturated soil). Soil moisture can also be expressed in terms of saturation degree:

$$\theta_{SD} = \frac{V_w}{V_v} \quad (2.3)$$

varying between 0 (no water content) and 1 (full saturation).

The porosity is needed to convert the saturation degree into volumetric soil moisture, and vice versa:

$$\theta_{vol} = n \cdot \theta_{SD} \quad (2.4)$$

Two characteristic soil moisture values are the field capacity and the wilting point, both depending on soil properties such as soil texture. Field capacity is the amount of water a soil can hold against gravity, i.e. volume of water retained after drainage due to gravity from a thoroughly saturated soil. Wilting point is the minimal soil moisture the plant requires not to wilt, i.e. below which water is held too strongly by the soil matrix that roots of plants cannot uptake it.

2.3 Soil moisture processes and variability

SM is a key state variable modulating energy and water exchanges at the land and atmosphere interface and derived from hydrological processes such as rainfall and snowmelt infiltration, evaporation and plant transpiration, movement of water in unsaturated zone (i.e. between soil surface and groundwater level) due to vertical percolation, lateral throughflow, capillary attraction from groundwater, or also groundwater movements (Western et al., 2002).

Figure 2.1a illustrates a standard one-dimensional conceptualization of the soil profile and the fluxes that influence the soil moisture. Generally, the exchanges between the atmosphere and the soil dominate water content dynamics, with the soil moisture being primarily replenished by infiltration and depleted by soil evaporation and plant transpiration. Fluxes between the soil and groundwater are also important. Drainage from the soil profile is the primary source of recharge for many groundwater systems, and capillary rise from shallow groundwater tables can be an important source of water replenishing the soil water store during drier periods. Included in Figure 2.1a is an example of soil moisture profiles measured at different times in the same ground station. Both the amount of soil moisture and its dynamics change with depth. In the upper 50 cm, soil moisture is strongly influenced by the fluxes between the active root

zone and the atmosphere; the moisture here is more variable than the moisture at depth. Surface soil moisture also responds more quickly and so has both short and long timescale variability, whereas the moisture at depth is less responsive to short term variations in the fluxes across the soil-atmosphere interface.

Figure 2.1*b* illustrates a standard conceptualization of a hillslope. The key difference with Figure 2.1*a* is in lateral flows (particularly in saturated zone) acting to redistribute soil water and also influence runoff processes. For significant lateral drainage to occur some conditions are necessary, such as topographic relief and an impeding layer limiting vertical drainage or anisotropy in hydraulic conductivity, as well as the process is more pronounced during wet conditions. In higher parts of the landscape lateral drainage can deplete the soil moisture store; this lateral drainage collects in convergent parts of the hillslope and replenishes the soil moisture store in those areas, often leading to soil saturation and generation of saturation excess runoff.

Thus, soil moisture is generally highly variable not only in time but also in space, both vertically and horizontally, and the proper SM description across spatial and temporal scales constitutes an important challenge, given a large variety of soil moisture applications requiring data from local (e.g., for agricultural support) to global scales (e.g., for climate change studies).

Considerable analysis efforts have been put on horizontal spatial SM variations, especially in the surface layer (also for what is said commenting on the Figure 2.1*a*), with several approaches proposed for problems such as the changes in variable values across scales (i.e. the upscaling and downscaling problems) or the choice of sites representative of larger area average behaviour. Regarding vertical changes within soil profiles, there is a lack of detailed analysis, with few studies that for example providing recommendations for a sampling strategy that accounts for hydrologically relevant profile sections (e.g. Pachepsky et al., 2005; Schwen et al., 2014).

Spatial variability can be random (i.e. not predictable in detail but showing predictable statistical properties) or organized (i.e. characterized by consistent spatial patterns) or a combination of the two (Western et al., 2003). The characteristics of spatial patterns change between landscapes and over time as different static and dynamic factors influence the soil moisture variations (Petropoulos et al., 2013). Previous studies highlighted that the main factors generally influencing the SM spatial variability are soil properties, topography, vegetation and climate (e.g. Mohanty & Skaggs, 2001; Wilson

et al., 2004; Crow et al., 2012), whose effects on SM determining processes are detailed below.

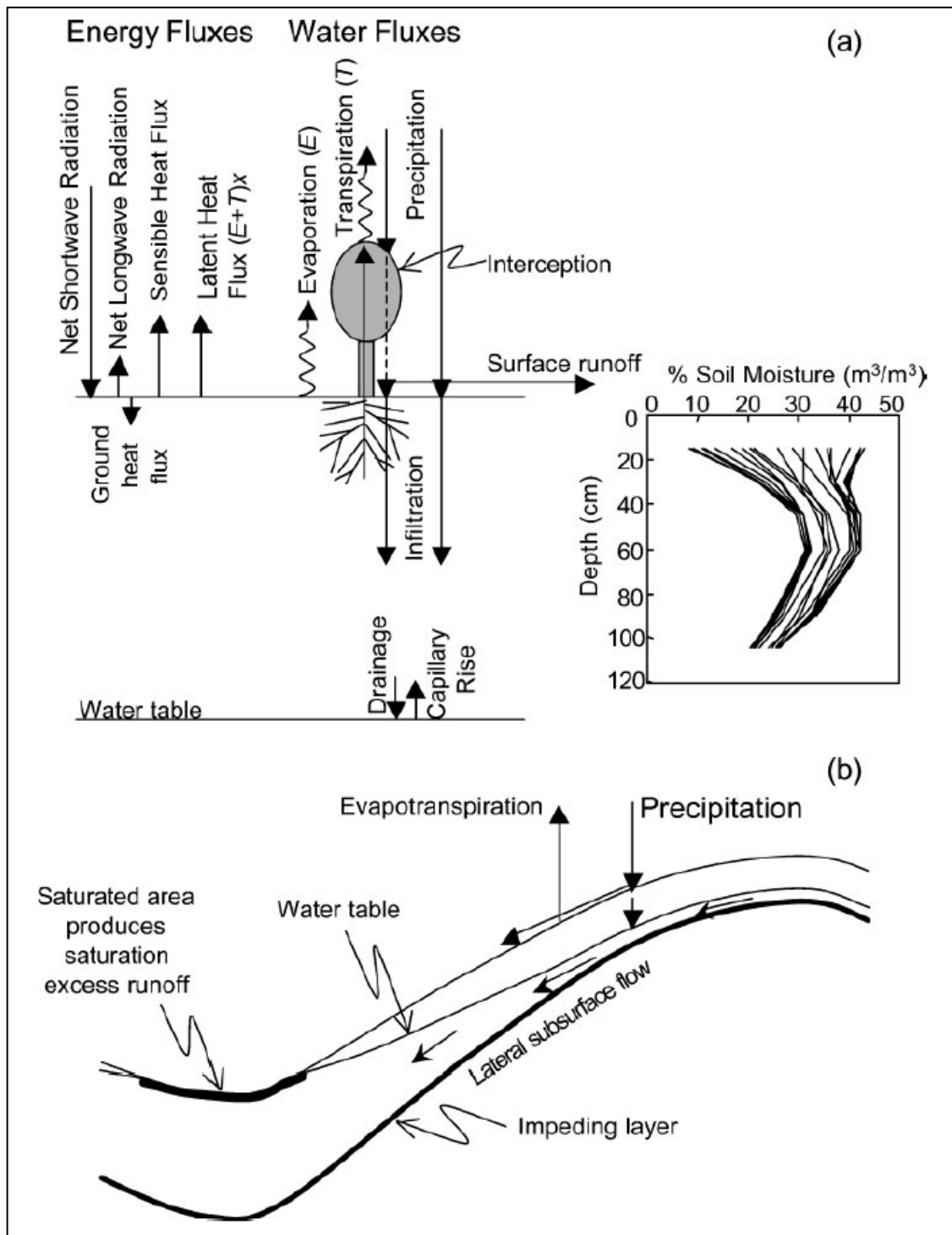


Figure 2.1. Figure taken from Western et al. (2002). (a) On the left: one-dimensional conceptualization of water fluxes affecting soil moisture, with also shown the related surface energy fluxes. On the right: soil moisture-depth profile series from an experimental site. (b) A two-dimensional conceptualization of fluxes affecting soil moisture in a hillslope.

Soil properties

Soil heterogeneity affects the distribution of soil moisture through variations in texture, organic matter content, porosity, structure, and macroporosity (Crow et al., 2012). Significant soil moisture variations may therefore exist over small spatial distances due to variations in soil particle and pore size distributions. The impact of soil hydraulic conductivity on soil moisture spatial patterns was demonstrated in numerous studies (e.g. Mohanty & Skaggs, 2001). Texture, in particular, can control the nature of water transmission and retention in the soil. Coarsely textured soils with a high proportion of sand will drain better than finely textured soils such as clays, and as such, will have a lower water-holding capacity and lower SM. Thus, soils with higher sand content exhibited persistently drier soil moisture conditions than soils with lower sand content. In addition, the organic matter content of soils directly influences soil albedo (and thus reflectance), in an inversely proportional way. By controlling soil albedo, soil organic matter influences evaporation rates especially from bare or lightly vegetated soil (Famiglietti et al., 1998). Soil colour can also influence albedo.

Topography

Topography-related parameters that affect the distribution of SM include slope, aspect, curvature, specific contributing area, and relative elevation (Petropoulos et al., 2013; Famiglietti et al. 1998). Slope influences processes such as infiltration, subsurface drainage, and runoff. Aspect and slope have been shown to have a direct control on the solar irradiance received, which, in turn, affects the rate of evapotranspiration from the land surface and, as a result, soil moisture. Land surface curvature is a measure of the landscape convexity or concavity and influences the convergence of overland flow. Areas characterized by high curvature tend to be characterized by a larger heterogeneity in SM than areas in which plan curvature is low. The specific contributing area is defined as the upslope surface area that drains through a unit length of contour on a hillslope. This parameter controls the potential volume of subsurface moisture that flows from a particular point on the land surface, affecting the distribution of soil surface moisture. Locations with larger contributing areas are generally expected to be wetter in comparison to locations with smaller contributing areas. Last but not least, relative elevation (so-called slope location) affects soil surface moisture directly by

affecting the degree to which orographic precipitation contributes to SM as well as indirectly due to its effect on soil water redistribution.

Vegetation

Vegetation type, density, and uniformity have all a strong effect on SM variability (Crow et al., 2012), which increased as a function of decreasing vegetation cover. Furthermore, the influence of vegetation on spatial variation in soil moisture is more dynamic as compared to soil and topographic factors. The presence and amount of vegetation influence the concentration of surface SM by adding organic matter to the soil surface layer and also by extracting water from the soil to be used for vegetation transpiration. The presence of vegetation cover also influences soil moisture via the throughfall pattern and shading of the soil layer that is imposed by the vegetation canopy, which, in turn, influences the rate of evaporation from the soil and soil hydraulic conductivity via the impact of root activity (Famiglietti et al., 1998). Effect of vegetation on SM variability is more pronounced during the growing season. In a field campaign reported in Panciera (2009), land cover was found to have a strong influence on soil moisture distribution; specifically, cropped areas exhibited persistently wetter-than-average conditions, and forested areas exhibited drier-than-average conditions, while grassland sites were more representative of the area average soil moisture conditions.

Climate

Space-time dynamics of SM are strongly influenced by a variety of climatological and meteorological factors, including incoming solar radiation, wind, humidity, and, most importantly, precipitation (Crow et al., 2012). Variations of incoming solar radiation and wind can both influence the rate of evapotranspiration from soils, either increasing or decreasing SM. At its most simple state, the characteristics of surface runoff, subsurface flow, and soil moisture depend on the characteristics of precipitation (phase, intensity, duration, etc.). In this sense climate, particularly the balance between potential evapotranspiration and rainfall at seasonal time scales, largely determines the seasonal average temporal pattern of soil moisture. The climatological factors, of course, strongly influence the dominant vegetation and soil type and the type of land use that

can be utilized in that location and are, in turn, themselves influenced by those factors as well as by topography (Famiglietti et al., 1998; Western et al., 1999).

Soil moisture variability generally reflects a combination of the effects of more than one of the above factors, and the processes that are dominant in determining the spatial pattern of soil moisture can change over time; moreover, factors controlling SM variations work on different spatial scales (Fig. 2.2), also affecting SM correlation structure and length. Thus, the significance of the exact relationship between SM and the different factors is variable and difficult to quantify precisely.

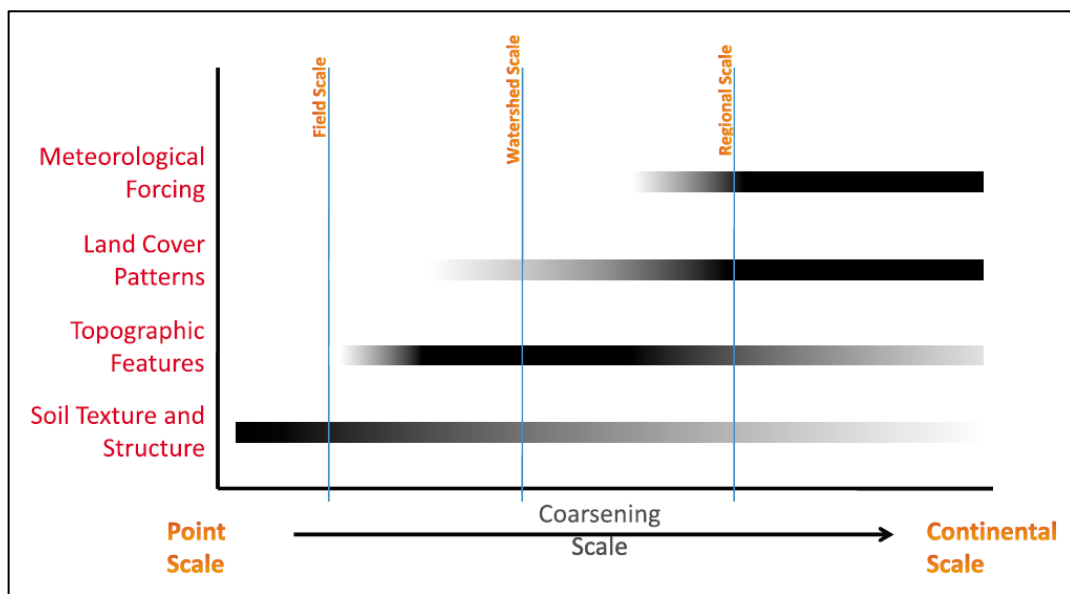


Figure 2.2. Figure taken from Crow et al. (2012), representing the dominant physical controls on soil moisture spatial variability as a function of scale ('land cover patterns' refer to vegetation characteristics). The grey shading of bars reflects the relative importance of each control at various scales with increasing intensity according to importance.

A large number of studies investigated the role of factors influencing soil moisture variability, both in time (e.g. Porporato et al., 2004) and in space (e.g. Riley & Shen, 2014), across the various scales. In time, soil moisture is mainly driven by precipitation and evapotranspiration, and its temporal variability is also a function of soil characteristics, vegetation, topography and groundwater (Porporato et al., 2004; Rosenbaum et al., 2002). In space, the same meteorological factors, i.e., precipitation and evapotranspiration, have a clear impact on soil moisture patterns at large scales. At finer spatial scales, static factors such as land cover, topography and soil texture/structure affects soil moisture spatial variability. Several authors compared soil

spatial patterns with these static factors (e.g. Brocca et al., 2007; Western et al., 1999) obtaining moderate to low predictive capability, largely varying across sites and climates. Therefore, our knowledge about the factors influencing spatial variability of soil moisture is still limited (Brocca et al., 2017b).

Currently, three different approaches are used for the monitoring of soil moisture from the point to the global scales: in situ observations, hydrological or land surface models, and remote sensing; they have different characteristics and will be described in detail below.

As mentioned, the various applications that need soil moisture information are focused on different spatial scales; in this sense, the high SM variability has important implications on modelling and measurement operations. Blöschl & Sivapalan (1995) suggested that both the measurement and the modelling scale can be defined as a scale triplet consisting of spacing, extent, and support (Fig. 2.3). Spacing is the separation between points at which measurements are made or between computational points in a model. Extent refers to the total coverage of the measurements or model. Support is the area over which a measurement averages the underlying variations, or over which a model assumes homogenous conditions.

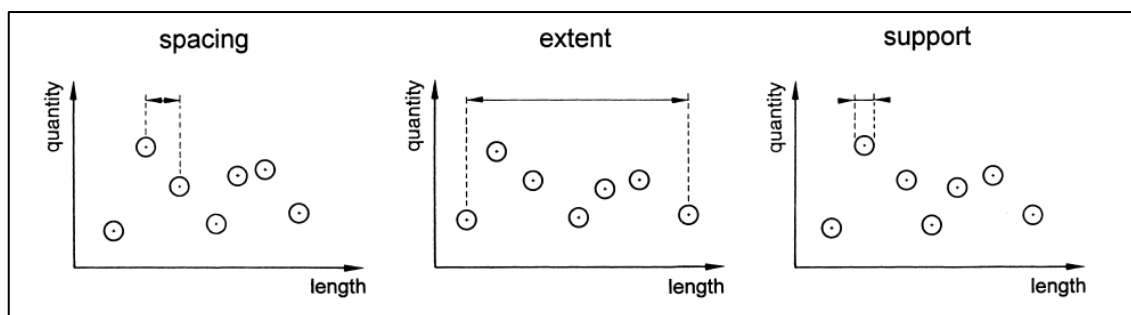


Figure 2.3. Figure taken from Western & Blöschl (1999), representing the definition of the scale triplet (spacing, extent and support). This scale triplet can apply to samples (i.e. measurement scale) or to a model (i.e. modelling scale).

The effect of measurement or model scale should be viewed as relative to the scale of interest of the process (e.g. point scale, watershed scale SM), in order to verify its consistency in investigating SM dynamics. In this sense, consideration of the scale triplet is important for example in both the choice of the monitoring approaches and the development of the sampling scheme. For example, in order to monitor the SM spatial variability in a plot (i.e. the extent), in situ instruments can be used (as they are

characterized by a small support) with a suitable number of measurement points (to contain the spacing and do not lose the small-scale variability); the direct use of remote sensing observations would not be equally appropriate, as they are commonly characterized by a large support (i.e. footprint sizes of several kilometres), where small-scale variations are averaged and lost.

2.4 Soil moisture key applications

Knowledge of the temporal and spatial variability of soil moisture is critical for understanding and predicting processes which are significantly influenced by this variable. Soil moisture has been widely recognized to have a fundamental impact on water, energy, and carbon cycles and, hence, plays a crucial role in the climate system and thus was included in the list of Essential Climate Variables (ECVs) (e.g. Dorigo et al., 2015). Soil moisture availability, in conjunction with atmospheric conditions, controls evaporation from bare soils and evapotranspiration from vegetated soils and the partitioning of incoming solar energy into latent and sensible heat fluxes (Seneviratne et al., 2010) (surface energy balance components are shown in Figure 2.1a in addition to the moisture fluxes). Because of the effects on atmospheric heating and moisture content, land-surface interaction and the role of soil moisture in this interaction have become an area of great interest to atmospheric modellers. Energy and water fluxes controlled by SM have important impacts on the atmospheric boundary layer dynamics and in turn on climate and weather and the prediction of each. Thus, soil moisture is deemed to be an essential variable to improve the accuracy of long term as well as seasonal scale numerical weather forecasts (de Rosnay et al., 2013; Dirmeyer & Halder 2016; Drusch & Viterbo, 2007). Several studies also analysed the role of SM on evapotranspiration (e.g. Martens et al., 2015; Miralles et al., 2011; Seneviratne et al., 2006), air temperature (e.g. Hirschi et al., 2014; Miralles et al., 2012, 2014), generation and location of precipitation (e.g. Findell et al., 2011; Guillod et al., 2015; Taylor et al., 2012), occurrence of heat waves (e.g. Ciais et al., 2005; Hirschi et al., 2011).

Then, soil moisture has an important influence on agricultural and ecological processes, as it plays a crucial role in the plant growth and hence in primary production in terrestrial ecosystems. By regulating plant growth, soil moisture is a valuable indicator for agricultural monitoring and crop yields forecast (also supporting management

practices), as well as for drought risk assessment. Moreover, SM has an important influence on a variety of soil processes including erosion (by controlling runoff), soil chemical processes and solute transport, and ultimately pedogenesis. Finally, soil moisture is also a regulator for various processes such as photosynthesis and soil respiration (e.g. Reichstein & Beer, 2008) and makes a strong control on variations in the global carbon cycle (e.g. Ahlström et al., 2013).

Soil moisture is also an important state variable for the quantification of the different components of the hydrological cycle. SM controls the partitioning of precipitation into runoff and infiltration, also playing a fundamental role in runoff modelling and flood forecasting. For this reason SM data are considered a valuable information to be integrated in flood early warning systems (e.g. Norbiato et al., 2008; Javelle et al., 2010; Van Steenberg & Willems, 2013). The degree of prior saturation is an important control on catchment response to rainfall and subsequent flood generation, with overland flow being larger and occurring more quickly on wetter soils and in catchments where areas of saturated soils are more extensive (e.g. Petropoulos et al., 2013). Thus, knowledge of soil moisture dynamics allows to more accurately model and forecast streamflow and flood events, and more generally the watershed dynamics.

Several studies analysed the role of soil moisture conditions in determining the hydrological response of a basin during extreme events such as floods (e.g. Berthet et al., 2009; Massari et al., 2014a), also in comparison with other potentially relevant runoff controlling factors and by considering different environments (e.g. Castillo et al., 2003; Penna et al., 2011; Zehe et al., 2010). In many cases the soil moisture conditions turned out to be a key factor in runoff production and explain much of the observed hydrological response (e.g. Zehe et al., 2010; Rodríguez-Blanco et al., 2012).

The sensitivity of the catchment response to soil moisture is influenced by the predominant runoff mechanisms, with the wetness conditions having a critical role on both surface and subsurface runoff generation. With regards to surface runoff component, there are two generation processes. Saturation excess runoff occurs when a drainage constraint combined with large accumulated rainfall depths results in a water table that reaches the surface, causing the whole profile to saturate and the production of runoff. Infiltration excess runoff occurs when the rainfall intensity exceeds the capacity of the surface soil to transmit infiltrating water, causing the soil surface to saturate and the production of runoff. In the case of infiltration excess runoff, for example, the actual

impact of the wetness conditions takes shape through the storm intensity relative to the current infiltration rates (e.g. Castillo et al., 2003). Specifically, when rainfall intensities are much larger or smaller than infiltration rates, initial wetness conditions are not critical; when these are of similar magnitude, runoff becomes highly sensitive to initial conditions. Thus, the relative influence of SM on peak discharge, response time and runoff volume can also depend on the characteristics of both the storm and the basin (e.g. topography, soil properties, presence of vegetation) (e.g. Penna et al., 2011); anyway, the various studies generally agree on the utility of having good estimators of soil moisture, as lack in models about soil water content representation typically results in substantial errors in runoff predictions (e.g. Berthet et al., 2009; Massari et al., 2014a).

The effect of wetness conditions on the runoff response was studied by using hydrological models (e.g. Castillo et al., 2003), observations (e.g. Zehe et al., 2010) or both (e.g. Massari et al., 2014b). Non-linear empirical relationship (e.g. Brocca et al., 2008) as well as neural networks (e.g. Tayfur et al., 2014) were proposed to link some runoff characteristics with both rainfall and pre-event wetness indicators. Simplified approaches such as the Soil Conservation Service (1972) infiltration method also compute the runoff depth associated to a rainfall event through a variable representing the soil retention capacity, which depends on the previous moisture conditions. The use of modelled or measured soil moisture was generally found more reliable in characterizing the basin antecedent wetness conditions during flood simulations, with respect to other proxies such as the antecedent precipitation and baseflow indices (e.g. Brocca et al., 2008, 2009).

Most of the basins showed a non-linear behaviour when prior wetness conditions and runoff characteristics are compared on an event basis, resulting characterized by a typical threshold relationship, with runoff coefficients (i.e. the ratio between runoff and rainfall volume) significantly increasing when a certain soil moisture value was exceeded (e.g. Western & Grayson, 1998; Zehe et al., 2010), as showed in Fig. 2.4. This behaviour was found in catchments with different topographic, climatic and land use characteristics, with varying values of the moisture threshold, likely due to factors such as soil type, soil depth and climatic conditions (Penna et al., 2011).

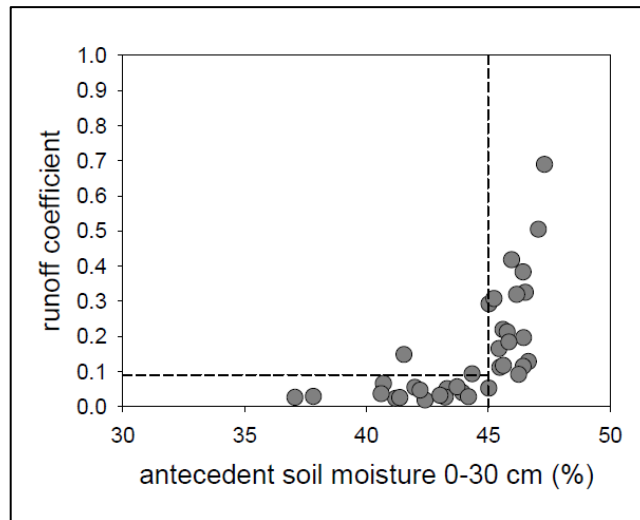


Figure 2.4. Figure taken from Penna et al. (2011), representing the threshold behaviour in the relationship between soil moisture prior to the rainfall event and the corresponding runoff coefficient.

These considerations suggest that the dynamics of soil moisture represents an important topic to be investigated to explain many processes of major significance in hydrological practice, especially the rainfall-runoff transformation. In this sense, the benefit of using soil moisture observations is not only limited to a proper initial conditions characterization in an event-based modelling approach (e.g. with respect to other wetness conditions proxies). Indeed, SM measurements could have a great potential in improving flood simulation through data assimilation techniques, that allow for updating the model soil moisture states in an optimal way.

2.5 Soil moisture modelling

Numerical simulations can be valuable practical tools for prediction and understanding of a system behaviour. In this sense, many types of models have been developed to represent soil moisture dynamics as effect of basic processes such as precipitation, evapotranspiration, infiltration, and drainage. Models predict relationships between physical system variables as a solution of specific mathematical structures, like simple algebraic equations or differential equations, which in this case reproduce the stores and fluxes within the system of interest. Given the specific model structure, accurate forcing and ancillary datasets are necessary, and model parameters need to be properly calibrated, representing characteristics of the system that are constant in time, or that can be assumed constant over the time scales of interest.

Models are usually developed to interpret a specific scientific question or solve a prescribed practical problem, so being very different in terms of model structure, spatial (horizontal and vertical) and temporal discretization, ways to simulate physical processes and corresponding parameterization, required inputs and then modelled soil moisture output. However, according to Koster et al. (2009), simulated soil moisture does not have an unambiguous meaning. It is a strongly model-specific quantity, essentially an “index” of the moisture state, with a dynamic range defined by the specific formulations utilized by the given model (Koster & Milly, 1997), in addition to model-specific soil parameters such as porosity, hydraulic conductivity, wilting point, and layer depth. Large differences are seen in the soil moisture products generated by different models, even when the models are driven with precisely the same meteorological forcing (Dirmeyer et al., 2006), owing to differences in physics and the meaning of parameters; however, once systematic deviations such as the different climatological statistics are accounted for, the different models tend to produce very similar information on temporal soil moisture variability (Koster et al., 2009).

Land surface models (LSMs) used in applications such as numerical weather prediction keep track of the soil moisture dynamics through state variables typically defined at a number of vertical subsurface levels, typically on a horizontal regular grid; these LSMs are addressed to simulate energy and water fluxes at the land surface to provide a lower boundary condition for the atmospheric model that is used to predict the weather. Hydrological models for rainfall-runoff simulations were instead characterized by a different processes representation and focused on basin scale. An analysis of the different approaches between land surface modelling in hydrology and meteorology can be found in Graham & Bergstrom (2000), and here summarized. Meteorologists are concerned primarily with solving the energy balance, whereas hydrologists are most interested in the water balance. Meteorological climate models typically have multi-layered soil parameterisation that solves temperature fluxes numerically with diffusive equations. Hydrological models are not usually so interested in soil temperatures, but must provide a reasonable representation of soil moisture to get runoff right. To treat the heterogeneity of the soil, many hydrological models use only one layer with a statistical representation of soil variability. Such a model can be used on large scales while taking sub-grid variability into account. Hydrological models also include lateral transport of water, which is an imperative if river discharge is to be estimated.

Examples of models simulating soil moisture that are developed for weather and climate modelling (e.g. Moran et al., 2004) are: NOAA-Noah Model (Chen et al., 1996; Koren et al., 1999), TOPMODEL-based land surface atmosphere transfer scheme (TOPLATS) (Famiglietti & Wood, 1994a; 1994b; Peters-Lidard et al., 1997), variable infiltration capacity model (VIC) (Liang et al., 1994, 1996), NCAR Community Land Model (CLM) (Bonan, 1998; Dickinson et al., 2006), Catchment LSM (Koster et al., 2000), Mosaic (Koster & Suarez 1996); examples of models developed for hydrological applications (e.g. Moran et al., 2004) include HYDRUS 1D and 2D models (Simunek et al., 1998; Simunek & van Genuchten, 1999), MIKE-SHE (Refsgaard & Storm, 1995), soil and water assessment tool (SWAT) (Srinivasan & Arnold, 1994; Arnold et al., 1998).

In general, models are able to provide regular soil moisture estimates, at the desired temporal and spatial resolution (e.g., subhourly and 100 m; de Rosnay et al., 2014; Bierkens et al., 2015). In the case of hydrological models, they can treat the basin as a single unit (lumped models) or discretize it into smaller units (distributed models), for example by introducing a regular grid, to take in account the spatial variability of forcings, ancillary data or parameters.

Models have so filled the gap in spatial (and also temporal) scales between in situ and remote sensing observations (Robinson et al., 2008), providing data at an intermediate scales that can be proper for the different applications, e.g. the basin scale in hydrological studies. In addition, model development can assist in interpreting observations (Western et al., 2003).

However, model predictions are affected by errors in model physics, parameters, input data (i.e., meteorological observations and ancillary information as soil type and land use), as well as by an inadequate treatment of sub-grid scale spatial variability (Walker & Houser, 2005; Reichle et al., 2004; Brocca et al., 2017a).

Not only the quality but also the resolution of meteorological observations (in space and time) and static information such as land use and soil texture maps plays a very important role, as they have to be consistent with the spatial and temporal resolution of the model (e.g. Brocca et al., 2017b). Models must anyway deal with scale effects related to the variability of soil moisture in an appropriate manner if they are to represent the system behaviour effectively (Western et al., 2003).

Models are being built with a higher degree of complexity in order to better represent processes; however, more complex modelling structure results in more parameters to be estimated, and may lead to overparameterized given the data available for parameter calibration (Walker & Houser, 2005). For instance, the saturated hydraulic conductivity, affecting the vertical and lateral movement of the water in the soil, is characterized by large variability (e.g. Morbidelli et al., 2006) and it cannot be observed over large areas and at different depths as it would be required (Hopmans et al., 2002). The same applies to other soil hydrological properties, e.g., porosity, pore size distribution. Additionally, many key hydrologic processes are extremely difficult to parameterize (e.g., irrigation, snow melting, and interception).

Therefore, modelled soil moisture data surely represents an important dataset that, however, needs to be used with caution. In order to remedy for the uncertainties in model estimates, data assimilation systems combine the complementary information from observations and the spatially and temporally complete information given by models into a superior estimate of soil moisture (e.g., Reichle & Koster, 2005).

Soil moisture is represented in models with two most widespread approaches, i.e. by solving the Richards equation or by using a bucket representation (Western et al., 2003; Romano, 2014; Vereecken et al., 2008). Their main features are briefly discussed.

Models based on the well-known Richards equation, or simplifications thereof, represent the vertical distribution of soil moisture and the movement of liquid water under the influence of gravitational and capillary (suction) gradients. Richards equation solution requires that hydraulic conductivity and water retention functions of soil matric potential, as well as the initial and boundary conditions, have to be defined. Such models are often applied at a point (the scale for which they were originally derived) and try to represent the physical processes and physically realistic soil moisture values. They assume that vertical flow is via the soil matrix and usually not consider lateral flow. Two- or even three-dimensional descriptions of water flow processes might be warranted when lateral processes such as overland and subsurface flow determine soil moisture dynamics. Flow in macropores and cracks can be important and needs additional model components for its representation. The Richards equation in one or more dimensions has been successfully applied to describe soil water fluxes at scales ranging from field (e.g. Clemente et al., 1994) to catchment (e.g. Schoups et al., 2005). Despite these successful applications, the appropriateness of the Richards equation is

still under debate. A detailed investigation of its validity is hampered by the inherent spatial variability of soils, the current inability to obtain spatially averaged measurements of matric potential, and the difficulty of accurately measuring boundary conditions at larger spatial scales. Often such models are used to represent spatial average soil moisture content over larger areas (such as in numerical weather prediction models), with spatial variability being largely ignored and uniform soil parameter values being used. Then, soil moisture data are often analysed using one-dimensional models, which assume that flow occurs only vertically; this for the sake of computational simplicity and because of the inability to appropriately populate multidimensional models with hydraulic properties, and the absence of reliable information about spatially varying boundary conditions. However, inappropriate selection of model dimensionality can reduce the accuracy of model predictions of soil moisture. Examples where consideration of horizontal flow might be important include analysis of soil profile moisture in humid landscapes where lateral flow is an important contribution to runoff.

In alternative, soil moisture is typically modelled by simplified water balance equations (e.g. Montaldo & Albertson, 2003), that are similar to the capacity type bucket models often used in conceptual rainfall-runoff models. Soil water balance models do not try to represent physical soil moisture that can be measured in the field but rather to represent the temporal variation in the depth-integrated content of the soil water store; they directly address moisture content as the variable of interest and assume gravity as the driving force for water flow, whereas models using the Richards equation also include the role of matric potential. This approach does not require a full characterization of the hydraulic conductivity and water retention functions.

With reference to a single-layer bucket model without lateral fluxes, the nonlinear differential equation of water balance for a soil layer of depth Z_r is usually written as follows (Milly, 2001; Rodríguez-Iturbe & Porporato, 2004):

$$nZ_r \frac{d\theta_{SD}(t)}{dt} = I[\theta_{SD}(t), t] - E[\theta_{SD}(t)] - T[\theta_{SD}(t)] - L[\theta_{SD}(t)] \quad (2.5)$$

Soil thickness Z_r is the control volume, n is the soil porosity, θ_{SD} is the degree of soil saturation (i.e., the volumetric soil water content normalized by soil porosity), I is actual infiltration rate from rainfall, E is actual evaporation rate, T is actual transpiration rate, and L is actual leakage rate from the bottom of the bucket. It is worth noting that

evaporation, transpiration and leakage rates are considered as a function of average soil saturation, θ_{SD} , whereas the surface runoff is generated only when the total rainfall rate $r(t)$ exceeds the infiltration capacity (which also depends from θ_{SD}) and only a fraction of the incoming precipitation is able to infiltrate into soil. Unlike the Richards equation, the bucketing approach incorporates a more simplistic representation of soil moisture dynamics and not fully resolve the local vertical variations in soil water contents.

Most of the models used to predict floods are based on these simplified “bucket style” models. Conceptual buckets connected in series are also sometimes used to represent shallow and deep layers within the soil profile. Such single- or multiple-layer bucket models can be applied at a point, but more often they are used within conceptual catchment models to represent soil moisture storage, averaged over the whole catchment (i.e. lumped models) or referred to a single grid cell or subcatchment (i.e. distributed models). Conceptual hydrological models describe in a simplified manner the processes that are assumed to dominate the hydrologic response of a system; using a distributed approach allows for the spatial variation of the simulated processes within the catchment while overcoming the inherent difficulties in the mathematical formulation and parameterization of the processes required instead by the models based on the Richards equation. In distributed models the soil water balance is implemented independently for each discrete element of the catchment, with lateral fluxes between stores taken in account in small scale applications.

Thus, there are several reasons for favouring a soil water balance approach. The estimation of effective soil water fluxes using Richards-based approaches heavily relies on knowledge of effective soil hydraulic parameters at the scale of interest and the validity of a capillary-based concept at larger scales. Although upscaling approaches enable estimation of effective properties from local-scale hydraulic properties, their validity is mostly limited to specific flow conditions. In contrast, simple water balance approaches have been postulated that may be considered scale-invariant, robust and parsimonious in terms of parameter specification and estimation. Then, it is very difficult to have information on soil matric potential, that is measured only at the point scale, in a destructive manner, and conventionally only covering the wet part of the moisture retention characteristic.

Despite these advantages, closure of the soil water balance models still requires constitutive equations for the unknown fluxes, such as lateral flow, capillary rise and

evaporation (Reggiani et al., 2000). These fluxes are usually derived on the basis of flux potential gradient relationships such as Darcy's law. However, this equation requires the specification of the hydraulic conductivity relationship, which is difficult to obtain. Also the prediction of root water uptake requires adequate knowledge of both the hydraulic conductivity and the water retention functions. Therefore, the issue of closing the water balance is often solved by using simplified descriptions that neglect specific processes (e.g. lateral flow and capillary rise in (2.5)) or by linearization of the constitutive equations, and to what extent this limits the validity of the obtained results is not really known.

An overview of comparative studies carried out on the Richards equation and the bucket scheme, for the purpose of highlighting the major advantages and limitations of these models, can be found in Romano (2014). Problems regarding the representation of SM variability in models (including the subgrid variability, also in distributed models), the modelling of spatial patterns, as well as the specific challenges that come out in large-scale SM simulations, are well discussed in Western et al. (2003).

2.6 In situ soil moisture observations

Ground monitoring networks have an important role in further understanding the processes influenced by soil moisture. In situ sensors have been predominantly developed to provide soil moisture data for monitoring agricultural and ecohydrological processes, generally controlled by field-scale soil water dynamics (Robinson et al., 2008). Common ground-based methods can be applied over any depth, accurately calibrated, and logged at any time scale. The high accuracy of in situ measurements also makes them a potential reference for calibrating and validating land surface models and satellite-based soil moisture retrievals. However, ground-based measurement techniques generally monitor soil moisture at the point scale, as they usually have $<1 \text{ m}^2$ support (Ochsner et al., 2013). Thus, in situ sensors are generally characterized by a very limited spatial representativeness, as there is a substantial amount of spatial variability in soil moisture (Western et al., 2002). It is worthy to highlight that some novel in situ techniques provide soil moisture measurements which are representative of the field scale (e.g. ground-penetrating radar, cosmic-ray neutron probes) (Jonard et al., 2018);

however, their broader applications still requires further research and technical investigations (Brocca et al., 2017b).

Given the small support and large spacing of typical ground-based networks, interpretation of the measurements at larger scales is challenging. A possible approach is to make a large number of measurements with sufficiently high spatial resolution to describe the spatial variability (e.g. Western & Grayson, 1998); an alternative is to relate the point scale measurements to areal soil moisture (e.g. Vachaud et al., 1985), which requires that a predictable time-stable relationship exist between point soil moisture and the spatial mean.

Comprehensive ground-based soil moisture observational networks are still very scarce, and only in few countries (e.g. United States) a good coverage of in situ stations is present. On a worldwide basis the number of networks and stations measuring soil moisture, in particular on a continuous basis, is relatively limited. Observation networks that measure in situ soil moisture are heterogeneous in terms of measurement principle and device, calibration techniques and installation methods, depth and placement of the sensors, and available measurement period. In many cases it is difficult to gain access to these datasets. Besides, there is a general lack of standardization of methods and protocolling which complicates the combined use of multiple networks in global studies. Furthermore, not every network applies quality control procedures to their data and if quality control mechanisms exist, they are not consistent.

Many of the available networks have been recently collected and harmonized as part of the International Soil Moisture Network (ISMN; Dorigo et al., 2011), which was initiated to establish and maintain a global in situ soil moisture database (Fig. 2.5). ISMN serves as a centralized data hosting facility where in situ soil moisture measurements from a large variety of individually operating networks are collected, harmonized, quality-controlled and made available to users. Incoming datasets provided by participating networks are harmonized in time and measurement units. For the ISMN a temporal resolution of 1 hour was considered sufficient for the common application purposes; observation datasets which are available at sub-hourly sampling rates have been thinned selecting the individual measurements at the hourly UTC reference time step. About the measurement units, soil moisture data provided to the ISMN are stored in the database as volumetric soil moisture ($\text{m}^3 \text{m}^{-3}$). Most current networks provide their data in volumetric soil moisture so often no conversion is

needed. If the data need to be converted, additional metadata (like bulk density or soil porosity) are required. Datasets contained in the ISMN are subjected to an automated quality control (Dorigo et al., 2013) and spurious observations are flagged, without modifying or removing it. There are two types of quality control procedures. The first category is based on the geophysical range and consistency of the measurements. About geophysical range, a threshold method is applied to flag values exceeding a plausibility range (0.0 to $0.6 \text{ m}^3 \text{ m}^{-3}$) or the local saturation point. To check the geophysical consistency, soil moisture measurements are confronted with precipitation and soil temperature. The second category of procedures (spectrum-based approaches) analyzes the shape of the soil moisture time series to detect outliers (spikes), positive and negative breaks, saturation of the signal, and unresponsive sensors. In 2015, the ISMN contained the data provided through 1400 measurement stations operated by 40 different networks (Dorigo et al., 2015). Available datasets include recently established operational networks as well as historical observations, like those contained in the Global Soil Moisture Data Bank (GSMDB; Robock et al., 2000), which was the first action to offer a centralized access point for globally available datasets and has now been closed.

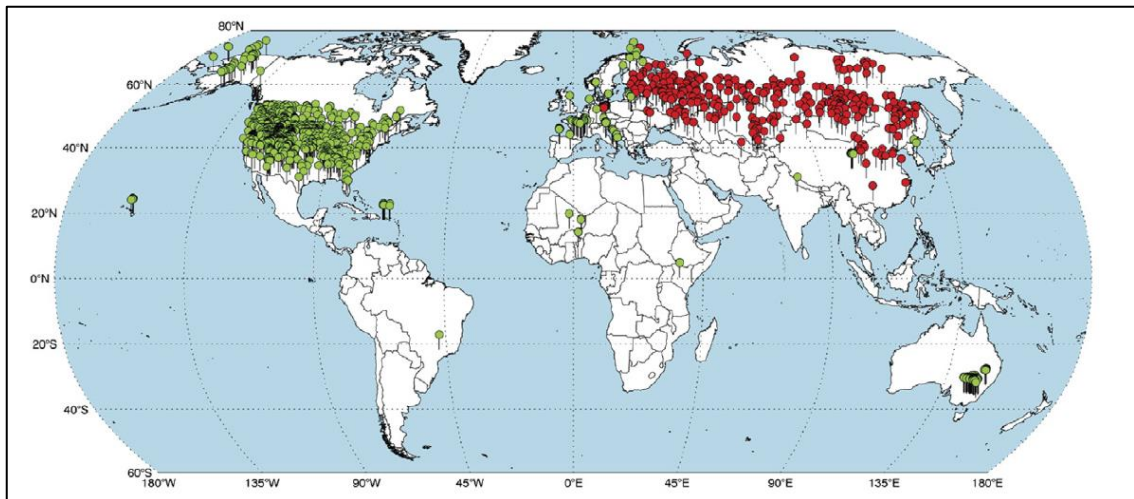


Figure 2.5. Figure taken from Ochsner et al. (2013), representing an overview of soil moisture stations contained in the International Soil Moisture Network (ISMN). Green dots show the stations from operational networks, red dots the stations that were imported from the Global Soil Moisture Data Bank. New networks have been added to the ISMN in the meantime.

As mentioned, several techniques are available for measuring soil moisture content in situ. The number of SM measurement methods has significantly increased over the last

years, with also the emergence of novel techniques. Most commonly, electromagnetic sensors (e.g. TDR) are used to establish continuous in situ soil moisture networks. Verstraeten et al. (2008) in their overview of the available conventional methods employed in estimating soil moisture concluded that even today there is not a single, clearly superior method suitable under all circumstances for measuring SM. As Dorigo et al. (2011) noted, even when one technique is consistently employed, SM measurements can be strongly influenced by several other factors including, for example instrument calibration, the installation conditions (e.g., installation depth and sensor placement), as well as the spatial resolution, geographical coverage, and representativeness of the measurements obtained, which are key factors for large area studies. Generally, required accuracy in the measurement of SM is subject to the considered application. Even though most sensor manufacturers claim to achieve the common accuracy requirement, these specifications are usually made for a default calibration, which is only valid for “typical conditions” achieved under laboratory conditions. A field-specific calibration (in place of universal calibration functions provided by the manufacturers) and accuracy assessment is indispensable to exploit the potential quality of the sensor and to quantify the actual accuracy. Additionally, an underrepresented but important issue of in situ sensors is related to their maintenance. Indeed, it is highly difficult to find long-term in situ soil moisture time series with good quality and consistency over time.

Overviews of the different in situ techniques have been given, for example, by Robock (2000), Robinson et al. (2008), Verstraeten et al. (2008), Dorigo et al. (2011), Dobriyal et al. (2012), Jonard et al. (2018), Romano (2014), Petropoulos et al. (2013). Some conventional frequently used methods, as well as some promising emerging techniques, are discussed below; however, this section does not aim to be comprehensive. The basic principles and the main characteristics of the different techniques, are presented, outlining its relative advantages and limitations.

Gravimetric method

Gravimetry refers to the measurement of soil water content by measuring the difference in weight between a soil sample before and after drying. Implementation of this technique is based on extracting a soil sample (usually 60 cm³) from the field which is then transferred to a soil analysis laboratory, where it is put into a drying oven at 105°C.

The soil sample is considered to be dry when its weight remains constant (usually after 24-48 h). The concentration of water in the soil is determined by subtracting the oven-dry weight from the initial field weight. The difference in mass gives the total soil moisture in the sample, which is converted to volumetric units using the density of the soil. Many different types of sampling equipment, as well as special drying ovens and balances, have been developed and used for this method. The gravimetric method is the oldest, the most direct and the most accurate method of measuring moisture content in the soil and remains the standard against which other methods (providing systematic measurements) are calibrated and compared. However, it cannot be used to obtain a continuous record of soil moisture at any one location because of the necessity of removing the samples from the ground for laboratory work. The key advantages of this approach are that it is easily and straightforwardly implemented using low cost technology and equipment. Nevertheless, the method has several disadvantages. It is a destructive method that precludes repeat sampling from exactly the same point. Furthermore, its implementation is generally time-consuming and labour-intensive, and difficult with rocky soils.

Neutron scattering

In this indirect method, the amount of water in a volume of soil is estimated by measuring the amount of hydrogen it contains, expressed as a percentage. Because most hydrogen atoms in the soil are components of water molecules, the backscatter of the slow neutrons after emission of fast neutrons from a radioactive source directly corresponds to water content in the soil, as the energy loss is much greater in neutron collisions with atoms of low atomic weight (i.e. hydrogen). A relationship with volumetric soil moisture content is obtained by calibrating the slow neutron counts with gravimetric samples. The soil volume measured by a neutron probe is bulb-shaped and has a radius of 1 to 4 m, according to the moisture content and the activity of the source. The neutron probe allows a rapid, accurate, repeatable measurement of soil moisture content to be made at several depths and locations. However, there are disadvantages such as the use of radioactive material requiring a licensed and extensively trained operator, the high equipment cost and the extensive calibration required for each site, as probes need to be calibrated to soil types and zones over a period of time with different soil moisture fractions. Furthermore, and perhaps most importantly, the probe's

relatively large volume of influence makes observations near the surface (top 20 cm) prone to errors, because fast neutrons can escape into the atmosphere, as well as adjoining air is also sampled. This technique is not common when frequent and automated observations are required, and its use has been proven most useful in measuring relative soil moisture differences rather than absolute SM (Dorigo et al. 2011).

Dielectric constant methods

Dielectric measurement takes advantage of the differences in dielectric constant (permittivity) values between different soil phases (solid, liquid, and gas). The dielectric permittivity is a measure of the capacity of a non-conducting material (such as soil) to transmit electromagnetic waves or pulses. Liquid water has a dielectric constant value of ~80 (depending on temperature, electrolyte solution, and frequency), air has a dielectric constant value of ~1, and the solid phase of 2 to 16. This contrast makes the apparent dielectric constant of soil very sensitive to variation in SM. The measurement of the apparent dielectric constant is then used to obtain the volumetric water content through calibration curves.

The first approach to be developed for measuring the soil dielectric constant and estimating the soil volumetric water content was the Time Domain Reflectometry (TDR) proposed by Topp et al. (1980). The TDR approach was found to provide accurate measurements for a wide range of soils and settings, also creating for the first time the possibility of automated in situ monitoring, and thus became the dominant technology for measuring soil moisture (Ochsner et al., 2013). Then, the emergence of capacitance type probes followed (Robinson et al., 2008).

The TDR measurement principle is based on the propagation velocity of guided electromagnetic waves emitted by a pulse generator and propagated along the waveguides of the TDR probe into the soil. The propagation velocity is determined from the measured travel time along the TDR probe (with a known length) which is dependent on the soil electromagnetic properties. The faster the propagation velocity, the lower the dielectric constant and thus lower soil moisture.

Capacitance probe electrodes and their adjacent soil form a capacitor with a capacitance that is a function of the permittivity of the soil and thus, also of the soil moisture content. In capacitance sensors based on Frequency Domain Reflectometry (FDR)

technique, an electromagnetic signal is propagated across the capacitor, and capacitance is evaluated by analysing changes in the reflected electromagnetic wave in the frequency domain.

Both type of sensors can be permanently installed to provide water content at different depths. Measurements are fast and easy, but they are very sensitive to installation because the sphere of influence is relatively small. The soil disruption can change water movement and water extraction patterns, resulting in erroneous data. Prongs can be damaged in hard or rocky soils. Accuracy is generally good and can be improved by specific calibration.

The capacitance probes typically operate at frequencies much lower than the effective frequency of TDR, and are also more sensitive to soil salinity and temperature errors. As a result, these probes are simpler and less expensive but also less accurate than TDR (Ochsner et al., 2013; Jonard et al., 2018). Because of the spatial variability of soil moisture, for some applications it has been suggested to use less accurate but cheaper sensors in order to increase the density of measurements within soil moisture networks. In addition to their cost, the high power consumption of TDR instruments may be another dissuasive argument for their use if a site has to be e.g. operated with stand-alone power supply.

Measurement of soil thermal properties

Heat dissipation and heat flux sensors make use of the principle that the thermal behaviour of the soil is closely related to its water content (e.g. Bittelli, 2011).

The heat dissipation technique uses a heat source (usually a heated needle) and temperature sensors (thermocouples or thermistors), immersed into a porous, ceramic block that equilibrates with the surrounding soil at a given water content. The needle is heated, and the rate of heat dissipation is measured by the temperature sensors. The thermal conductivity is then obtained, which depends on the block water content.

In the heat flux method, the pulse of heat is applied at one location and its arrival at another location is determined by measuring the soil temperature at the other location. The time required for the pulse of heat to travel to the second location is a function of soil thermal conductivity, which is related to water content.

As heat transport depends on soil thermal properties and so on water content, soil moisture can be also estimated from the distributed temperature sensing (DTS) method

(e.g. Ochsner et al., 2013), where fiber-optic cables work as linear sensors, with temperatures being recorded along the cable with a typical spatial resolution of 1 m, on measurement distances of also several kilometres. DTS uses a laser pulse emitted into the fiber-optic cable and partly scattered back all along the cable; some characteristics of the backscattered signal are temperature dependent, and the corresponding position on the cable can be determined from the travel time. Observing temperature dynamics with more DTS cables can provide a way to determine the soil moisture. Temperature changes can be due to the diurnal radiation cycle (the passive DTS method described in Steele-Dunne et al., 2010) or due to a heat pulse transmitted from the metal housing of one of the cables (the active DTS method proposed in Sayde et al., 2010). The main advantages are the large spatial extent and resolution that this technique offers and the low power requirements. Disadvantages include the difficulty of placing the fibers, for example at consistent depths; passive DTS remains challenging under conditions where the thermal response to the diurnal temperature cycle is not large enough to allow accurate estimation of soil moisture content (e.g., under dense vegetative canopy, at depths beyond the top few centimetres of the soil column, cloudy days).

Gamma-ray attenuation

Gamma ray attenuation is a radioactive method capable of determining the moisture content in the upper soil layers. The intensity of a gamma ray that passes through a soil section undergoes an exponential decrease that principally depends on the apparent density of the soil (assumed constant), the water contained in the soil and the coefficients of attenuation of the soil and of the water (both constants). The method consists of concurrently lowering a gamma-ray source (generally Caesium 137) and a gamma-ray detector (scintillator-photomultiplier) down a pair of parallel access tubes that have been installed in the soil. At each measurement level, the signal can be translated into the apparent wet density of the soil or, if the apparent dry bulk density of the soil is known, the signal can be converted into a measure of the volumetric soil-moisture content. The measuring equipment permits tracking of the evolution of wet density profiles and of the volumetric soil moisture at several centimetres of depth below the soil surface if the dry density does not vary with time. Gamma ray attenuation is unaffected by the state of moisture in the soil and can be also be used with frozen soil. Additionally, this technique is much easier to calibrate as it does not have to be site

specific. However, the measurements are not specific to water alone, as the apparent variations in dry density can confound the measurements of soil moisture. Then, its implementation requires the use of relatively expensive instrumentation and a greater level of user expertise. The high cost and difficulty of use limit the applicability of this technique in the field. In highly stratified soils, large variations in the bulk density and moisture content have been noticed while using this technique (Pires et al., 2005).

Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) is a physical phenomenon that can yield molecular properties of matter by irradiating atomic nuclei, in a magnetic field, with electromagnetic radio waves (Blümich et al., 2011). NMR is particularly sensitive to the presence of hydrogen protons and is therefore suited to investigate processes that involve water. When applied to the study of soil, NMR can be used to infer soil water content, total porosity, and pore-size distribution and to quantify bound and free fluid fractions. This technique subjects water in the soil to both a static and an oscillating magnetic field perpendicular to each other. A radio frequency detection coil, tuning capacitor, and electromagnetic coil are used as sensors to measure the spin echo and free induction decays. Generally, advantages and disadvantages of this technique are similar to those of the neutron scattering technique discussed earlier.

Cosmic-ray neutrons

Recently, a non-invasive method has been proposed based on SM relationship with low-energy cosmic-ray neutrons (CNR) that are generated within the soil and diffused back to the atmosphere (Zreda et al., 2008, 2012). More specifically, secondary high-energy cosmic-rays neutrons (originated by primary cosmic-rays interaction with atmosphere) penetrate the ground surface and interact with nuclei in soil to generate fast neutrons. The number of fast neutrons above the soil surface depends strongly on the number of hydrogen atoms in the surroundings because hydrogen atoms have a very high capacity to moderate fast cosmic-ray neutrons (that means to slow them down and turn them into low-energy thermal neutrons, so effectively removing the fast neutrons from the system). The number of hydrogen atoms increases with increasing soil water content and hence soils with high water contents re-emit fewer fast neutrons than soils with low water content. Thus CRN sensors can infer soil water dynamics from the change of fast

neutron intensity with time in near-surface environment. Portable neutron detectors are placed a few meters above the soil surface and allow for a sampling interval of several minutes to hours. About supporting volume, soil moisture information is inferred over a depth of 10 to 70 cm below surface (Zreda et al., 2008), depending on soil type, water content and distance from the sensor (Köhli et al., 2015). In contrast to the methods presented above, cosmic-ray measurements integrate soil moisture measurements over much larger horizontal scales (i.e. areas of a few hundred square meters, with radius that can decrease with increasing air density and humidity, with increasing vegetation density and with increasing soil moisture, according to Köhli et al., 2015) and hence have the potential to bridge the spatial mismatch between point measurements using contact methods and remote sensing estimates over large areas. Limitations include the need to isolate the signal of SM from other sources of hydrogen source cosmic rays such as some minerals (e.g., clay), vegetation, and organic matter as well as surface and atmospheric water. Measurements of these variables are required in order to calibrate the soil moisture measurements.

Ground penetrating radar

Ground Penetrating Radar (GPR) measurements are based on the transmission and reflection of an electromagnetic wave in the soil (Chanzy et al., 1996). The transmitter antenna of the radar systems generates radio-waves propagating in a broad beam. The receiver detects variations in the electrical properties of the sub-surface by detecting the part of the transmitted signal that is reflected. A number of radar antennas are moved over the ground surface simultaneously when assessing moisture content. GPR is a fast and non-destructive technique with high resolution that can penetrate beyond the surface layer and is capable of covering large areas with differential hydrological conditions in a short time. Soil volume estimations are relatively quick. This method bridges the gap between point measurements and remote sensing. The performance of GPR varies across soil types (Doolittle & Collins, 1995). Many soil types, due to their high electrical conductivity, are radar opaque and dissipate radar energy, thus restricting its use (Doolittle & Collins, 1995). Steep and rocky slopes limit its use due to the large sizes of the antennas. The use of GPR in forests is difficult because trees behave as reflectors generating erroneous data (Schrott & Sass, 2008).

2.7 Remote sensing soil moisture observations

Soil moisture estimate from remote sensing has experienced a growing development over the last four decades (Karthikeyan et al., 2017a, 2017b; McCabe et al., 2017; Paloscia & Santi, 2013). Satellite observations are attracting increasing interest because of their continuous monitoring of land surface over very large areas, with frequent revisiting time. Global coarse-scale surface soil moisture (SSM) retrievals are obtained from sensors on board polar satellites, through algorithms that estimate SSM states from raw measured variables such as backscattering coefficient or brightness temperature, in order to obtain products for research and operational purposes. Although not specifically designed for measuring soil moisture, several satellite instruments have proved to be very useful for this purpose over the years, most notably microwave sensors operating at suboptimal frequencies (e.g. ASCAT, AMSR-E), with good correspondences being found between satellite retrievals and ground observations taken over a large variety of environmental conditions (e.g. Albergel et al., 2009; Brocca et al., 2010b, 2011a; Jackson et al., 2010), and in recent years two soil moisture monitoring dedicated missions have been launched, i.e. Soil Moisture Ocean Salinity (SMOS) in 2009 and Soil Moisture Active Passive (SMAP) in 2015. Significant developments have been made in SSM retrieval algorithms accuracies, as well as in increasing the temporal and spatial resolutions (that are usually approximately daily and of the order of some tens of kilometers, respectively) of the most widely used satellite measurements; in this regard, current research strives to further improvements, also by exploring the blending of data from different sensors to overcome the single specific limitations (e.g. Entekhabi et al., 2010b; Liu et al., 2012a).

A variety of studies has shown that SM at the surface layer can be measured to some degree by all regions of the electromagnetic spectrum (Petropoulos et al., 2015); in particular, different methods were developed for the retrieval of soil moisture from microwave, optical and thermal satellite sensors.

Optical remote sensing techniques based on soil surface reflectance measurements make use of visible, near-infrared (NIR), shortwave infrared (SWIR) bands, with spectral information being related to SSM as a function of spectral absorption features. Given that water absorbs energy, a soil with higher SM will, in theory, have less reflective intensity than soils containing less moisture. Various studies have explored the

relationship between spectral reflectance and soil moisture (e.g. Stoner & Baumgardner, 1981; Lobell & Asner, 2002) and have confirmed that, for bare soils, increase in SSM generally leads to a decrease in soil reflectance, although Liu et al. (2002) found that the opposite was true at high moisture levels. In addition to moisture content, reflectance measurements are also strongly affected e.g. by the soil composition and physical structure.

The estimation of SSM can be also made by measuring the thermal infrared (TIR) emittance, related to the land surface temperature. The common scheme most often utilised in thermal remote sensing of SSM is to decouple the surface thermal properties from ambient temperature (daily temperature cycle) by calculating the thermal inertia (i.e. a physical property that characterises the surface resistance to ambient temperature change) (Petropoulos et al., 2015). The thermal inertia, in turn, is dependent upon the thermal conductivity and the heat capacity, which both increase with soil moisture. However, the relationship between thermal infrared signal and soil water content is a function of soil type and is largely limited to bare soil conditions (e.g. van de Griend et al., 1985; Zhao & Li, 2013).

Optical and thermal sensing of surface soil moisture have several limitations, that include the shallow soil penetration (only the top millimetres of the surface) and the poor temporal resolution (e.g. Moran et al., 2004; Zhang & Zhou, 2016). Then, the relationships to soil water content are weak in presence of vegetation cover. Furthermore, unlike the longer microwave wavelengths, optical and thermal signals have limited ability to penetrate clouds and vegetation canopy. Optical methods are not suitable to be applied at night time, while SSM retrievals from TIR methods are highly influenced by atmospheric effects (Barrett & Petropoulos, 2013). Efforts to retrieve SSM have achieved success only when models are fit for specific soil types in the absence of vegetation cover (Petropoulos et al., 2015), and most of the models are empirical in nature and thus vary across time and land cover types and are a function of local meteorological conditions (Moran et al., 2004), with lack of transferability to other regions.

However, recent advances in thermal and optical remote sensing methods for soil moisture estimation have been made (e.g. Kustas & Anderson, 2009; Rahimzadeh-Bajgiran & Berg, 2016; Zhang & Zhou, 2016), in addition to the good spatial resolution and the presence of multiple operating satellites, showing promises also related to the

availability of hyperspectral sensors for reflectance-based methods, as well as to the strong relation between TIR signal and SSM in thermal methods. Furthermore, other opportunities are provided by the synergistic use of measurements from optical and thermal bands, as well as of optical/TIR with microwave data (e.g. Barrett & Petropoulos, 2013).

Most progress has been made utilising the microwave domain of the electromagnetic spectrum (Engman & Chauhan, 1995; Njoku & Entekhabi, 1996; Jackson et al., 1996; Shutko, 1982; Ulaby et al., 1996), particularly within the low frequency range (<6GHz), where microwave sensors can detect fine changes within the dielectric permittivity of the soil, and the penetration depth through soil and vegetation is higher. Microwave observations are not limited by clouds, daytime and/or atmospheric conditions, but are influenced by the vegetation cover and the roughness of the soil surface; however, at lower frequencies the effects of vegetation and roughness on the remote sensing signal decrease (Robinson et al., 2008). Microwave measurements are the most widely used for SSM retrievals (Fig. 2.6), also providing operational products (e.g. Wagner et al., 2013b; Mecklenburg et al., 2016).

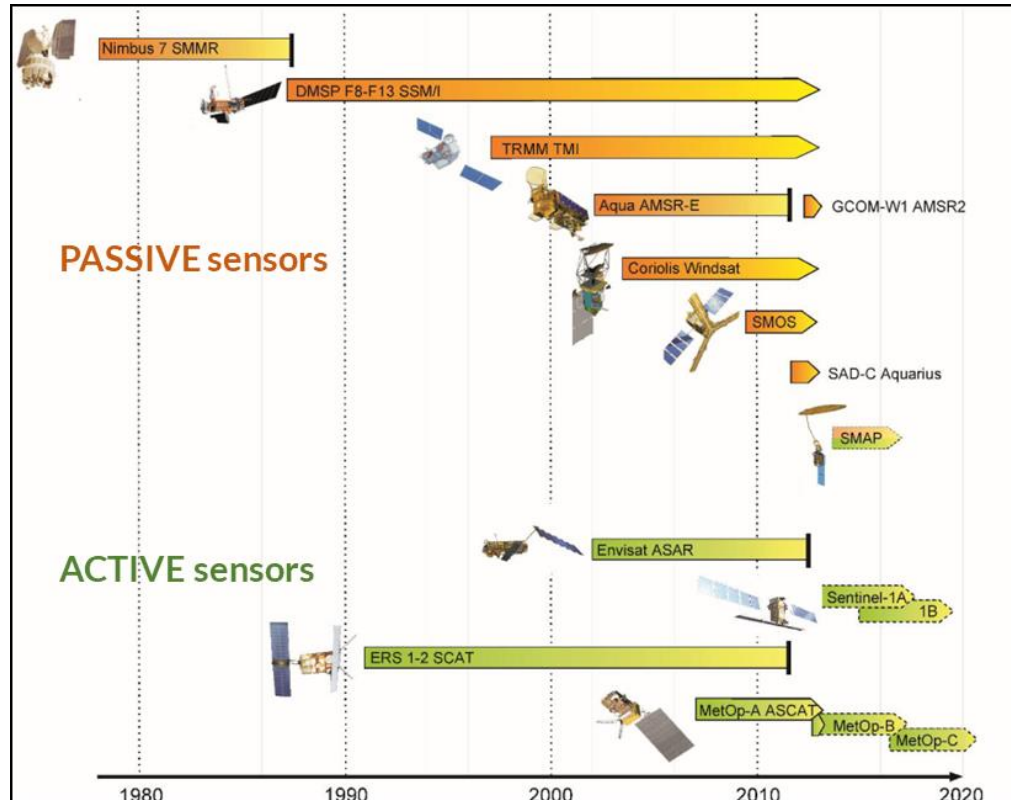


Figure 2.6. Active and passive microwave sensors contributing to soil moisture monitoring (adapted from <http://www.esa-soilmoisture-cci.org/>).

On the basis of their energy source, microwave sensors can be grouped into two categories: active (radar, scatterometer) and passive (radiometer) sensors. The passive sensors detect the natural microwave radiations emitted from the land surface in the form of brightness temperature measurements (T_B). The active sensors send microwave pulses towards the land surface and detect the reflected and scattered signals as backscattering coefficients (σ_0). Microwave measurements are sensitive to soil moisture due to the large contrast between dielectric constants of dry soil (typically ~ 3.5) and water (~ 80). The brightness temperature of the soil surface is proportional to the emissivity, ϵ , while the backscattering coefficient is related to the reflectivity, R . Emissivity, according to the Kirchhoff's law, complements the surface reflectivity, $\epsilon = 1 - R$, and both are strongly influenced by dielectric properties of soil surface and thus by soil moisture variations. More precisely, increase in SSM leads to increases in R and σ_0 and to decreases in ϵ and T_B .

From a user's perspective, passive microwave radiometers offer the advantage of overpassing the same surface almost every day; however, they are hampered by the coarse spatial resolution, which generally is not below 25 km. Among active microwave sensors, it is important to distinguish between Synthetic Aperture Radars (SARs) and scatterometers. The former (SAR) are able to detect land surface information at a high spatial resolution (e.g. < 1 km) but with a long revisit period (e.g. > 10 days). The latter (scatterometer) are characterized by nearly the same spatial-temporal resolution of radiometers.

Different portions of the microwave region of the electromagnetic spectrum were commonly used for estimating soil moisture, i.e. X (8-12 GHz), C (4-8 GHz), and L (1-2 GHz), with the latter identified as optimal (e.g. Kerr et al., 2012). In particular, observations within L-band at 1.4 GHz show the maximum sensitivity to surface soil moisture (Njoku & Chan, 2006), minimizing the impacts of vegetation and surface roughness on the signal; for this reason, radiometers onboard the dedicated missions SMOS and SMAP operate at this specific frequency.

The depth of the investigated surface layer is not precisely defined, since differences in microwave frequencies and soil moisture conditions primarily lead to different emitting or penetration depths. Several studies suggest that the contributing depth for microwave sensing is between 0.1 and 0.25 times the wavelength (Jackson et al., 1996). In practice,

the thickness of the layer whose moisture is estimated by L-band sensor is usually assumed of the order of 5 cm, while for C-band (~5 GHz) and X-band (~10 GHz) it is not higher than 2 and 1 centimetres, respectively. The measuring depth of microwave sensors also depends on the soil moisture profile: it was noted that the effective retrieval depth is shallowest during wet soil moisture conditions (e.g. Escorihuela et al., 2010). Furthermore, other factors influenced the effective contributing depth, such as the soil characteristics and temperature profile (e.g. Shutko, 1982; Lv et al., 2018).

Microwave measurements generally depend on both soil surface characteristics and known sensor configuration (such as frequency, polarization and incidence angle). An important practical issue for soil moisture retrieval is that the part of microwave signal related to soil surface characteristics responds to the dielectric constant (largely determined by soil moisture) but also to other factors such as the soil surface roughness and the vegetation canopy, making it difficult to separate the soil moisture contribution to the signal from other factors. In this sense, the retrieval of SSM from microwave measurements can be considered an ill-posed problem, because, in general, more than one combination of SSM, roughness and vegetation cover has the same electromagnetic response (e.g. Paloscia et al., 2008); several retrieval algorithms were developed, that address the problem with different approaches (see Sect. 2.7.1 and 2.7.2).

Soil surface roughness tends to decrease the sensitivity of both σ_0 and T_B to the soil water content (e.g. Robinson et al., 2008). Vegetation has a double effect on the surface emission and backscatter, both attenuating the signal from the soil and adding its own contributions, and may eventually completely obscure the soil moisture signal above wavelength-dependent vegetation water content density thresholds (Dorigo et al., 2017). More precisely, the masking effect of vegetation increases with frequency, and it is generally considered that soil moisture can be monitored for levels of vegetation water content lower than about 3-5 kg/m² at L-band and 1.5 kg/m² at C-band (Wagner et al., 2007). With regard to passive sensors, attenuation is due to both vegetation absorption and scattering, that can be effectively modelled with two parameters, the vegetation optical depth and the single-scattering albedo, respectively (Mo et al., 1982). With regard to active sensors, total backscatter from vegetated surfaces is composed not only from underlying ground surface backscatter (subject to attenuation in the canopy), but also from canopy volume scattering and multiple path interactions between canopy and ground. In this case the vegetation geometric structure is also a key factor; dense forests

and shrubs are usually opaque to C-band radar (which is the common frequency of the considered active sensors), while sparse forest, grassland and agricultural crops are partly transparent (Wagner et al., 2013b).

Then, retrievals are impossible under snow and ice or when the soil is frozen (as dielectric properties of the water changes dramatically), while complex topography, surface water, and urban structures have an adverse effect on the retrieval quality (e.g. Dorigo et al., 2017). Passive microwave observations can also be affected by human-induced Radio Frequency Interference (RFI). Flags and attributes are usually included in SSM datasets to detect retrievals impacted by these factors.

Currently, SSM data are provided from the following quasi-operational (i.e. with data available either in near real time or few days after sensing) coarse resolution satellite sensors (Brocca et al., 2017b):

- (1) the Soil Moisture Active and Passive (SMAP) mission (L-band radiometer) starting from April 2015 with ~36 km/2-day spatial/temporal resolution (Entekhabi et al., 2010b);
- (2) the Advanced Microwave Scanning Radiometer 2 (AMSR2) onboard the Global Change Observation Mission for Water, GCOM-W1, satellite (C- and X-band radiometers) starting from July 2012 with ~25 km/1-day spatial/temporal resolution (Kim et al., 2015);
- (3) the Soil Moisture and Ocean Salinity (SMOS) mission product (L-band radiometer) starting from January 2010 with ~50 km/2-day spatial/temporal resolution (Kerr et al., 2016);
- (4) the Advanced SCATterometer (ASCAT) onboard MetOp satellites (C-band scatterometer) starting from January 2007 with ~25 km/1-day spatial/temporal resolution (Wagner et al., 2013b).

It is worthy to note that the National Aeronautics and Space Administration (NASA) SMAP mission originally involved the synergistic use of active and passive instruments; unfortunately, active radar on SMAP stopped working after only 3 months of operation, so only radiometer data are currently available.

Additionally, global soil moisture products were released within the ESA (European Space Agency) Climate Change Initiative (CCI), by merging multiple active and passive microwave sensors (Liu et al., 2012a; Dorigo et al., 2017), providing a nearly daily dataset starting from 1978 with a grid spacing of 0.25°.

In the next months, higher resolution (~1 km) soil moisture products will become available from SARs on Sentinel-1 satellites in the framework of the Copernicus Global Land Service (Bauer-Marschallinger et al., 2019); it is worthy to note that thus far a high-resolution soil moisture product from SAR sensors has never been made available. Others satellite sensors can be currently employed for soil moisture retrieval, but without the delivery of the corresponding soil moisture products.

Several processing levels are typically adopted, according to the definitions provided by the Committee on Earth Observation Satellites (CEOS). Each level represents a step in the abstraction process by which data relevant to physical information (e.g. backscattering coefficient) are turned into data relevant to geophysical information (e.g. surface soil moisture), and finally turned into data relevant to thematic information (e.g. root zone soil moisture). CEOS (<http://ceos.org/>) has identified five levels of data products:

- Raw Data - Data in their original packets, as received from a satellite.
- Level 0 - Reconstructed unprocessed instrument data at full space time resolution with all available supplemental information to be used in subsequent processing appended.
- Level 1 - Unpacked, reformatted level 0 data, with all supplemental information to be used in subsequent processing appended. Optional radiometric and geometric correction applied to produce parameters in physical units. Data generally presented as full time/space resolution. A wide variety of sub level products are possible.
- Level 2 - Retrieved environmental variables (e.g. soil moisture) at the same resolution and location as the level 1 source data.
- Level 3 - Data or retrieved environmental variables which have been spatially and/or temporally re-sampled (i.e. derived from level 1 or 2 products). Such re-sampling may include averaging and compositing.
- Level 4 - Model output or results from analyses of lower level data (i.e. variables that are not directly measured by the instruments, but are derived from these measurements).

2.7.1 Active microwave sensing of surface soil moisture

Active microwave sensors for SSM retrieval can be classified into two categories, imaging and non-imaging sensors. Radio detection and ranging (radar) is the most common imaging sensor; with the Synthetic Aperture Radar (SAR) technique, backscattered signals from short coherent microwave pulses are processed together to simulate a very long aperture capable of high surface resolution. Scatterometers (designed to measure wind speed and direction over the sea surface but also used for SSM retrieval) are distinguished as non-imaging sensors.

While scatterometers are characterized by nearly the same spatial-temporal resolution of radiometers, i.e. approximately daily revisit time and coarse scale (~25 km), SAR systems are designed to acquire high spatial resolution backscatter images and allow for soil moisture mapping at fine scales without recurring to downscaling procedures. SARs provide data with pixel sizes in the order of tens of meters, although aggregation is required to overcome the problem of noise and fulfil the accuracy required by many applications. This increases the effective pixel size (e.g. 500 m - 1 km), which however remains much lower than those derived from radiometers and scatterometers. However, spaceborne SAR systems were generally not designed to obtain repetitive and continuous coverage (Pathe et al., 2009). A significant limitation of SAR was that satellites provide only weekly repeat coverage and even longer for the same orbital path; for example, the first generation SAR systems onboard the European Remote Sensing satellites (ERS-1/2) had a scheduled repeat pass every 35 days for the same orbital path. Moreover, SAR systems usually alternatively operate in different ‘conflicting’ modes (generally acquiring data with different spatial resolutions and swath widths), and due to high power consumption the number of fine spatial acquisitions per orbit was limited, further worsening the temporal resolution. More generally, spaceborne SAR systems have to reach compromises between spatial coverage, spatial resolution and revisit time, that are hardly compatible with an operational use of SSM retrievals. Due to this reason, despite the large volume of research conducted on the derivation of soil moisture from SAR, routinely produced soil moisture maps have long been unavailable and few literature studies addressed the problem of operational soil moisture monitoring using SAR (e.g. Pierdicca et al., 2013). However, due to recent advancements in observational capabilities (in terms of spatial/temporal resolution, radiometric accuracy, very large swath) that characterize the

European Space Agency (ESA) Sentinel-1 (S-1) mission (Torres et al., 2012), SAR soil moisture products will be made available in the near future (Bauer-Marschallinger et al., 2019).

At present, most active remote sensing systems operate only at the C-band, which is not the best band for retrieving SSM, as at this frequency the effects of surface roughness and vegetation cover on the backscattering coefficient are high and an estimate of SSM with the accuracy required by many applications is still problematic (Satalino et al., 2002; Kerr, 2007; Paloscia et al., 2008).

Several backscattering models have been suggested for retrieving the surface soil moisture from active microwave measurements. According to a common classification (Petropoulos et al., 2015; Karthikeyan et al., 2017a), these models can be grouped into four classes: theoretical or physical, semi-empirical, empirical and change detection models.

Theoretical models allow for simulating the backscattering coefficients by providing a physical-based description of the interactions between microwave radiation and soil; the dielectric constant of the soil surface, and hence the SSM, is estimated from the mathematical inversion of these models. Amongst their advantages is the independence on local site conditions and sensor characteristics, while downsides include the requirement for a large number of input parameters which make their parameterisation difficult and complex (Moran et al., 2004). A dominant physical based radiative transfer model is the Integral Equation Model (IEM) originally developed by Fung et al. (1992), that essentially quantifies the backscattering coefficient of a bare soil as a function of the unknown dielectric constant, surface roughness, and known sensor configuration (wavelength, polarisation and local incidence angle). Application of IEM to real world simulations has generally shown mixed results (e.g. Baghdadi & Zribi, 2006; Álvarez-Mozos et al., 2007; Paloscia et al., 2008), with the main difficulty being related to the sensitivity to the surface roughness parameters and the difficulty associated with their correct characterization (Zribi & Dechambre, 2003), which is only achievable through intensive roughness measurement campaigns over large regions.

Semi-empirical backscattering models offer a compromise between the complexity of the theoretical backscattering models and the simplicity of the empirical models. They provide relatively simple relationships between surface properties and backscatter metrics that reflect, to a certain extent, the physics of the scattering mechanisms. Their

main advantage is that they are not site dependent and may be applied when little or no information on surface roughness is available. The most widely used semi-empirical models are those developed by Oh et al. (1992) and Dubois et al. (1995). The Oh model (Oh et al. 1992, 2002; Oh, 2004) relates the ratios of radar backscatter in separate polarizations (HH, HV, and VV) to volumetric soil moisture and soil surface roughness. The main advantage of the Oh model is that only one surface parameter is required, and when multipolarized data are available, both the dielectric constant and surface roughness can be inverted without the need for field measurements (Álvarez-Mozos et al., 2007). Similar to the Oh model, the Dubois model accounts for copolarized backscatter only (HH and VV) and do not require fully polarimetric systems. Both Oh and Dubois models are generally valid only over bare soil surfaces and has shown mixed results in real applications (van Oevelen & Hoekman, 1999; Baghdadi & Zribi, 2006; Álvarez-Mozos et al., 2007; Panciera et al., 2014), although some studies have reported good results over sparsely vegetated soil surfaces (e.g. Neusch & Sties, 1999). Empirical backscattering models, generally based on experimental measurements, gain insight into the interaction of microwaves with natural surfaces through simple retrieval algorithms (e.g. Mathieu et al., 2003; Holah et al., 2005; Álvarez-Mozos et al., 2007). For example, many of those empirically based studies have shown that a linear relationship between the backscattering coefficient and SSM is a reliable approximation for one study site under bare soil conditions, assuming that roughness does not change between successive radar measurements. The use of empirical backscattering models is characterised by some limitations. Due to the lack of a physical basis, empirical models are generally confined to the local specific datasets and implementation conditions for which they were derived (i.e. surface conditions and radar parameters at the time of the experiments) (Chen et al., 1995). However, various authors have proposed calibration approaches for adjusting those empirical models to other implementation conditions (e.g. Baghdadi et al., 2008; Zribi et al., 2005). Another limitation is that empirical models require many high-quality in situ soil moisture measurements, obtained over time, for calibration, which could be a costly and challenging task, and not always attainable; in particular, large databases over a variety of study sites are essential to ensure that developed models are robust and transferable (Baghdadi et al., 2002). The models described under physical, semi-empirical and empirical approaches are valid only under bare soil conditions. The contribution of vegetation on backscatter

coefficient can be taken into account by coupling a vegetation scattering models. Models are of variable complexity and accuracy and their application is based on availability of the required in situ data (Kornelsen & Coulibaly, 2013). Models such as the Michigan Microwave Canopy Scattering (MIMICS) (Ulaby et al., 1990) and that of Bracaglia et al. (1995) can simulate a variety of species within a vegetation type; more complex models exist (e.g. Stiles & Sarabandi, 2000) that reproduce a single crop type in a highly physically representative manner, however they are not often used due to the extensive parameterization of the vegetation. A significantly less complicated model is the Water-Cloud Model (WCM) (Attema & Ulaby, 1978), which is semi-empirical with some vegetation parameters locally calibrated (e.g. Bindlish & Barros, 2001), and has found widespread use in a variety of vegetation and climate regimes (Kornelsen & Coulibaly, 2013).

While the physical, semi-empirical and empirical models rely on backscatter coefficient information obtained in a single time period, change detection approaches use the multi-temporal passes made by an active sensor at a location to obtain the relative change in soil moisture, i.e. these techniques try to interpret backscatter changes without attempting to explain the absolute backscatter level. These approaches are based on the assumption that the temporal variability of surface roughness and vegetation biomass is generally present at a much longer time scale than that of soil moisture; therefore, the change in backscatter coefficient between repeat passes results from change in soil moisture conditions. Thus, a multi-temporal radar dataset can be used to minimise the influence of surface roughness and vegetation biomass, and maximise the sensitivity of radar backscatter to changes in SSM. Change detection is an attractive technique because it presents a simple way of accounting for surface roughness and vegetation effects; however, it has some drawbacks. It should be noted that such methods are not easy to apply for cultivated areas, as surface roughness and vegetation biomass change dramatically over short time periods. Until recently, significant limitations were related to the SAR applications due to their coarse repeat coverage time, that means lack of dense and long-term time series of acquisitions with the same sensor configuration. Finally, one big challenge remains the characterization of the influence of seasonally changing vegetation on the backscatter signal. Several change detection methods have been proposed (Barrett et al., 2009). In particular, change detection algorithms were developed to obtain SSM retrievals from spaceborne scatterometers such as ASCAT

(Naeimi et al., 2009), that currently provides several soil moisture operational products (Wagner et al., 2013b). The so-called TU-Wien change detection model, originally proposed for the ERS scatterometer by Wagner et al. (1999) and based on a normalization that accounts for both roughness (assumed to be constant over time at scatterometer spatial scale) and vegetation seasonal effects (considered invariant from year to year), is used for this purpose. The model parameters, describing maximum dry and wet soil conditions in term of backscatter values at a reference incidence angle for the different days of the year, are locally calibrated through the statistical analysis of long-term radar observations time series; for the relative soil moisture estimation, the actual backscatter value is normalized to the reference incidence angle and then compared with dry and wet backscatter references for that day of the year.

With regard to the space programs contributing to global scale soil moisture retrievals through active microwave sensors, the first influential one was the ESA European Remote Sensing (ERS). ESA's two ERS satellites, ERS-1 and -2, were launched in 1991 and 1995, and ended in 2000 and 2011, respectively. ERS was the first ESA program in Earth observation with the overall objectives to provide environmental monitoring. Satellites payloads included the Active Microwave Instrument (AMI), that incorporated two separate radars (Attema, 1991): a Synthetic Aperture Radar (SAR) and a wind scatterometer (SCAT), both in the C-band. AMI-SAR and -SCAT observations were also used for SSM retrievals (e.g. Moran et al., 2000; Wagner et al., 1999). Both ERS-1 and ERS-2 satellites have a standard orbit repeat cycle of 35 days, with a revisit time on a location of 2–7 days (Karthikeyan et al., 2017b); AMI scatterometer mode cannot be operated in parallel to the AMI-SAR modes (Pathe et al., 2009).

The follow-up mission of ERS, that was initially conceived to cover research aspects of environment as well as provide operational meteorological data, was splitted by ESA into the dedicated Environmental Satellite (EnviSat) and Meteorological Operational (MetOp) satellites programmes.

MetOp series include three operating satellites (launched respectively in 2006, 2012 and 2018) carrying the Advanced Scatterometer (ASCAT). Even though ASCAT was not designed for soil moisture monitoring, it currently provide SSM retrievals for operational uses (see Sect. 2.7.1.1), and the continuation of the soil moisture products is also ensured in a long-term perspective by scatterometers onboard the MetOp Second

Generation (MetOp-SG) satellites (Brocca et al., 2017a), which will be launched in the next years.

EnviSat was launched in 2002 and ended in 2012; it carried the Advanced Synthetic Aperture Radar (ASAR), operating in C-band. Different methods were proposed to retrieve SSM from EnviSat ASAR measurements (e.g. Paloscia et al., 2008; Pathe et al., 2009). EnviSat was replaced by the ESA Sentinel series of satellites, which are elements of the European Global Monitoring for Environment and Security (GMES) program currently known as Copernicus; in particular, the Sentinel-1 (S-1) mission has taken over the SAR duties of EnviSat.

In contrast to the ERS and EnviSat missions, Sentinel-1 is based upon an operational concept, resulting in revisit time and spatial coverage that are dramatically improved and also providing high reliability and rapid data dissemination (Torres et al., 2012). The S-1 mission currently comprises a constellation of two polar-orbiting satellites (1A and 1B, launched in 2014 and 2016, respectively), each of them carrying a C-band (5.4 GHz) SAR sensor. Sentinel 1A and 1B share the same orbit plane with a 180° orbital phasing difference, offering altogether six days exact repeat. Over land, radar data will be predominantly acquired in the Interferometric Wide-swath mode (IW), in either VV+VH or HH+HV dual polarization, combining a large swath width (250 km) with a high spatial resolution (5×20 m on ground). Currently, several algorithms have been proposed to retrieve soil moisture operationally from Sentinel-1 data (e.g. Hornacek et al., 2012; Paloscia et al., 2013; Pierdicca et al., 2014). Change detection techniques can provide a good compromise between retrieval accuracy and processing time, thus allowing compliance with the timeliness requirements; in the framework of Copernicus Global Land Service, a SSM product at 1 km resolution will be made available, retrieved from S-1 data following the TU-Wien change detection method (Bauer-Marschallinger et al., 2019).

2.7.1.1 ASCAT

The Advanced Scatterometer (ASCAT) is the successor of scatterometers deployed on ERS, and currently is one of the instruments flown on board the MetOp satellites (Figa-Saldaña et al., 2002).

The MetOp is a series of three meteorological satellites developed by ESA and operated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT). MetOp satellites are in a near-circular sun-synchronous polar orbit, with equator crossing time at ~9:30 (local solar time) A.M. and P.M. in descending and ascending orbit direction, respectively. The first satellite, MetOp-A, was launched in 2006. The temporal coverage was irregular, ranging from 0 to 2 acquisitions on a single day, with an average revisit time of 1–2 days. In 2012 MetOp-B was launched, doubling the coverage frequency; finally, the third and last satellite, MetOp-C, has been launched in November 2018.

ASCAT is a real aperture C-band (5.3 GHz) radar that uses two sets of three antennas in VV polarization to continuously illuminate separated 550 km wide swaths on either side of the satellite ground track. In each swath, a regular grid of points (nodes) is defined, with one triplet of independent backscatter measurements per node, taken at different azimuthal and incidence angles. The nominal spatial resolution of the ASCAT backscatter measurements is 50 km on a nodal grid of 25 km, however a research product at a higher resolution of about 25 km (actual resolution varies somewhat across the swath from 25 to 34 km) on a 12.5 km nodal grid is also available (Wagner et al. 2013b).

ASCAT is developed with the primary objective of determine wind vector fields at sea surface on a global basis; however, it is also proven to be very useful in monitoring soil moisture. ASCAT measurements are used to generate high-quality soil moisture products for operational hydrological applications within the framework of the Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF) programme, which is part of the EUMETSAT application ground segment. Detailed reviews of ASCAT soil moisture products are given in Wagner et al. (2013b) and in Brocca et al. (2017a).

ASCAT backscatter measurements can be related to soil moisture content in the first 2 cm below the surface. SSM retrievals, in degree of saturation between 0 and 100, are obtained through the TU-Wien change detection method, originally developed by Wagner et al. (1999) and subsequently improved for ASCAT by Naeimi et al. (2009). The soil moisture retrieval algorithm is implemented within a processing software called Water Retrieval Package (WARP).

In the framework of the H-SAF project (<http://hsaf.meteoam.it/>) several soil moisture products are generated on a regular basis and distributed to users.

ASCAT measurements at the highest spatial resolution (25–34 km) are processed via the offline WARP chain in order to obtain the SSM time series on a discrete global grid (identified as WARP5 and characterized by a 12.5 km grid spacing). More details about the retrieval algorithm and processing strategy can be read in H-SAF (2014).

Operational near-real-time (NRT) SSM products are generated soon after each satellite (currently MetOp-A or -B) orbit completion, and given in swath geometry (at both 25 and 12.5 km sampling), by using a dedicated software package called WARP-NRT, that employs the model parameters derived off-line within WARP (Wagner et al., 2013b).

A small scale NRT SSM product (1 km resolution) results from downscaling of ASCAT SSM NRT data, based on a high-resolution, static backscattering characterization layer, computed from statistical analysis of multi-annual EnviSat ASAR imagery (Wagner et al., 2008). However, the added value of this product is not very clear, given that the downscaling parameters are static and all information about the soil moisture temporal variability still comes from the large scale ASCAT measurements (Wagner et al., 2013b).

Finally, two root-zone soil moisture products (a near-real-time one and an offline time series) have been developed, both based on the assimilation of ASCAT SSM observations into the ECMWF Land Data Assimilation System (de Rosnay et al., 2013), providing global daily soil moisture estimates at a spatial resolution of 25 km for four soil layers (0-7, 7-28, 28-100 and 100-289 cm).

2.7.2 Passive microwave sensing of surface soil moisture

Passive microwave remote sensing utilizes highly sensitive radiometers to measure the naturally emitted microwave radiation at a particular wavelength, expressed as brightness temperature (T_B). Differences of nearly 100 K in brightness temperature can be observed between very dry and wet soils, that is large in relation to the precision of a typical microwave radiometer (≤ 1 K), making it possible to measure changes in SSM of less than 1% (e.g. Robinson et al., 2008). All of these instruments are typically characterized by broad spatial coverage and high temporal resolution but coarse spatial resolutions (25–50 km), because, in order to detect the low quantities of emitted

radiation, the field of view for passive sensors must be large enough to detect sufficient energy to record a signal, resulting in a low spatial resolution (Barrett & Petropoulos, 2013).

The emitted energy detected by microwave radiometers generally includes contributions from the atmosphere (that also attenuates the emission from soil surfaces); however, atmospheric effects (that can be accounted for, e.g. Kerr et al., 2012) are considered negligible at low microwave frequencies, i.e. $\lambda > 5$ cm (Jackson et al., 1996). The interpretation and use of passive microwave signatures is still made complex by the influence of surface variables, e.g. the increase in soil roughness and vegetation biomass that reduce the sensitivity of T_B to soil moisture. However, towards the longer-wavelength region of the microwave spectrum (i.e. $\lambda > 10$ cm), the effects of vegetation and roughness are greatly reduced (e.g. Petropoulos et al., 2015), and L-band at 1.4 GHz was identified as optimal, as previously mentioned. Another main surface variable that should be accounted for in the retrieval method is the surface temperature (e.g. Wigneron et al., 2003); a general model assumption is that, for vegetated surface, the temperatures of soil and vegetation are approximately the same (Karthikeyan et al., 2017a). Soil texture and variability in the temperature of both soil and vegetation also affect microwave retrieval, yet have much less of an influence (Guerriero et al., 2012).

Different soil moisture retrieval approaches have been developed to deal with the various effects contributing to the surface microwave emission (Wigneron et al., 2003; Paloscia & Santi, 2013; Petropoulos et al., 2015; Karthikeyan et al., 2017a).

The radiative transfer theory is the basis of most SSM retrieval algorithms from passive microwave observations, with a radiative transfer model (RTM) being used to relate T_B with soil emissivity. The latter can be in turn related with the soil dielectric constant, through the Fresnel reflectivity equations for smooth soil surface, eventually adjusted to take in account the soil roughness, e.g. with the model proposed by Wang & Choudhury (1981). Finally, soil dielectric constant are linked with soil moisture using a dielectric mixing model, chosen among the several ones available in literature (e.g. Wang & Schmugge, 1980; Topp et al.; 1980; Hallikainen et al.; 1985).

With regard to the RTMs, Mo et al. (1982) developed a simplified form of the radiative transfer theory to quantify vegetation effects, known as the “tau-omega” model. In the tau-omega model, the brightness temperature, as observed from above the canopy for a given polarization, is due to the sum of three terms: (1) the radiation from the soil

attenuated by overlying vegetation, (2) the upward radiation directly from the vegetation, (3) the downward radiation from the vegetation, reflected upward by the soil and again attenuated by the canopy. The model is based on two parameters, the single-scattering albedo (ω) and the vegetation optical depth (τ). The former describes the ratio of scattering efficiency to total extinction efficiency of the soil emissivity within vegetation layer, and is a function of the vegetation type and the leaf characteristics (van de Griend & Wigneron, 2004). It is generally assumed that the scattering effects are minimal in the case of low frequency microwave emissions, resulting in low values being assigned to ω (Karthikeyan et al., 2017a); computations of the single-scattering albedo gave a range of values from 0.04 to about 0.13 (Mo et al., 1982; Owe et al., 2001). The vegetation optical depth (VOD) parameterizes the vegetation transmissivity, that also depends on the known sensor view angle; for frequencies lower than 10 GHz, τ can be expressed as a linear function of vegetation water content (Jackson et al., 1982). Over some vegetation water content thresholds, e.g. 1.5 kg/m² at C-band (Paloscia & Santi, 2013), the signal becomes totally saturated by vegetation components and retrieval of soil moisture is not possible. The tau-omega model involves the use of soil and canopy physical temperature. The surface temperature is dependent on soil properties and moisture content. Hence, some algorithms compute soil temperature as a function of soil type, soil temperature at surface and at a depth, soil moisture or soil dielectric constant, and frequency of radiation (these inputs are obtained from ancillary data sources) (Choudhury et al., 1982; Holmes et al., 2006; Raju et al., 1995; Wigneron et al., 2008; Wigneron et al., 2001). In general, soil and canopy temperature are assumed to be approximately equal (as previously mentioned), although few methods use temperature values obtained from independent sources for soil moisture retrievals (e.g. Wigneron et al., 2007).

The tau-omega model acts as a baseline for many retrieval algorithms (e.g. Jackson, 1993; Wigneron et al., 2007; Owe et al., 2008; Kerr et al., 2012), with main differences usually lie in the treatment of the observations (e.g. by using different frequencies, polarizations, or multiple overpasses or view angles), in the methods used to account for the effects of the main geophysical variables (soil roughness, vegetation, surface temperature), and finally in the conversion of the soil dielectric constant to soil moisture (i.e. the adopted dielectric mixing model) (Paloscia & Santi, 2013; Dorigo et al., 2017).

The first generation retrieval methods were developed for airborne observations with a mono-configuration sensor (i.e., single polarization/frequency channel and view angle) (Wigneron et al., 2003) and were found capable of very accurate analysis over well-defined and well-controlled areas (Petropoulos et al., 2015). These algorithms retrieve only soil moisture with an objective of minimizing the error between observed and modelled brightness temperatures in the horizontal or the vertical polarization. Physical parameters such as surface temperature, vegetation optical depth and surface roughness were typically obtained from ancillary data or empirical sources; however, the task became very challenging when these retrieval techniques were adapted to satellite instead of airborne observations.

An example of the algorithms proposed in this regard is the Single Channel Algorithm (SCA) by Jackson (1993), initially developed to support the relatively simple instrument configurations that were available on aircraft platforms. This approach assumes that the single-scattering albedo is negligible and that the atmospheric contributions are minimal, thus significantly simplifying the tau-omega model. The soil moisture is therefore estimated by sequentially performing temperature normalization, removing the attenuating effects of the overlying canopy and atmosphere, and estimating the associated smooth (i.e. removed surface roughness effects) surface emissivity using ancillary data.

Since the passive microwave satellite sensors can usually provide multi-configuration measurements (i.e. measurements for several configuration systems of the sensor in terms of polarization, view angle, and microwave frequency), other parameters such as vegetation attenuation effects and surface temperature can be retrieved along with soil moisture, and therefore, less ancillary information is required in the retrieval process. Since the vegetation optical depth is the most important variable that needs to be computed in the retrieval process, certain algorithms such as the Dual Channel Algorithm (DCA) (Owe et al., 2001), the 2-Parameter L-band Microwave Emission of the Biosphere (L-MEB) model (Wigneron et al., 2000), the Land Parameter Retrieval Algorithm (LPRM) (Owe et al., 2008), and the revised Land Surface Microwave Emission Model (LSMEM) (Pan et al., 2014), made the retrieval of VOD possible by the aforementioned procedure. They are called two-parameter retrieval models, to distinct them from algorithms that also estimate the surface temperature (e.g. Njoku & Li, 1999), i.e. the three-parameter retrieval models (Wigneron et al., 2003).

Furthermore, the effects of retrieving parameters such as surface roughness (Parrens et al., 2016) and single-scattering albedo (Konings et al., 2016), along with soil moisture and VOD have also been successfully explored. The two-parameter retrieval algorithms obtain surface temperature data from ancillary sources of thermal infrared or high-frequency microwave brightness temperature measurements (Karthikeyan et al., 2017a). However, implementation of this kind of approach requires a good parameterisation of the dependence of retrieved surface variables on the sensor configuration parameters (Wigneron et al., 2003). Numerous studies have reported good retrieval accuracies and relatively low error distribution in comparison to in-situ and modelled observations (Li & Rodell, 2013; de Jeu et al., 2014; Mladenova et al., 2014). Generally such studies have indicated that model performance is generally higher over sparse to moderately vegetated regions, where a decrease in accuracy is evident when transitioning to denser vegetated regions (Petropoulos et al., 2015).

Apart from the models having physical basis in the radiative transfer theory, soil moisture is derived from passive microwave measurements using statistical regression techniques towards a reference soil moisture dataset. Some regression models are derived from standard RTM (e.g. Al-Yaari et al., 2016), while few models involve either simple linear relationship (e.g. Jackson et al., 1999) or neural network based procedures (e.g. Rodríguez-Fernández et al., 2015).

Passive microwave remote sensing has been extensively used to retrieve soil moisture, with C- and X-band measurements that date back to 1978. A range of radiometers have indeed been in operation (Paloscia & Santi, 2013; Karthikeyan et al., 2017b), starting from the Scanning Multichannel Microwave Radiometer (SMMR), the Special Sensor Microwave Imager (SSM/I), and the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI). These sensors were essentially devoted to meteorological applications due to the relatively high frequency employed and the very coarse ground resolution; however, the possibility of obtaining surface feature information gave rise to both theoretical and experimental research.

The Scanning Multichannel Microwave Radiometer (SMMR), launched onboard the NASA Nimbus-7 satellite in 1978, was the first passive microwave satellite sensor that had the capability of retrieving surface soil moisture. Primarily, it was launched for obtaining sea surface temperature, surface wind speed, water vapor, and cloud liquid water content information. Brightness temperature was measured at five dual-polarized

frequencies from 6.6 to 37 GHz, and the global repeat coverage was up to six days. SMMR continued to provide data until 1987.

The Special Sensor Microwave Imager (SSM/I) carried aboard the United States Air Force Defense Meteorological Satellite Program (DMSP) satellites continued the legacy of SMMR since 1987. Its purpose was to obtain ocean surface wind speed, water vapor, cloud liquid water, and rain rate information. Measurements are made at four frequencies (19.3, 22.2, 37.0 and 85.5 GHz) in both horizontal and vertical polarizations, except the 22.2 GHz which is sampled in the vertical only. With an improved swath width and reduced power consumption, SSM/I achieved global coverage almost daily.

The Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) was one of the sensors onboard the TRMM satellite launched in 1997 jointly by NASA and Japan Aerospace Exploration Agency (JAXA). The satellite was launched primarily for measuring rainfall and energy exchange in the tropical and subtropical regions; TMI provides information on the integrated column precipitation content, cloud liquid water, cloud ice, rain intensity, and rainfall types (e.g. stratiform or convective). The TMI operated at five frequencies: 10.7, 19.3, 37, and 85.5 GHz at dual polarization and 22.2 GHz at vertical polarization. The satellite made two daily passes (ascending and descending) at a location with a near-equatorial orbit. The TMI had a higher spatial resolution than that of SSM/I due to its larger antenna and lower orbital altitude, and provided data until 2015.

The goal of estimating surface soil moisture became more attainable with the launch of the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), onboard the NASA's Aqua satellite that was launched in 2002. Aqua is along a sun-synchronous orbit and overpass times are near 1:30 A.M. (in descending orbit) and P.M. (ascending) local time at the equator. The AMSR-E instrument, that stopped operating in 2011 due to an antenna failure, measured geophysical variables related to the earth's water cycle, including: precipitation rate, cloud water, water vapor, sea surface winds, sea surface temperature, sea ice concentration, snow water equivalent, and also soil moisture. AMSR-E was developed jointly by JAXA and NASA. The sensor measured microwave emissivity in six frequencies (6.9, 10.7, 18.7, 23.8, 36.5, and 89.0 GHz) in both horizontal and vertical polarizations. The polar orbiting AMSR-E achieved global coverage within two days separately for ascending

and descending passes. Spatial resolution varied depending on the frequency, from 5.4 km at 89.0 GHz to 56 km at 6.9 GHz, and improved significantly over SMMR (from 150 km to 50 km for C-band radiometer) while also containing all channels of the previous sensors (SMMR, SSM/I and TMI).

AMSR-E was the first satellite sensor to incorporate soil moisture as a standard product, also specifying an accuracy goal less than $0.06 \text{ m}^3\text{m}^{-3}$ (Jackson et al., 2010). Several retrieval algorithms (using different physical formulations, parameters, ancillary data, and AMSR-E channels) have been developed to provide soil moisture products, by NASA (Njoku et al., 2003; Njoku & Chan, 2006) and JAXA (Koike et al., 2004; Lu et al., 2009) and other research groups, e.g. the previously mentioned LPRM, developed by the Vrije Universiteit Amsterdam (VUA) in collaboration with NASA (Owe et al., 2008).

Another sensor that contributed to SSM monitoring is the WindSat onboard the Coriolis satellite, launched in 2003 and developed by the Naval Research Laboratory Remote Sensing Division, the Naval Center for Space Technology for the U.S. Navy and the National Polar-orbiting Operational Environmental Satellite System (NPOESS) Integrated Program Office (IPO). WindSat was meant to demonstrate the capabilities of a fully polarimetric radiometer to measure the ocean surface wind vector. In addition to wind speed and direction, the instrument also measures other parameters such as sea surface temperature, total precipitable water, integrated cloud liquid water, and rain rate over the ocean. The radiometer operates in 5 frequencies (6.8, 10.7, 18.7, 23.8 and 37.0 GHz); all of them are fully polarimetric (i.e. in the six principal polarizations: V/H, $\pm 45^\circ$ linear, and left/right circular), except the 6.8 and 23.8 GHz that have only dual polarization (vertical and horizontal). The revisit time is almost daily, and the horizontal resolution is 25 km. A SSM retrieval algorithm was specifically developed for WindSat by Li et al. (2010), while Parinussa et al. (2012) adapted the LPRM to WindSat observations in order to obtain SSM retrievals that were consistent with the soil moisture products derived from AMSR-E (so potentially increasing the overall temporal resolution).

With the sensors shown so far, the SSM retrieval was obtained from brightness temperature measurements taken in C- and X-bands. More recently, the launch of the dedicated soil moisture missions SMOS and SMAP, in 2009 and 2015 respectively, marked significant improvements because of their more suitable observation frequency

(L-band at 1.4 GHz). At the same time, continuity of previously existing products is guaranteed by the follow-up of AMSR-E, i.e. the AMSR2 sensor, launched in 2012 on the Global Change Observation Mission for Water research (GCOM-W1) satellite.

2.7.2.1 SMOS

The Soil Moisture Ocean Salinity (SMOS) is the first dedicated mission intended for global scale soil moisture and ocean salinity retrievals (Kerr et al., 2001). SMOS is a sun-synchronous polar orbiting satellite, launched in 2009 by ESA, with a revisit period of 1–3 days and equatorial ascending/descending overpasses at 6:00 A.M./P.M. local solar time. SMOS carries a single payload, a novel two-dimensional L-band (1.4 GHz) interferometric radiometer (Microwave Imaging Radiometer with Aperture Synthesis, MIRAS), with dual polarization and multiangular (0° – 55°) viewing capabilities. The horizontal spatial resolution of the instrument is in the range of 30-50 km, while the usage of L-band wavelength results in an ability to retrieve soil moisture from greater depth (~5 cm).

The SMOS retrieval algorithm was described by Kerr et al. (2012). In order to account for the effects of canopy layer and soil roughness to land emissivity, the algorithm makes full use of the SMOS dual-polarised and multiangular brightness temperature acquisitions to retrieve soil moisture and vegetation optical depth, through the inversion of the L-band Microwave Emission of Biosphere (L-MEB) model (Wigneron et al., 2007). Soil moisture units are m^3m^{-3} , with a project target accuracy of $0.04 \text{ m}^3\text{m}^{-3}$.

SMOS Level 2 soil moisture products are generated in the SMOS Data Processing Ground Segment (DPGS) located at the European Space Astronomy Centre (ESAC), and distributed through the ESA SMOS Online Dissemination Service (<https://smos-diss.eo.esa.int/oads/access/>). The latest processor version of the SMOS retrieval algorithm is employed to generate reprocessed and operational (with a latency of 8-12 hours) SMOS Level 2 soil moisture (L2SM). L2SM contains the retrieved soil moisture, vegetation optical depth and other ancillary data derived during processing (surface temperature, roughness parameter, dielectric constant and brightness temperature retrieved at top of atmosphere and at surface) with their corresponding uncertainties. A near-real-time product is also generated, that meet the requirement of data provision within 3-4 hours from sensing. It uses the SMOS NRT brightness temperature product

as an input to a neural network that was trained on historical time series of L2SM data (Rodriguez-Fernandez et al., 2015). NRT product also provides an estimation of the soil moisture uncertainty and the probability that a soil moisture value is contaminated by Radio Frequency Interference. For both L2SM and NRT products, swath measurements are geo-located in 15 km ISEA 4H9 equal-area grid system.

SMOS Level 3 and 4 soil moisture products are also generated and distributed, by the Centre Aval de Traitement des Données SMOS (CATDS) (<http://www.catds.fr/>) and the Barcelona Expert Centre (BEC) (<http://bec.icm.csic.es/>).

Level 3 CATDS products firstly include daily global maps of soil moisture and its associated parameters (e.g. vegetation optical depth, dielectric constant), for ascending and descending orbits processed separately. The product is derived with a multi-orbit approach. The retrieval is done using three successive orbits within a seven-day moving window. When several retrievals are available for a given day, the best estimation of soil moisture is selected for each grid point. Then, some aggregated products are generated from these daily global maps, i.e. the 3-day, 10-day and monthly global map products of soil moisture and its associated parameters. The CATDS also hosts the Level 3 SMOS INRA-CESBIO (SMOS-IC) product. The SMOS-IC algorithm (Fernandez-Moran et al., 2017) was designed by INRA (Institut National de la Recherche Agronomique) and CESBIO (Centre d'Etudes Spatiales de la Biosphère). It is based on the two-parameter inversion of the L-MEB model, similarly to the L2SM retrieval approach but using some simplifications and with a reduced set of auxiliary information. The product contains daily global map of soil moisture and vegetation optical depth, for ascending and descending orbits. All Level 3 CATDS products are provided over the 25 km Equal-Area Scalable Earth Grid (EASE-Grid).

Level 3 BEC products firstly include daily global maps of soil moisture, vegetation optical depth and dielectric constant for ascending and descending orbits. The product is available on 15 km (ISEA 4H9) and 25 km (EASE) grids. The product on ISEA 4H9 grid is derived from the SMOS L2SM data without spatial or temporal averaging. The product on EASE-Grid is derived by quality-filtering, quality weighting and re-gridding the L2SM data. The 3-day, 9-day, monthly and annual products are generated by performing a temporal aggregation of the daily maps on the 25 km EASE-Grid for ascending and descending orbits.

Level 4 CATDS products consist in data obtained from SMOS Level 3 CATDS products combined with data from other sensors or models. Two Level 4 soil moisture products are currently available, both provided on the 25 km Ease-Grid. The first is a daily global map of root zone soil moisture at 0-1 m depth, for ascending and descending orbits, derived from the Level 3 CATDS soil moisture data and through the usage of surface temperature information from ECMWF model reanalysis. The second is a long term consistent dataset of daily global SSM maps, consisting of AMSR-E retrievals for the period 2003-2010 and SMOS Level 3 retrievals for the period 2010-2017. SSM retrievals from AMSR-E brightness temperatures are obtained through neural networks, trained on SMOS Level 3 data on the concurrent missions period (~1.5 years) and then applied to the past AMSR-E observations (Rodriguez-Fernandez et al., 2016).

Level 4 BEC products include a high resolution soil moisture dataset over Iberian Peninsula, with daily map at 1 km resolution for ascending and descending orbits. The product is derived from SMOS brightness temperature data and the use of land surface temperature and vegetation index obtained from models (ECMWF ERA-interim) and satellite observations (NASA Terra/Aqua MODIS).

2.7.2.2 SMAP

The Soil Moisture Active Passive (SMAP) satellite mission is the most recent dedicated to soil moisture monitoring (Entekhabi et al., 2010b). It was launched in 2015 by NASA, with the objective of generating global fields of soil moisture and landscape freeze/thaw state. The spacecraft is in a sun-synchronous near-polar orbit, with equator crossings at 6:00 A.M. (descending node) and 6:00 P.M. (ascending node) local time.

SMAP was designed with two components: an active L-band SAR (1.2 GHz with VV, HH, and HV polarizations) and a passive L-band radiometer (1.4 GHz with dual polarizations). The instruments share a rotating conically scanning antenna system with a surface incidence angle of 40°. The wide swath (1000 km) provides near-global revisit every 2–3 days. The radiometer provides accurate soil moisture data at moderate spatial resolutions (~40 km), while the SAR was designed to provide measurements with high spatial resolution (1-3 km over outer 70% of swath) but resulting more influenced by roughness and vegetation. The mission concept was to combine the attributes of the

radar (higher spatial resolution but lower soil moisture accuracy) and radiometer observations (higher soil moisture accuracy but lower spatial resolution): the concurrent coincident measurements were expected to provide soil moisture accuracy approaching radiometer-based retrievals but with intermediate spatial resolution (~10 km) approaching radar-based one. The target accuracy of SMAP mission is $0.04 \text{ m}^3\text{m}^{-3}$. Unfortunately, after 11 weeks of operation, the radar instrument encountered a failure, while the radiometer continues its operations.

The SMAP project generally aimed to produce three types of surface soil moisture products (also including ancillary data and quality-assessment flags), i.e. retrieved from passive, active, and active-passive measurements, respectively. They are based on the resampling of sensors measurements on the Equal-Area Scalable Earth Grid version 2.0 (EASE-Grid 2.0) at different spatial resolutions, i.e. 36 km for the passive, 3 km for the active, and 9 km for the active-passive soil moisture products. Level 2 products contain SSM retrievals based on half orbits (with a temporal resolution of 49 minutes), so only cells that are covered by the actual swath are written in the product; Level 3 products are daily global composites of the Level 2 soil moisture data; finally, a Level 4 product provides root-zone soil moisture based on the assimilation of SMAP observations into a land surface model.

The SMAP project consider several retrieval algorithms, with a designated baseline. For example, five retrieval algorithms were selected for brightness temperature measurements: the Single Channel Algorithm considering horizontally or vertically polarized data respectively (SCA-H and SCA-V), the Dual Channel Algorithm (DCA), the Microwave Polarization Ratio Algorithm (MPRA), and the Extended Dual Channel Algorithm (E-DCA). They are all based on the tau-omega model but differing in their approaches to solve for soil moisture. The SCA-H, designated as pre-launch baseline retrieval algorithm, was replaced by the SCA-V due to its better performances in preliminary Calibration/Validation (Cal/Val) analyses (Chan et al., 2016). However, all five algorithms are continuously assessed and the choice of the operational algorithm for the validated release of the product is evaluated on a regular basis as analyses of new observations and Cal/Val data become available, and algorithm parameters are tuned based on a longer SMAP radiometer brightness temperature time series record.

The baseline retrieval algorithm for SMAP radar measurements inverts a forward scattering model, through a multichannel approach based on a look-up table

representation (Kim et al. 2012; Kim et al. 2014). Finally, the active-passive algorithm (Das et al., 2014) is based on the disaggregation of the radiometer brightness temperatures using the radar backscatter spatial patterns within the radiometer footprint; once the disaggregated brightness temperatures are produced, the retrieval algorithm used for the radiometer-only soil moisture product is also here applied.

SMAP soil moisture products are made available publicly through NASA National Snow and Ice Data Center (NSIDC, <https://nsidc.org/daac/>). The following products, among those initially planned, are currently generated: Level 2 and Level 3 Passive (L2_SM_P and L3_SM_P) and Level 4 Surface and Root-Zone (L4_SM) Soil Moisture products. The SMAP L4_SM product is currently generated by assimilating SMAP brightness temperature observations into the NASA Catchment land surface model; the product provides estimates of surface (0-5 cm) and root-zone (0-100 cm) soil moisture at 9 km grid resolution. Then, other SMAP soil moisture products were introduced, i.e. the enhanced Level 2 and Level 3 passive (L2_SM_P_E and L3_SM_P_E) and the Level 2 SMAP/Sentinel active-passive (L2_SM_SP) products. The new products are developed post launch to recover capabilities lost when the SMAP radar ceased operation. In the enhanced Level 2 passive product, SSM retrievals are derived from brightness temperatures interpolated at 9 km grid resolution (Chan et al., 2018), while the enhanced Level 3 is the daily global composite of the enhanced Level 2 data. The SMAP/Sentinel active-passive product is based on the SMAP active-passive algorithm accommodated to use collocated Sentinel-1A and -1B data as backscatter fields for SMAP brightness temperature disaggregation and high-resolution soil moisture retrieval. The product is provided on global 3-km and 1- km EASE-Grid 2.0; the less frequent coverage of Sentinel-1 data results in more gaps for this match-up product than in the standard SMAP time series.

2.7.2.3 AMSR2

The Advanced Microwave Scanning Radiometer 2 (AMSR2), developed jointly by NASA and JAXA, was launched in 2012 onboard the JAXA's Global Change Observation Mission - Water satellite (GCOM-W1, also nicknamed "SHIZUKU"), as a continuation to AMSR-E (Imaoka et al., 2010). The general objective of the mission is still to observe the global water cycle variables: integrated water vapor, integrated cloud

liquid content, precipitation, sea surface temperature, sea surface wind speed, sea ice concentration, snow depth and soil moisture. The satellite has a revisit time of one day and is in a sun-synchronous orbit, with equator crossing time at 1:30 P.M. (local time on ascending node) and 1:30 A.M. (descending), to continue the AMRS-E observations. AMSR-2 succeeded most of the characteristics of AMSR-E: it has all the frequency channels and similar sensor configuration with an additional C-band (7.3 GHz) dual polarization frequency to improve the mitigation of the Radio Frequency Interference (RFI).

Several SSM products are available, generated with different retrieval algorithms, and specifically the JAXA standard algorithm, the Single Channel Algorithm (SCA), the Normalized Polarization Difference algorithm (NPD) and the Land Parameter Retrieval Model (LPRM). Due to RFI issues observed in the C-band, especially over the U.S., the soil moisture products are mainly derived using X-band brightness temperature data from the lowest frequency available, i.e. 10.65 GHz. However, LPRM products provide soil moisture retrievals also for the C-band, employing the frequency (6.9 or 7.3 GHz) less contaminated by RFI (Parinussa et al., 2015). In addition, 36.5 GHz data are also employed in some retrieval algorithms, e.g. to retrieve land surface temperature (Holmes et al., 2009). The footprint size at 10.65 GHz is 24 x 42 km, while for C-band it is 35 x 62 km.

JAXA generates SSM products using a modified version of its retrieval algorithm developed for AMSR-E, with vegetation fractional area as ancillary data (Koike et al., 2004; Fujii et al., 2009). The method utilizes the polarization ratio at 10.65 GHz and the normalized brightness temperature difference between the 36.5 and 10.65 GHz horizontal channels. JAXA Level 2 soil moisture product is distributed near-real-time in swath geometry at ~50 km spatial resolution, while Level 3 products are provided as daily and monthly temporal average grid data with 0.1° and 0.25° spatial resolutions in Equidistant Cylindrical Projection. JAXA standard products are currently distributed through its Globe Portal System (<https://gportal.jaxa.jp/gpr/index/index>).

NASA, using JAXA near-real-time brightness temperatures as input, provides several SSM products. Two products are distributed by NASA National Snow and Ice Data Center (NSIDC) (<https://nsidc.org/>), with input brightness temperatures being resampled to the global cylindrical EASE-Grid with a nominal grid spacing of 25 km. In the first product SSM is retrieved by SCA (Jackson, 1993), using horizontally polarized

brightness temperature observations at 10.65 GHz. In the second one, SSM is obtained with the NPD algorithm, that is an updated version of the NASA AMSR-E standard algorithm (Njoku & Chan, 2006), involving both polarizations at 10.65 GHz frequency. Then, NASA and VUA provide soil moisture products retrieved with LPRM (Owe et al., 2008) and distributed through the NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) (<https://disc.gsfc.nasa.gov/>). Here, LPRM uses the dual-polarized 6.9/7.3 or 10.65 GHz data for the retrieval of both surface soil moisture and vegetation water content, while the land surface temperature is derived separately from the vertically polarized 36.5 GHz channel. Level 2 datasets are provided in swath geometry, at 46 and 31 km spatial resolutions for C and X bands, respectively, as well as downscaled to 10 km by 10 km by using a smoothing filter-based intensity modulation technique. Level 3 gridded datasets are also distributed, at daily temporal resolution and 10 km and 25 km grid resolutions, with separate products for ascending (day-time) and descending (night-time) satellite passes, as the geophysical conditions are different and so the expected SSM retrievals characteristics.

2.7.3 ESA-CCI soil moisture products

ESA CCI SM (<https://www.esa-soilmoisture-cci.org/>) combines various single-sensor active and passive microwave soil moisture datasets into three harmonised global products: a merged ACTIVE, a merged PASSIVE, and a COMBINED that is obtained by merging the ACTIVE and PASSIVE ones. The origins of the ESA CCI SM products date back to the inclusion of soil moisture in the list of Essential Climate Variables (ECVs) identified by the Global Climate Observing System (GCOS), due to its critical role to the characterization of the climate system and its changes (Dorigo et al., 2015, 2017). A minimum record length of 30 years was among the required specifications for satellite-based soil moisture climate data record, and individual satellite missions are too short to fulfil it. Besides, differences in system and mission design as well as the use of different retrieval algorithms have led to varying quality and consistency over space and time. Yet, various studies have indicated the complementarity of active and passive products over different land cover types, with generally a better performance of radiometers over dry areas, a better performance obtained by scatterometers over more

densely vegetated areas, and a similar performance over semi-arid regions (Albergel et al., 2012; de Jeu et al., 2008; Dorigo et al., 2010; Taylor et al., 2012).

A first attempt to utilize the synergy between different active and passive microwave products and merge them into a single multi-decadal soil moisture dataset was made by Liu et al. (2011, 2012). The product, which was initially developed under ESA Water Cycle Multi-Mission Observation Strategy (WACMOS) project, is then being extended and improved within ESA Climate Change Initiative (CCI) (Wagner et al., 2012), with the first version of ESA CCI SM being released in 2012. Later versions of ESA CCI SM included a large number of enhancements, incorporated various new satellite sensors, and extended the products temporal coverage. In Dorigo et al. (2017) the version v03.2, covering the period 1978–2015, was presented, which is also used in the application reported in Chapter 6 and is therefore described here.

The architecture for the ESA CCI SM v03.2 production system (Fig. 2.7) is partly similar to the one originally proposed by Liu et al. (2011, 2012) and Wagner et al. (2012), with the main algorithmic improvement lying in the merging step.

In the original scheme (Liu et al., 2012), used for the previous ESA CCI SM releases, a simpler approach was applied to merge the several input time series. The ACTIVE and PASSIVE products were generated by choosing the (expected) highest quality observations in the intervals where more than one input product was available at pixel scale, e.g. the lower the measurement frequency, the more accurate soil moisture retrievals can be expected from passive sensors. Similarly, the merging of ACTIVE and PASSIVE into the COMBINED product was based on their expected relative performances with respect to the vegetation density. Over areas with a low vegetation the PASSIVE product was chosen, while over areas with moderate vegetation density the ACTIVE one was considered; in transition areas, both products were being used in a synergistic way: on time steps where only one of the products was available, the estimate of the respective product was used, while on days where both provided an estimate, their observations are averaged (Dorigo et al., 2015). The vegetation density classification was based on the vegetation optical depth values obtained from reference AMSR-E C-band observations. The original merging scheme was suboptimal from a merging perspective as it ignores the information contained in the retrievals that are not selected; moreover, it resulted in several product issues (Dorigo et al., 2017), in terms of data coverage and quality.

The merging scheme in ESA CCI SM v03.2 is instead based on the error characterization of the input datasets (Gruber et al., 2017), made through the triple collocation analysis (Stoffelen, 1998). The method involves the use of three collocated datasets with independent random errors (e.g. SSM estimates from a Land Surface Model and retrievals from active and passive sensors, respectively) in order to estimate the random error variance for each of them (see Sect. 4.6.2 for details on triple collocation analysis). In ESA CCI SM production system the triple collocation is applied to estimate the error variances of the individual input products, as well as of the ACTIVE and PASSIVE datasets, for each blending period separately. Surface soil moisture estimates from the GLDAS-Noah v1 LSM (Rodell et al., 2004) provide the third dataset. A weighted average of all available measurements is made to compute the merged soil moisture estimate in an optimal way (i.e. resulting in lowest output error variance), with weights inversely proportional to the random error variances of the input datasets. It is worthy to note that the triple collocation cannot be used to characterize the errors of the final COMBINED dataset, since after blending ACTIVE and PASSIVE an additional dataset with independent error structures would be required to complement the triplet. In this case, random error variances for COMBINED product are however computed through a classical error propagation scheme, i.e. by propagating the error variance estimates of inputs ACTIVE and PASSIVE through the merging scheme (error propagation approach is described in Sect. 4.6.3).

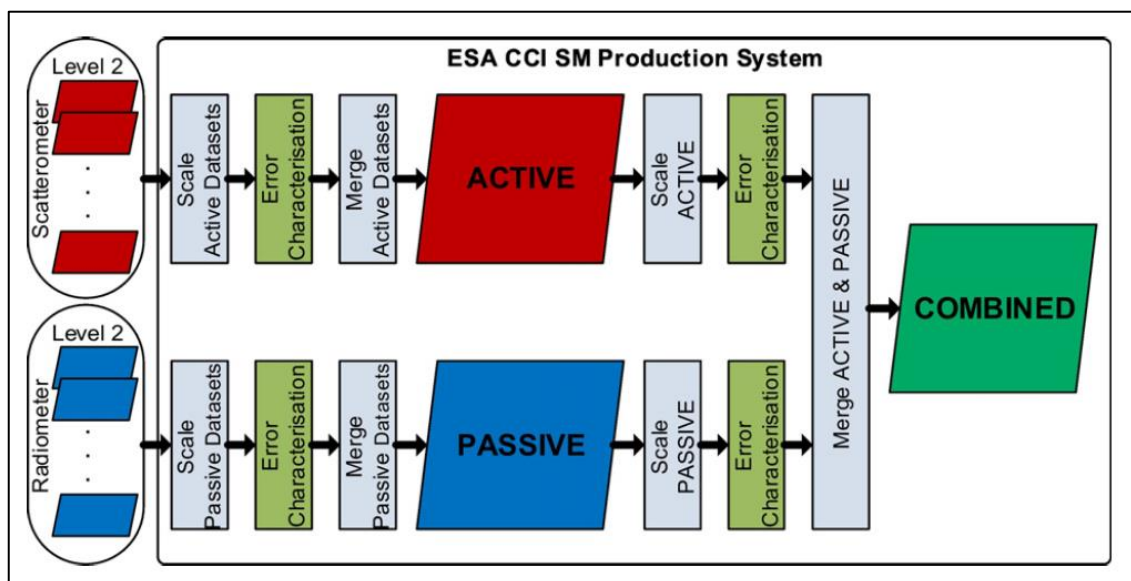


Figure 2.7. Figure taken from Dorigo et al. (2017), representing the schematic overview of ESA CCI SM v03.2 production system.

As showed in Fig. 2.7, the ESA CCI SM approach starts from publicly available Level 2 soil moisture retrieval data records, while the several merging steps are preceded by the rescaling and error characterization of the input products. Active microwave soil moisture products retrieved with the TU Wien change detection method and passive microwave products generated with the Land Parameter Retrieval Model are used because of their consistency in methodology across sensors. In v03.2 passive microwave observations come from SMMR, SSM/I, TMI, WindSat, AMSR-E, AMSR2 and SMOS, while active microwave observations derive from ERS-1/2 AMI wind scatterometers and MetOp-A and -B ASCAT. Retrievals based on synthetic aperture radars (SARs) yield higher spatial resolutions but at the expense of reduced revisit times and are therefore not considered appropriate for global climate data record production (Dorigo et al., 2017).

The input Level 2 SSM products undergo the following processing steps, which are described below:

1. Temporal Resampling
2. Spatial Resampling
3. Rescaling passive and active Level 2 observations into radiometer and scatterometer climatology, respectively
4. Error characterisation of rescaled passive and active Level 2 products
5. Merging rescaled passive and active time series into PASSIVE and ACTIVE products, respectively
6. Rescaling PASSIVE and ACTIVE products sets into common climatology
7. Error characterisation of PASSIVE and ACTIVE
8. Combining rescaled PASSIVE and ACTIVE products into single COMBINED product

ESA CCI SM datasets are provided on a regular grid with a spatial resolution of 0.25° in both latitude and longitude extension, and have a daily temporal resolution with time stamp centred at 0:00 UTC (however, the actual data availability varies in space and time due to the varying spatial and temporal availability of the single-sensor input products). For this reason, the first processing steps are the temporal and spatial resampling. All input observations are assigned to the closest daily 0:00 UTC reference time step and resampled to the 0.25° regular grid points (steps 1 and 2).

Then, on a pixel basis the temporally rebinned passive and active soil moisture datasets are scaled into a common radiometer and scatterometer climatology, respectively, by applying a propagating cumulative distribution function (CDF) matching approach (step 3). CDF-matching technique is described in Sect. 4.5.2; in this case, AMSR-E and ASCAT observations at pixel scale are used as ‘initial’ scaling references for the other passive and active datasets, respectively. The earlier datasets are successively matched to the references; in case the overlapping period between sensors is too short to allow for a direct scaling, e.g. for ERS with ASCAT, the datasets are matched based on their assumed similarity in seasonality (Liu et al., 2012).

The error characterisation of the rescaled passive and active input products was made by triple collocation (step 4), as previously described. Finally, the rescaled passive and active input datasets are respectively merged into a single radiometer- (PASSIVE) and scatterometer-based (ACTIVE) soil moisture products (step 5), according to the weighted blending scheme that takes into account the estimated input error variances. The units of measurement of ACTIVE is degree [%] of saturation while PASSIVE is provided in volumetric units [m^3m^{-3}], according to the input SSM retrievals.

Subsequently, the systematic differences between ACTIVE and PASSIVE products (including the different units) are corrected for by matching for the CDF of each pixel against long-term volumetric soil moisture estimates from a Land Surface Model, by using GLDAS-Noah v1 as reference (step 6). The choice of using a modelled soil moisture product and not one of the microwave-based products as scaling reference has been motivated by the fact that none of the latter has global coverage and spatially consistent quality; even though the rescaling procedure affects the absolute soil moisture values, temporal variability and trends of the original datasets are generally well preserved (Liu et al., 2012). Triple collocation is then performed on rescaled PASSIVE and ACTIVE products (step 7).

In the final step (8), the rescaled ACTIVE and PASSIVE products are merged into the COMBINED one, again based on their error characteristics. The units of measurement of COMBINED is volumetric units [m^3m^{-3}]. Error variance estimates for the COMBINED product are computed by error propagation.

3. DATA ASSIMILATION METHODS

3.1 Introduction

Data assimilation (DA) procedures are developed to combine complementary information from model simulations and independent observations to find the best representation of the dynamic behaviour of a system.

Data assimilation was originally used for meteorology and oceanography, while its application in hydrology is relatively recent; the development of remote sensing significantly enhance the application of DA.

In hydrology DA is widely applied in order to update the model state variables; it is also used to improve model parameter estimates or to update model output to match the latest observations, but these aspects will not be detailed here.

Data assimilation methods are designed to optimally merge state estimates from model and observations (both containing errors) considering the respective uncertainties, in order to minimize the error in analysis state estimates. The information contained in the observations is so also propagated forward or backward in time, depending on the specific application and therefore on the appropriately chosen DA method.

Many assimilation techniques have been developed, generally differing in their numerical cost, optimality, and suitability for given applications, with the sophistication of the merging algorithm that can vary widely.

In this chapter an overview of data assimilation methods is given, according to different classification criteria, and some commonly used techniques are introduced. Then the topic of soil moisture data assimilation is discussed, particularly with regard to the integration of remotely sensed observations into hydrological models in order to improve streamflow predictions. Finally, the mathematical approach is illustrated for the techniques used in most of the soil moisture DA applications.

3.2 An overview of data assimilation methods

The DA process can be generally divided into two phases: (1) a propagation step which evaluates changes in state estimates according to the dynamical model, and (2) an

analysis (or update) step which modifies state estimates to account for information contained in measurements (Evensen, 2009).

DA methods generally rely on the assumption of unbiased errors, i.e. errors must be strictly random and on average the model estimates and the observations must agree with the true fields (Reichle, 2008). The goal to combine model predictions and observations to determine an error variance minimizing estimate of the true state of the system (analysis) can be achieved by solving an appropriate least-squares minimization problem (Todini & Biondi, 2016).

Most assimilation algorithms are based on updated states that are a linear combination of the observation and the model estimates (Lahoz & Schneider, 2014). In other terms, analysis estimates are defined by corrections to the background (i.e. the a priori model estimates) which depend linearly on background-observation departures, weighted by a gain operator. The optimal gain operator corresponds to the BLUE (Best Linear Unbiased Estimator) solution. The BLUE analysis can be equivalently obtained as a solution to the variational optimization problem where a cost function is minimized, having two terms which account for the distance between the state vector and the background and observation vectors respectively (see Sect. 3.4).

The DA methods can be classified according to different criteria, starting from the way in which the analysis values are obtained.

In methods such as the Kalman Filter (KF) and its derivatives (Kalman, 1960; Evensen, 2009), the solution of BLUE problem is calculated using explicit linear algebra for the direct determination of the gain operator (also called Kalman gain matrix); this allows also to update the error covariance, then explicitly propagated in time to obtain its background value for the next assimilation step.

In large systems, like those in the present meteorological and oceanographical applications, the computational load associated with the dimension of the problem makes the use of numerical solutions preferable (Bouttier & Courtier, 1999). Methods such as the three-dimensional variational (3D-VAR) and four-dimensional variational (4D-VAR) were developed, with analysis solution numerically obtained by using iterative minimization techniques on the variational cost function. In 3D-VAR the 3 space dimensions are considered at the fixed time and model error covariance is considered to be static (so matrix is inverted just one time); in 4D-VAR the time is considered as the 4th dimension, with error covariance that implicitly evolves within an assimilation

interval (i.e. the analysis estimates are obtained without ever explicitly computing their error covariance).

Another distinction can be made concerning the assimilation interval in which the analysis is performed (Figure 3.1). In filtering problems, observations are assimilated sequentially, i.e. the analysis estimation is done for any instant a new observation is available, which is usually technically convenient. Examples are the KFs (detailed below) and the 3D-VAR.

In smoothing problems, observation distributed within a time window are considered, and the correction to the analysed state is smooth in time, which is physically more realistic. An example is the 4D-VAR, whose standard formulation assumes model is perfect (also known as strong-constraint 4D-VAR, i.e. the model physic is imposed as a strong-constraint in the minimization of the cost function), so it reduces to finding the initial condition that produces the trajectory that fits observations best throughout the whole assimilation window (Ide et al., 1997). Weak-constraint 4D-VAR methods are being developed to relax the assumption that the model is perfect (Reichle, 2008), by including an additional term in the cost function involving errors in the model formulation (in this case it is not enough to find an optimal initial condition, since it does not uniquely determine the trajectory).

When assimilation is based on filtering, there is propagation of the information contained in the observations only from the past into the future, i.e. after analysis step each observation influences the estimated states only at later times; in smoothing assimilation the system is instead updated in times that precede the most recent observation.

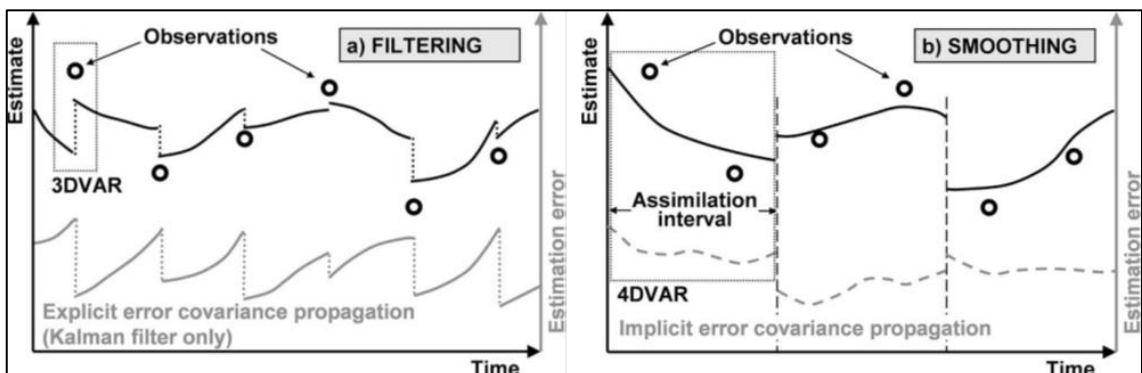


Figure 3.1. Figure taken from Reichle (2008), representing schematic of continually operating data assimilation systems based on (a) filtering (for example 3D-VAR, Kalman filter) and (b) smoothing (for example 4D-VAR).

In literature, the filtering and smoothing assimilation problems (McLaughlin, 2002; Reichle, 2008) are often referred respectively to as sequential and variational (implicit meaning through time) (e.g. Ide et al., 1997; Talagrand, 1997), or also as direct and dynamic observer (Walker & Houser, 2005) assimilation.

DA methods were developed to take also in account the characteristics of the dynamical model, which influences the propagation step.

Focusing on filtering problem (which is the most common case in hydrological DA), for linear systems the optimal sequential assimilation method is the Kalman filter (KF), originally proposed by Kalman (1960), which introduces an equation for the time evolution of the error covariance matrix. However, there are not many cases in which the linear KF could be applied to complex hydrological models (Sun et al., 2016).

The Extended KF (EKF) (e.g. Jazwinski, 1970) and the Ensemble KF (EnKF) (Evensen, 1994) were both developed to extend the application of the KF to nonlinear systems; also, they are the two major descendants of the linear KF.

The EKF applies a straightforward Taylor extension scheme to linearize the nonlinear system, but often becomes unstable when applied to complex nonlinear hydrological models (e.g. Reichle et al., 2002a). The EnKF avoids direct linearization by introducing an appropriate ensemble to represent the statistical properties of the model states with a Monte Carlo approach (essentially, the covariance matrix is replaced by the sample covariance). The EnKF avoids many of the problems associated with the EKF and is one of the most widely used hydrological DA methods (Moradkhani, 2008).

If the background and observation error pdfs are Gaussian (fully characterized by their error covariances, having assumed unbiased errors), the KF analysis scheme corresponds to the maximum likelihood estimator of the true state (the same applies to KF variants, within the limits of the approximations introduced) (Bouttier & Courtier, 1999; Evensen, 2009).

Other filtering techniques, such as Particle Filter (PF) (e.g. Moradkhani, 2008), avoid the restrictive assumption of normal distribution of errors in state variables, and the corresponding characterization in time of error covariance only. PF aims at providing a complete representation of the posterior probability distribution of state variables given the model and the measurements, by using a set of discrete random samples (called particles). Model estimates are not updated but rather their probability distributions are evolved through time. The prediction phase implies routing the samples through the

dynamical model; the probability distribution of model predictions is then calculated as weighted combination of the ensemble members, with associated weights that essentially are the normalised value of the pdf of the observation given the model estimates.

More generally, in DA algorithms there are other common assumptions on the errors statistical properties that are often violated in the practical application.

The assumption that observation and model errors are unbiased to the truth is the most restrictive assumption, most commonly violated assumption, and most detrimental assumption in terms of predictive performance (Walker & Houser, 2005). In practice there are often significant biases in the background fields and in the observations which is difficult to remove prior to data assimilation (Reichle, 2008). The common data assimilation systems designed to correct random, zero-mean errors are called bias-blind; it is not possible to produce an unbiased analysis from a biased background and/or biased observations with a bias-blind analysis method (Dee, 2005). Bias-aware assimilation methods instead incorporate specific assumptions about the source and nature of (some of) the biases in the system, characterized in terms of some well-defined set of parameters, in order to estimate and correct them during the analysis step (Dee & da Silva, 1998). The first challenge for these methods is to correctly attribute a detected bias to its source; then they require the formulation of a useful model for the bias, as well as a reference dataset from which to estimate the bias model parameters (both requirements involve difficult choices). In practice, however, it is extremely difficult, if not impossible to attribute the bias conclusively to either the model or the observations, and subjective assumptions need to be made. If the source of an evident bias is uncertain, bias-blind assimilation may be the safest option (Dee, 2005).

Another common assumption is in model and observation random errors to be uncorrelated with the true state, mutually uncorrelated and uncorrelated in time (white noise) (e.g. Reichle et al, 2002a, 2002b). The hypothesis of mutually uncorrelated errors is the most justifiable because the causes of errors in the background and in the observations are supposed to be completely independent. However, one must be careful about observation preprocessing practices that use the background field in a way that biases the observations toward the background: it might reduce the apparent background departures, but it will cause the analysis to be suboptimal.

3.3 Data assimilation of soil moisture

Effective quantification and reduction of uncertainties are fundamental for a comprehensive and reliable forecasting of hydrological variables, such as river discharge (Li et al., 2016). Rainfall-runoff models are used to predict streamflow at future time given knowledge of the current state of the system; in this respect, good estimates of the state variables are necessary to enable the model to produce relatively accurate and reliable flow forecasts (Liu et al., 2012b). The application of data assimilation in hydrology holds considerable potential for improving prediction accuracy and quantifying uncertainty, also as a consequence of the abundance of new hydrologic observations (in-situ or remotely sensed) (Montzka, 2013). In recent years, increasing availability of satellite observations with appropriate spatial and temporal resolutions has brought great interest into assimilating remotely sensed retrievals of various quantities, such as soil moisture (Ni-Meister, 2008). Due to the important role of soil moisture in hydrologic, meteorological and ecological cycles, soil moisture data assimilation has received increasing attention; drought monitoring, runoff modelling and flood forecasting, numerical weather prediction, land surface and climate models assessment, agricultural monitoring and crop yield forecast are among the most important applications benefiting from assimilation of soil moisture observations (Brocca et al., 2017b). In-situ techniques provide fairly accurate soil moisture measurements at the point-scale for a suitable depth; however potential benefits in assimilating this kind of observations are restricted due to the very limited spatial coverage coupled with the very small spatial support. In a recent work by Gruber et al. (2018), in situ soil moisture measurements from a large-scale monitoring network are assimilated into a continuous model domain, with results demonstrating that a fourfold increase in existing station density is needed to provide a level of skill comparable to that provided by existing satellite-based surface soil moisture retrievals. In fact, most DA applications employ remotely sensed soil moisture observations at the surface layer (1–5 cm). However, information on the moisture condition in the root zone and subsurface layers is more critical for understanding and simulating many hydrologic processes including evapotranspiration and surface runoff (Han et al., 2012; Vereecken et al., 2008). By the utilization of satellite SSM observations in a DA framework containing a hydrological or land surface model, it is possible not only to update SSM

model states but also to improve the prediction of soil moisture profiles. An important prerequisite is that the water content of the near-surface soil layer be coupled with the deep soil layer (Brocca et al., 2012a; Flores et al., 2012; Walker et al., 2002), otherwise the soil moisture profile cannot be retrieved. A decoupling can for example be observed during extended drying periods when a divergence between the drying rates at the soil surface and deeper levels occurs; moreover, frozen soil conditions manifest in the data as low soil moisture observations, which can have a significant detrimental impact on the assimilation and need to be excluded (Draper et al., 2011). There are basically two different options to assimilate remotely sensed soil moisture observations: (1) direct assimilation of (not always available) high-level data products, i.e. SSM (eventually preprocessed), and (2) direct assimilation of raw or low-level sensor data (brightness temperatures or backscatter) (Montzka, 2013). The advantage of direct assimilation of high-level products is that it is not necessary to implement an additional SSM retrieval model or a complex observation operator to internally calculate soil moisture (Reichle, 2008).

Many applications rely on assimilating SSM observations into land surface models to improve soil moisture simulations. While the impact of SSM DA for the estimation of soil moisture profile is reported to be generally positive, it still remains controversial if it introduces added value for further hydrological applications such as the runoff prediction (Montzka, 2013; Brocca et al., 2017b).

As soil moisture is a key control variable in runoff modelling, it is a logical step to update soil moisture states into the models through satellite observations assimilation. Addressing uncertainties in state variables can improve streamflow simulations, also by enhancing the estimation of model initial conditions for flood forecasting.

However, as most operational models are highly conceptualized, it is not promised that addressing errors in soil moisture states is the most beneficial approach to maximizing the benefit in streamflow forecasts (Li et al., 2016). Different from streamflow data assimilation, soil moisture affects the flow forecasts through the catchment hydrologic processes, which are typically not fully represented by forecasting models. Therefore, improvement in soil moisture estimates does not necessarily lead to improvement in streamflow predictions.

The analysis of the scientific literature reveals that the actual added-value in using satellite soil moisture for runoff modelling is still unclear (e.g. Matgen et al., 2012;

Massari et al., 2015). Some authors obtained moderate to significant improvement through the assimilation of satellite data soil moisture in hydrological modelling, while in other studies degradation of performances was reported.

Brocca et al. (2012a) showed that the assimilation of satellite-derived root-zone soil moisture has a stronger impact than direct assimilation of SSM products on streamflow prediction. Draper et al. (2011) found that the assimilation of SSM data improved the streamflow prediction to some extent, but the improvement may mainly result from the correction of large bias from precipitation, which was suggested to be addressed through bias-aware data assimilation approaches. Han et al. (2012) found that the improvements in streamflow were much weaker than in soil moisture, and not consistent in all sub-areas. Matgen et al. (2012) showed that satellite-derived root-zone soil moisture was not able to enhance runoff prediction if the model is well calibrated by streamflow gauges. Chen et al. (2011) obtained in a synthetic experiment moderate results in evapotranspiration estimation, while deeper soil moisture, surface runoff, as well as stream flow predictions were not improved; the results of the real-world experiment were also not successful, caused by significant underprediction of the vertical soil water coupling. Alvarez-Garreton et al. (2014, 2015) applied satellite soil moisture assimilation for both lumped and semi-distributed models, but runoff prediction was not strongly improved; when assimilation were performed in distributed models (Wanders et al., 2014; Lievens et al., 2015), improvements were identified especially in ungauged areas. Ridler et al. (2014) showed that assimilation of SSM products resulted in overcorrection of errors in streamflow and flood peaks, which may be caused by the bias between observations and model estimates.

These contrasting results have to be attributed to the inherent uncertainties and issues involved in the use of satellite soil moisture data in hydrological modelling, including the bias between remote sensing data and model states, the discrepancy between soil moisture depth in the model and the depth observed by remote sensing data, the spatial mismatch between the model and the remote sensing products, the particular model structure, the assessment of the magnitude and the structure of the errors in the hydrological model and in the observations (Brocca et al., 2017b). These issues are still under-research and need to be further addressed before remote sensing data can be operationally implemented to constrain the flood forecasting models. It is important to understand whether and how much benefit can be achieved when a single issue is

addressed. There will be a greater potential to improve flood forecasting when better understanding and more effective solutions to those specific challenges are achieved (Massari et al., 2015).

Even the choice of the most appropriate DA technique can have a significant impact on final results. DA algorithms are becoming increasingly sophisticated, from simple rule-based, direct insertion methods (i.e. the forecast model states are directly replaced with the observations) to advanced smoothing and sequential techniques as well as the various variants of these techniques (Li et al., 2016).

After pioneering works by Goodrich et al. (1994) and Otlé & Vidal-Madjar (1994), who simply integrated remotely sensed SM retrievals into a hydrological model through direct insertion, several other studies followed (Houser et al., 1998; Galantowicz et al., 1999; Li & Islam, 1999; Margulis et al., 2002; Montaldo et al., 2001; Pauwels et al., 2002; Walker et al., 2002), not only focused on runoff modelling. This stage was mainly characterized by the use of data such as the ERS/SAR measurements, with low temporal frequency (i.e. about 35 days) resulting in an insufficient impacts on flow simulation. At that period, various approaches were examined in order to identify what type of data assimilation methods are most suitable for satellite SM assimilation, including direct insertion, statistical correction, successive correction, Newtonian nudging (Houser et al., 1998; Walker & Houser, 2005).

Along with the development of high temporal resolution satellites (with revisits of 1–3 days), it was gradually realized that Kalman filtering approaches are relatively strong in both uncertainty reduction ability and computational efficiency (Li et al., 2016).

Although traditionally dominant in numerical weather forecasts, variational methods have not been widely used in hydrological soil moisture assimilation (Sun et al., 2016), and there are few studies available in this regard (e.g. Reichle et al., 2001; Jones et al., 2004; Yang et al., 2007; Calvet & Noilhan, 2000; Sabater et al., 2007).

The majority of studies chose the Ensemble Kalman Filter (EnKF) (e.g. Reichle et al., 2002b; Alvarez-Garreton, 2015; Brocca et al., 2012a; Chen et al., 2014; Han et al., 2014; Lievens et al., 2015; Loizu et al., 2018), while a minor part used the Extended Kalman Filter (EKF) (e.g. Draper et al., 2011), and also simple nudging approaches (e.g., Brocca et al., 2010a) where the expected weights of model and observation are assumed constant in time. The Particle Filter (PF) could be a strong alternative to EnKF, which is subject to some limitations including the linear updating rule and assumption

of jointly normal distribution of errors in state variables and observation (Moradkhani, 2008; Matgen et al., 2012; Montzka, 2013). However, the PF has not been widely implemented in flood forecasting studies. One reason is that the PF can lead to ensemble collapse and needs to be resampled for ensemble forecasting, while the EnKF is more efficient to meet the operational requirement of low computational cost (Li et al., 2016).

3.4 System formulation and analysis scheme

Considering two unbiased estimates related to the true state vector x^t at a given time, defined by:

$$x^b = x^t + \eta^b \quad (3.1)$$

$$y = Hx^t + \varepsilon \quad (3.2)$$

where:

- x^b is the background model state vector;
- η^b is the (unknown) background model state error, with zero-mean and (known) covariance matrix P^b ;
- y is the observations vector (there are usually fewer observations than variables in the model);
- H is the ‘observation operator’ (here considered linear as in the Kalman filter theory) that relates states to observations;
- ε is the (unknown) observation error, with zero-mean and (known) covariance matrix R ;
- model and observation errors are mutually uncorrelated.

The Best Linear Unbiased Estimator (BLUE) of the true state, that provides x^a , the analysis estimate of x^t :

$$x^a = x^t + \eta^a \quad (3.3)$$

by requiring that the total analysis error variances are minimum, is determined by the following equation:

$$x^a = x^b + G(y - Hx^b) \quad (3.4)$$

where the linear operator G is the gain matrix (also called Kalman gain matrix), given by:

$$G = P^b H^T (H P^b H^T + R)^{-1} \quad (3.5)$$

The error covariance of the analysed model state vector, P^a , is reduced with respect to the error covariance of the background state as:

$$P^a = (I - GH)P^b \quad (3.6)$$

The BLUE analysis is equivalently obtained as a solution to the variational optimization problem:

$$x^a = \operatorname{argmin}[J(x)] \quad (3.7)$$

with:

$$J(x) = \frac{1}{2}(x - x^b)^T(P^b)^{-1}(x - x^b) + \frac{1}{2}(y - Hx)^T R^{-1}(y - Hx) \quad (3.8)$$

where J is called the variational cost function and measures, in a weighted sense, the distance between an estimate x and the forecast x^b , plus the distance between the estimate and the observations y .

The analysis problem can be also formalized using the conditional probabilities, i.e. by looking for the maximum of the conditional probability of the model state given the observations. If the background and observation error pdfs are Gaussian, then the error variance minimizing estimate x^a is also the maximum likelihood estimate of the true state.

The observation operator, representing functions from model state space to observation space such as interpolation operators from the model discretization to the observation points, or conversions from model variables to the observed parameters, in practice may not be linear, i.e. $H(x)$ and not H . $H(x)$ can be used to the predicted measurement from the predicted state in (3.4), but cannot be applied directly in (3.5) to derive the gain operator. It usually makes physical sense to linearize the observation operator $H(x)$ in the vicinity of the background state (Bouttier & Courtier, 1999), introducing the tangent linear operator H , i.e. assuming that the variations of the observation operator in the vicinity of the background state are linear: for any x close enough to x^b , then $H(x) - H(x^b) \approx H(x - x^b)$. The effects of approximations in the observation operator may in any case be represented in the observation error covariance; the observation operator indeed generates the values that the observations would take in the absence of any error, according to (3.2).

3.5 Sequential data assimilation

Sequential data assimilation methods use the analysis scheme previously described to update the model state when new observations become available. A major issue concerns the time evolution of the forecast error covariance P^b , according to the assigned dynamical model.

3.5.1 Linear dynamics: the Kalman Filter

For linear dynamics the optimal sequential assimilation method is the Kalman filter (KF), originally proposed by Kalman (1960), which introduces an equation to predict error statistics for the model background.

Given a best estimate for x^t at time t_{k-1} , x_{k-1}^a , a forecast x^b is calculated at time t_k , as:

$$x_k^b = Mx_{k-1}^a \quad (3.9)$$

where M is a linear model operator, including the model parameters and forcings.

It is assumed that the true state vector x^t evolves according to:

$$x_k^t = Mx_{k-1}^t + q_{k-1} \quad (3.10)$$

where q_{k-1} is the (unknown) model error over one time step, with model error covariance matrix Q_{k-1} . The unbiased model error summarizes all the uncertainties in model formulation, parameters, forcing data, initial conditions.

Subtracting (3.9) to (3.10) results in:

$$x_k^t - x_k^b = M(x_{k-1}^t - x_{k-1}^a) + q_{k-1} \quad (3.11)$$

The error covariance matrix for the background at time t_k becomes:

$$P_k^b = MP_{k-1}^a M^T + Q_{k-1} \quad (3.12)$$

At times when new observations are available:

$$y_k = Hx_k^t + \varepsilon_k \quad (3.13)$$

it is thus possible to calculate the gain matrix with the (3.4) and then update the state and the error covariance using (3.5) and (3.6), respectively.

3.5.2 Nonlinear dynamics

Considering now a nonlinear model:

$$x_k^t = \mathcal{M}(x_{k-1}^t) + q_{k-1} \quad (3.14)$$

where $\mathcal{M}(x)$ is a nonlinear model operator and q is again the unknown model error over one time step.

For nonlinear dynamics the Extended Kalman filter (EKF) may be applied, in which an approximate linearized equation is used for the prediction of error statistics.

A prerequisite to apply the EKF is that the nonlinear model should be continuously differentiable. Under this condition, it is possible to obtain the converted linear dynamic matrix by expanding the nonlinear functions at the estimated point.

The tangent linear operator is applied:

$$\mathcal{M}'_{k-1} = \left. \frac{\partial \mathcal{M}(x)}{\partial x} \right|_{x_{k-1}} \quad (3.15)$$

assuming that the contribution from higher order terms are negligible.

Similarly to (3.12), the following equation is then used for error covariance propagation:

$$P_k^b = \mathcal{M}'_{k-1} P_{k-1}^a \mathcal{M}'_{k-1}^T + Q_{k-1} \quad (3.16)$$

which represents an approximate solution due to the linearization and the discarding higher order terms (closure approximation). Thus, the usefulness of the EKF will depend on the properties of the model dynamics, with approximations that could be too severe for strongly nonlinear dynamics (Moradkhani, 2008).

In many cases, nonlinear systems do not have explicit analytical solutions and the derivatives can only be calculated numerically (Sun et al., 2016).

The Ensemble Kalman Filter (EnKF) was originally proposed as a stochastic or Monte Carlo alternative to the deterministic EKF by Evensen (1994).

The EnKF is based on representation of the state error covariance needed at analysis times through the statistical properties of an appropriate ensemble of model states, randomly generated and propagated through the model operator independently.

The EnKF has gained popularity because of its simple conceptual formulation and relative ease of implementation for practical applications (Evensen, 2003). It avoids the need to linearize the model equations for the propagation of the error covariance, so there are no derivation of a tangent linear operator nor closure approximation used; moreover it requires no storing a full error covariance matrix nor its propagation in time, making the computational demand feasible also for high-dimensional state vectors. A critical aspect is the size of the ensemble, which determines the difference

between the ensemble covariance and the error covariance. However, in many cases a limited ensemble size provides a sufficient representation (Sun et al., 2016).

3.5.2.1 Ensemble Kalman Filter

The EnKF is based on the representation of error statistics at a given time by using an ensemble of model states.

While the error covariance matrices for the background and the analysed estimate are in the Kalman filter defined in terms of the true state as:

$$P^b = \overline{(x^b - x^t)(x^b - x^t)^T} \quad (3.17)$$

$$P^a = \overline{(x^a - x^t)(x^a - x^t)^T} \quad (3.18)$$

where the overbar denotes an expectation value, in EnKF since the true state is not known, it is more convenient to consider ensemble covariance matrices around the ensemble mean \bar{x} :

$$P_e^b = \overline{(x^b - \bar{x}^b)(x^b - \bar{x}^b)^T} \quad (3.19)$$

$$P_e^a = \overline{(x^a - \bar{x}^a)(x^a - \bar{x}^a)^T} \quad (3.20)$$

where now the overbar denotes an average over the ensemble.

This leads to an interpretation of the EnKF as a purely statistical Monte Carlo method where the ensemble of model states evolves in state space with the mean as the best estimate and the spreading of the ensemble as the error covariance.

Therefore the EnKF does not need to store a full error covariance matrix, nor its propagation is necessary for the prediction of error statistics, as the time evolution of the probability density of the model state is represented by integrating ensemble members forward in time according to the model dynamics till the next observation time.

At analysis times, observations are treated as random variables (Burgers et al., 1998) by adding random perturbations drawn from a distribution with zero mean and covariance equal to the measurement error covariance matrix:

$$y_j = y + \varepsilon_j \quad (3.21)$$

where j counts from 1 to the number of model state ensemble members N . In this way an ensemble of observations is generated, from a distribution with mean equal to the actual measurements and covariance equal to R , that then is used in updating the ensemble of model states.

The Kalman gain matrix is calculated as in (3.5) apart from the use of P_e^b instead of P^b :

$$G_e = P_e^b H^T (H P_e^b H^T + R)^{-1} \quad (3.22)$$

A new ensemble representing the analysed state is then generated by updating each ensemble member individually using the traditional analysis equation:

$$x_j^a = x_j^b + G_e (y_j - H x_j^b) \quad (3.23)$$

that implies:

$$\bar{x}^a = \bar{x}^b + G_e (y - H \bar{x}^b) \quad (3.24)$$

As an effect of the use of perturbed observations, the analysed error covariance is:

$$P_e^a = (I - G_e H) P_e^b \quad (3.25)$$

resulting to be consistent with (3.6).

The general idea of the EnKF is illustrated in Figure 3.2.

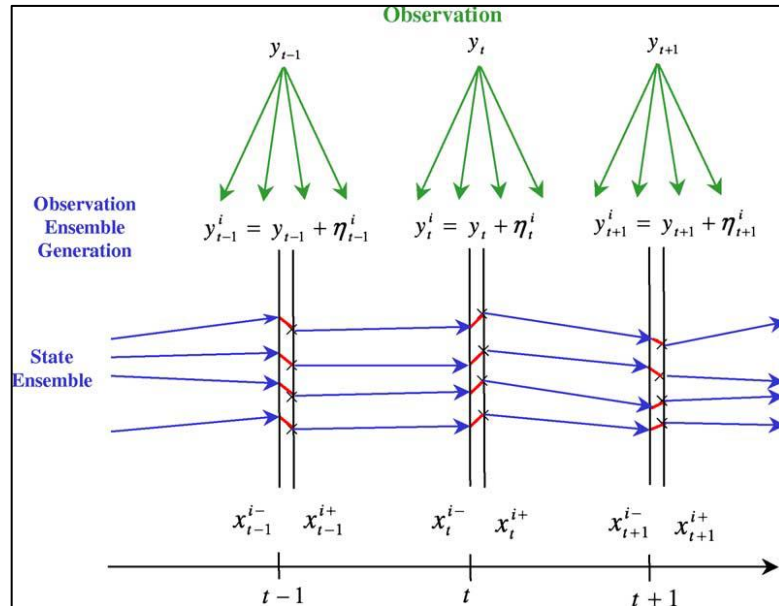


Figure 3.2. Figure taken from Moradkhani et al. (2005), representing EnKF schematic. x_t^{i-} : forecasted state ensemble member and x_t^{i+} : updated state ensemble member.

4. OBSERVATION PREPROCESSING AND ERROR CHARACTERIZATION

4.1 Introduction

Some operations on surface soil moisture retrievals are typically necessary for the purpose of an effective data assimilation, in order to consistently compare the satellite-based data with the model estimates, as well as to obtain optimal analysis solutions.

As showed in the previous chapter, an observation operator, H , is introduced that maps the model variables into observation space, which then must be in a linear form in order to calculate the gain matrix. DA provides optimal solutions according to underlying hypothesis on the observation error structure: observations are assumed to be unbiased with additive zero-mean random errors, temporally-uncorrelated and mutually uncorrelated with the analogous model errors. The use of the actual observation error variance guarantees the optimal weighting of model backgrounds and observations.

In this sense, the preprocessing and error characterization of satellite-based soil moisture observations are crucial for the successful implementation of data assimilation systems, and represent a challenge also in terms of compatibility with the underlying hypothesis of the common DA techniques.

Following a common practice in soil moisture DA, an ‘inverse’ observation operator can be used to process the satellite observations into model variables, with these ‘preprocessed’ observations being subsequently considered in DA analysis scheme. During DA, the correspondent observation operator thus assumes the form of a matrix with elements being equal to 0 or 1 (depending on whether the observation is convertible or not in state vector elements) and can be used in order to compute gain matrix. More generally, the preprocessing is addressed to face up to a number of known weaknesses which can affect the optimality of the DA, including the masking of low quality observations and the reduction of relative biases between model and satellite-based estimates. The preprocessed observations are then analysed in the error characterization step, aimed to assess their actual random error variance prior to DA.

All these tasks can be performed by choosing among different methodologies, and final choices can determine not only the optimality of DA experiments but even the ability to improve the background model estimates. In the following, some important issues

involved in remotely sensed soil moisture DA process are summarized, with a focus to those concerning the satellite observations and to the corresponding commonly used solving strategies. Then, for the identified steps (i.e. quality check, propagation to deeper layers, bias correction, error characterization) the most used approaches are described.

4.2 Overview of steps prior to data assimilation

As discussed in the previous chapter, the relative improvement of assimilating remotely sensed soil moisture into flood forecasting models is yet difficult to be quantified (e.g. Matgen et al., 2012), and much more research is still needed to understand the benefits and limitations in streamflow prediction by integrating satellite soil moisture data. The contrasting results obtained in various studies also revealed a number of scientific and practical issues involved in the use of satellite soil moisture data in hydrological modelling that should be addressed. A big challenge is concerned with the “mapping” between observed and modelled variables, as their characteristics rarely coincide. In this sense issues include the bias between remote sensing data and model states, the discrepancy between soil moisture depth in the model and the shallow layer observed by remote sensors, and also the spatial mismatch between observations and modelled data. Then, other critical aspects are related to the very low quality of soil moisture retrievals under certain surface conditions (e.g. dense vegetation, frozen soils, snow, urban areas, open water), the particular model structure, and the assessment of the errors magnitude and structure in the hydrological model and in the observations.

While these limitations may prevent the use of remotely sensed soil moisture in practical applications, some approaches have been already developed to overcome these issues. In this sense, several choices can be made in a data assimilation study and they might have a significant impact on final results, with even the same relevance of the considered observational dataset or data assimilation algorithm (Massari et al., 2015).

With regards to the model structure, it is worthy to highlight the development of rainfall-runoff models specifically targeted to the assimilation of satellite data, e.g. by considering a thin soil layer close to the surface. The assimilation of soil moisture data is also useful to identify and overcome structural deficiencies in the hydrologic models themselves. For example, in Chen et al. (2011) the calibrated SWAT model

significantly underestimated the vertical coupling of soil moisture between upper and lower soil layers, and the SSM assimilation was not effective in improving estimates of deep soil moisture and then streamflow. The particular challenge of correctly representing linkages between soil moisture across two or more soil layers has been identified as a key concern.

DA methodologies even rely on a proper representation of model errors, based for example on the statistical descriptions of errors in model parameters and forcings. However, model error characterization is often made in a somewhat arbitrary or subjective fashion, depending on uncertain, user-defined error assumptions; as a result, the relative weighting applied to modelled and observed soil moisture information is arguably subjective and does not necessarily reflect an optimized integration of independent data sources (e.g. Crow & Van Loon, 2006). Efforts have been made to overcome this weakness; for example, in EnKF applications several ensemble verification metrics can be used to assess the ensemble representation of SM while considering model uncertainty, also in order to verify or calibrate error assumptions on parameters and forcings.

The remaining of this chapter is focused on the issues related to preprocessing (i.e. quality check, propagation to deeper layers, bias correction) and error characterization of satellite-based soil moisture observations, aimed at their subsequent assimilation in prediction models.

As already mentioned, remotely sensed observations can have different levels of quality, due to various factors involved in the processing that can affect the accuracy of retrievals. Thus, before being used, observations are generally subjected to a quality control aimed to reject particularly poor measurements. Since no observation is free of error, the challenge is to mask only those observations that are below acceptable quality thresholds while providing reliable error estimates for the remainder. SSM products usually include several additional data fields, containing attributes and flags which can be useful in judging the reliability of remotely sensed estimates. Using these indicators, quality check procedures can be set for specific applications. For example, sets of threshold values are used to check SSM values preliminarily to their assimilation in land surface models into numerical weather prediction systems (e.g. Dharssi et al., 2011; de Rosnay et al., 2013), in order to detect areas where retrieval methods do not work well or to mask single spurious SSM observations.

Another critical issue is related to how propagate information from the surface layer observed by satellite sensors to a deeper layer. The sensing depth of satellite sensors is very shallow, around 2 and 5 cm for the used C-band and L-band sensors, respectively, and also varying with wetness conditions. These depths are much lower than that generally considered by models which simulate soil moisture states for various applications. When sensor and model are not representing the same geophysical variable (i.e. soil moisture states at a given depth), a mapping becomes necessary before the comparison. The task to propagate remotely sensed surface soil moisture information to the deeper layers was targeted in many ways, as several application fields require knowledge of soil moisture in the so-called root-zone. Contributions to the estimation of root-zone soil moisture from surface measurements include regressive relationships (e.g. Arya et al., 1983; Srivastava et al., 1997), filtering approaches (Brocca et al., 2013) such as the exponential filter (Wagner et al., 1999), physically-based formulations (Manfreda et al., 2014), data driven methods (e.g. Kornelsen & Coulibaly, 2014; Pal et al., 2016). These approaches can potentially be used to convert observations to model state space, before data assimilation. However, by modifying the structure of the models including a shallow soil layer close to the surface, satellite soil moisture information can be extended to the root zone in a data assimilation framework, with surface observations directly integrated that also update the root-zone soil moisture states. In all cases, if surface and root-zone soil moisture levels are decoupled, as usually occurs in very dry conditions, the use of satellite measurements can be of limited utility. A simplified method such as the exponential filter was also used when prediction model already includes a surface soil layer (Fairbairn et al., 2017; Massari et al., 2018); this because, in order to avoid numerical problems, the assigned model surface layers usually have a minimum thickness (typically 10 cm) that is greater than satellite sensing depth (Manfreda et al., 2014).

Even if the model has a surface layer, or the remote sensing data have been converted to the root-zone SM estimates, there may still be significant bias between modelled and remotely sensed soil moisture. Following typical SM DA practices, these systematic differences are corrected prior to data assimilation, in order to compare consistent information; this facilitates data assimilation corrections for random errors, otherwise analysis estimates will be affected by biases between the compared datasets. It is worthy to remember that DA methods usually rely on observations and model estimates that are

unbiased with respect to the unknown “truth”; thus, the proper treatment of bias is critically important to the success of a data assimilation system. As mentioned in the previous chapter, modelling the bias respect to truth is a difficult task, and in practice it is not possible to attribute the bias to either the model or the observations, as both are likely to be affected by systematic errors. Some types of systematic differences between model and observation are usually removed, making subjective assumptions with the risks to lose some information in the bias-corrected datasets. The common approach is to rescale the observations in order to match some statistical properties of the model estimates. These so-called a priori rescaling methods are easy to implement as a preprocessing step to the data assimilation system, and different techniques are available, generally requiring a relatively long-term of record to properly derive the rescaling statistics.

Finally, another challenge is the error characterization of remote sensing observations, which is a prerequisite for DA analysis. Error is defined as the difference between the measured (or estimated by a model) value and the corresponding (unknown) true value. In general, all estimates of a variable are considered to be affected by errors, both those obtained by model and by measurement (in situ or remotely sensed). Error can be viewed as having two components, a random and a systematic one. As already seen, proper corrections for systematic errors could be made with reliable bias modelling, alternatively some types of systematic differences with a reference dataset can be effectively removed. Random errors reduction is instead the main target of data assimilation methods, which optimally combine model and observation information. The success of data assimilation also relies on an adequate representation of observation errors, while on the contrary an incorrect error characterization can seriously degrade the prediction model performance. Because of a general lack of information on the errors magnitude and structure, usually simplistic assumptions are made, with DA results being related to the reliability of the errors assumption/estimation. Several error characterization methods are available, generally not only limited at quantifying the observation random error variance which is usually needed for DA applications. Different methods are based on different error assumptions and provides different information. In the following it will be described how the error characterization is carried out by the use of standard statistical measurements (like the correlation

coefficient or the root-mean-square-error) and by the triple collocation and the error propagation methods.

4.3 Quality check

Data assimilation techniques can fail in estimating the true fields if observations grossly inconsistent with the model state are used (Waller et al., 2018). The presence of measurements affected by gross errors (or more generally by errors that can be considered characterized by different statistics compared to the rest of observational dataset) can be due to various causes. With regards to remotely sensed soil moisture, the quality of individual observations is impacted on one hand by factors related to the sensor properties and operation, as well as to the retrieval algorithm skill. On the other hand, the quality can be also affected by many environmental conditions, which are variable through space (e.g., topography), time, or both (e.g., frozen soil conditions, vegetation cover). Some factors may entirely impede a realistic retrieval (e.g., snow/ice coverage) while the majority adds some degree of random error and bias to the obtained estimate, the amount of which depends on the nature, intensity, and subpixel area affected by the specific factor (e.g., vegetation, open water) (Dorigo et al., 2017).

To make the measurements useful to users, information on data quality are included in many satellite-based soil moisture datasets. Data quality information are provided as additional data fields, in the form of flags (referred to some specific environmental condition or retrieval process aspect) or also quantitative uncertainty measures. According to Wagner et al. (2013b), these indicators can be grouped into ‘quality flags’ and ‘advisory flags’ with respect to their functionality. The quality flags are directly derived from the data processing and describe the intrinsic quality of the soil moisture retrieval (e.g., noise, processing, and correction flags in ASCAT SSM). On the other side, the advisory flags support the user in judging the validity of the soil moisture product in those situations that could not have been identified during the retrieval (e.g., snow, frozen soil, topography, wetland); therefore this second kind of flags originates from external datasets and complements the quality flags.

Elementary quality control operations are commonly performed by applying thresholds to quality level descriptors or uncertainty information, to determine which data to use. For example, Numerical Weather Prediction (NWP) centres adopted a suitable set of

quality flags, which allows automatic selection of data of sufficient quality for assimilation; hence quality screened data are preferred at the cost of some data loss (Wagner et al., 2013b). The decisions made in this context are generally application specific and may also be based on the amount of available data, as they may result in a reduction in spatiotemporal coverage and can affect the representativeness of the dataset (Loew et al., 2017).

In hydrological DA applications, quality check (QC) procedures were also set up for remotely sensed SM data, aimed at recognizing the effects of known sources of error, peculiar to the active and passive sensors. QC procedures differ for attributes considered and threshold values adopted. Then, there is another aspect that is usually taken into consideration, in addition to the provided quality indicators of SSM estimates. Surface observations occurred during night and early morning passes are expected to be more appropriate in order to update the soil moisture profile, because the soil is most likely to be in hydrologic equilibrium conditions, avoiding daytime decoupling due to various factors such as evaporation (e.g. Albergel et al., 2008; Wanders et al., 2012). Depending on the specific polar satellite, ascending and descending passes occurred in specific hours of the day, and a check can so be made on the observation orbit direction.

However, such a priori quality controls may not be sufficient for the success of data assimilation (Lahoz et al, 2014). According to Reichle (2008) data assimilation systems should include on-line quality control routines, which cross-compare observations, incorporate information from the geophysical model, and discard inconsistent observations. More generally, additional checks may be necessary such as checking the physical plausibility of a given measurement, visual inspection of the data, or tests for temporal consistency.

In this sense, Walker et al. (2005) suggested the following additional checks:

- consistency or sanity checks, to see if the observation absolute value or time rate of change is physically realistic;
- buddy checks, which compare the observation with comparable nearby (space and time) observations of the same type and reject the questioned observation if it exceeds a predefined level of difference;

- background checks, which examine if the observation is changing similarly to the model prediction. If the user has some reasonable confidence in the model, observations that result in anomalous observation-minus-background residuals are discarded.

4.4 Surface soil moisture propagation to deeper layers

Root-zone soil moisture dynamics are a key factor in numerical weather prediction (NWP), rainfall–runoff processes, agricultural applications, drought risk assessment, ecosystem health. Significant efforts are therefore being made to extend satellite based surface soil moisture data into the root zone. In this sense, a first key challenge was to identify or create models that are structured in a way that is optimal for the assimilation of surface soil moisture data.

The first user community to be attracted by opportunities offered by satellite soil moisture observations was the NWP community, due to importance of soil moisture for modelling land-atmosphere interactions. In particular, root zone soil moisture influences the energy and water exchange occurs continuously at the interface between the land surface and the lower atmosphere. Given the recognized role in weather forecasts played by land surface processes, the NWP community started to invest in developing more physically based Land Surface Models, which also accurately simulate the profile soil moisture dynamics. Hence, efforts have been made in the integration of observations to improve soil moisture representation, also by assimilating remotely sensed surface soil moisture together with conventional measurements of screen-level (near surface) air temperature and relative humidity. For example, the European Centre for Medium Range Weather Forecasting (ECMWF) and the UK Met Office (UKMO) started to operationally assimilate ASCAT SSM data (de Rosnay et al., 2013; Dharssi et al., 2011) with sequential schemes in the shallow layer of their LSMs (7 cm depth for ECMWF, 10 cm for UKMO), after a bias correction phase which makes satellite data consistent with the model. A great challenge faced by soil moisture assimilation in NWP is that improving the soil moisture estimates may not immediately improve atmospheric forecasts due to errors in the model physics; in this sense, the assimilation of satellite data can help to identify and address errors in the model surface flux processes (Ochsner et al., 2013).

Using satellite SSM data assimilation to improve the modelled profile soil water content, some root-zone soil moisture analysis products were released. In the context of EUMETSAT H-SAF, ASCAT-based root zone soil moisture products have been developed through the ECMWF Land Data Assimilation System (de Rosnay et al., 2013). These products are referred to as SM-DAS-2 or H14 (a near-real-time product) and SM-DAS-3 or H27 (an offline product, currently from 1992 to 2014 also using ERS scatterometer data) and provide estimates for four soil layers from surface down to 3 meters. In both products remotely sensed SSM data are assimilated in H-TESSSEL Land Surface Model through an Extended Kalman Filter; screen-level temperature and humidity observations are also used in the analysis. Another product is the SMAP Level 4 Surface and Root-Zone Soil Moisture (L4_SM), which is instead generated by assimilating SMAP brightness temperature observations into the NASA Catchment land surface model using an Ensemble Kalman Filter. The SMAP Level 4 product provides estimates of surface (0-5 cm) and root-zone (0-100 cm) soil moisture. There are some limits (e.g. the fixed soil depths) in the use of these root-zone products as observations in a data assimilation framework with a second prediction model, above all the fact that products provide analysis estimates, thus including the first model errors. This could be reflected in errors correlated with the second model background estimates, if both models share some elements in physics, parameters or forcings.

In some cases, also the structure of hydrologic models was modified to maximize the benefits of SSM integration. Soil moisture is indeed one of the key variables in flood forecasting models, as it plays an important role in partitioning rainfall into runoff and infiltration. Different remotely sensed soil moisture products are used along with various hydrologic models, including conceptual models where the representation of physical processes is simplified compared to that in LSMs, for example by typically simulating the water storage of a single soil column to represent the unsaturated zone. As satellite data requires a model tailored to their use, some conceptual hydrological models were modified by introducing a surface soil layer, such as the GRHUM (Loumagne et al., 2001), GRKAL (Francois et al., 2003) and MISDc-2L (Brocca et al., 2012a), making possible to directly assimilate surface data. During DA analysis steps, also deeper layers are usually updated; however, in some cases (e.g. Lievens et al., 2015, 2016) bottom layers SM were not included in state vector to avoid stability problems. According to Kumar et al. (2009), the effectiveness of deeper layer SM

updates via surface assimilation depends on vertical coupling between soil layers, which is, in turn, determined by the specific model used, model parameters and climate. However, the quality of various layers updates has different impact in runoff improvement, as each individual SM layer can have a different control on runoff. Using an alternative approach, in Brocca et al. (2012a) root-zone SM was estimated from surface data through the application of an exponential filter, and both satellite-based surface and root-zone data were then assimilated into a two-layer rainfall-runoff model. In their study the assimilation of satellite-based root-zone estimates in the bottom layer produced a significant positive impact on runoff simulation, while assimilation of SSM product in the top layer had only a small effect.

Other methods were proposed to propagate SSM information to the depths of interest for the model, including the exponential filter, regressive relationships, artificial neural networks, and physically based formulations.

A widely used approach is the above mentioned exponential filter (Wagner et al., 1999) to obtain a root-zone soil moisture estimate, typically referred to as soil water index, SWI, from available remotely-sensed surface measurements. The method is widespread for its simplicity as it involves using only one parameter, and has been extensively employed also to evaluate satellite-based root-zone soil moisture estimates by comparison with in situ and modelled data (e.g. Ceballos et al., 2005, Albergel et al., 2008, Paulik et al., 2014, Cho2015, Tobin2017). The exponential filter is the predominantly used within a more general category of filtering techniques (Brocca et al., 2013), which infer the root zone estimate from the available SSM values. The Copernicus Global Land Service provides a SWI product from ASCAT SSM datasets (Paulik et al., 2014), according to different exponential filter parameter values in order to represent soil moisture at various depths; with respect to the previous mentioned satellite-based root zone products, SWI is much closer to the original satellite measurements because it is obtained without the use of a complex land surface modelling scheme. Satellite-based SWI was used as indicator or variable in forecasting models for drought and crop production (e.g. de Wit & van Diepen, 2007; Zhao et al., 2008; Zribi et al., 2010; Amri et al., 2012; Qiu et al., 2014; González-Zamora et al., 2016), landslide (Brocca et al., 2012b) and soil loss (Todisco et al., 2015). The exponential filter is also commonly used to obtain root-zone soil moisture observations for data assimilation in hydrological models (e.g. Brocca et al., 2010a, 2012a; Matgen et

al., 2012; Alvarez-Garreton et al., 2014, 2015, 2016; Massari et al., 2015; Laiolo et al., 2016; Cenci et al., 2016, 2017; Loizu et al., 2018), as the SWI method is easy to be implemented for all conceptual models without requiring modifications to soil layers structure.

Regression approaches simply fit mathematical relationship between surface and deeper soil moisture, while artificial neural networks (ANN) are a data-driven method that uses interconnected processing elements to represent the nonlinear and heterogeneous nature of the soil moisture variation across depths. ANN approaches usually requires multisource information (e.g. meteorological variables) to enhance model accuracy, and/or the use of a land surface model for the training (e.g. Kornelsen & Coulibaly, 2014). In a comparative study using in situ data, Zhang et al. (2017) obtained that exponential filter performs better than linear regression and ANN for vertical extrapolation of soil moisture. Although the ANN performances depend significantly on the used configuration, they are more computationally demanding and requiring higher amount of input data when compared to other considered methods.

Finally, a simple physically based approach for estimating root-zone soil moisture from surface observations was proposed by Manfreda et al. (2014). The method, referred to as SMAR, requires the identification of few parameters that are related to the physical characteristics of the area under investigation. The SMAR method is not designed to be followed by a bias correction of root zone estimates, unlike the exponential filter where SWI mainly works as trend indicator of the profile soil moisture. Both remotely sensed SSM and in situ surface measurements are used within the SMAR model, obtaining a good description of root zone soil moisture dynamics after proper calibrations (Faridani et al., 2016). In Baldwin et al. (2017) an ensemble Kalman filter was used to support SMAR, by estimating a regional-scale satellite bias parameter using reference in situ data. The bias parameter is added to input SSM retrievals before their use in SMAR; in this way, the optimized SMAR-EnKF model predicted root zone soil moisture with a lower degree of error respect to previous SMAR calibration efforts.

4.4.1 The exponential filter

Wagner et al. (1999) proposed an exponential filter to estimate root-zone SWI from remotely sensed SSM time series, making use of one parameter only, T , representing a

characteristic time length. The method has the effect of smoothing and retarding the SSM time series, mimicking the diffusion process of the water into the deeper soil layers. In this sense, the exponential filter reproduces a simple two-layer water-balance model, where the soil moisture dynamics of the lower layer are captured by convoluting the input top layer SM measurements. In the soil water balance equation the water fluxes between the two layers are controlled by a constant pseudodiffusivity term and assumed to be linearly related to the difference in soil water content; many important processes like transpiration are not explicitly considered. The soil water index at time t_n , SWI_n , is defined by equation (4.1), where SSM_i is the surface soil moisture at time t_i :

$$SWI_n = \left(\sum_i^n SSM_i e^{-\left(\frac{t_n-t_i}{T}\right)} \right) / \left(\sum_i^n e^{-\left(\frac{t_n-t_i}{T}\right)} \right) \quad (4.1)$$

In alternative, the SWI can be calculated using a mathematically equivalent recursive formulation (Albergel et al., 2008):

$$SWI_n = SWI_{n-1} + K_n(SSM_n - SWI_{n-1}) \quad (4.2)$$

with the gain K_n , comprised in the interval $0 \div 1$, also expressed in a recursive form:

$$K_n = K_{n-1} / \left(K_{n-1} + e^{-\left(\frac{t_n-t_{n-1}}{T}\right)} \right) \quad (4.3)$$

In this recursive formulation, initial values SWI_0 and K_0 are set to SSM_0 and 1, respectively.

A comprehensive analysis of the exponential filter was made by Ceballos et al. (2005). Given its high variability, SSM time series resembles a white-noise process (uncorrelated in time), while the derived SWI time series has a spectrum similar to that of a red-noise process (serially correlated in time). The characteristic time length T acts a surrogate parameter that takes into account the different processes that affect the temporal scale of soil moisture (layer depth, soil hydraulic properties, evaporation, transpiration, runoff, soil layers, etc.); given the limited understanding of seasonal effects on soil moisture time scales, T is usually treated as a constant, although it is expected to vary seasonally.

In practice, T is usually determined by optimizing the correlation between SWI and reference root-zone soil moisture data, e.g. model estimates. Despite its simplicity, the exponential filter has been found to approximate the profile soil moisture content quite well, with T depending mainly on the soil depth.

In the case of extensive temporal SSM data gaps relative to the filter parameter T , equation (4.3) shows that the gain for the next available observation tends toward unity.

As a result, when a new SSM value is available, the previous SWI estimate is disregarded and the SWI estimate at that time takes the value of the new SSM observation. It is evident in the recursive formulation that the SWI estimated through the exponential filter approach suffers from prolonged data gaps in SSM data, compared to the filter parameter value T . Conversely, a sufficiently low gain K_n implicitly indicates an adequate number of input observations in the temporal range of interest for the filter estimation.

It is worth noting that, although widely used in DA frameworks, SWI estimates present time-correlated errors that, theoretically, are not compatible with the common data assimilation assumptions.

4.4.2 The SMAR method

Manfreda et al. (2014) proposed a soil moisture analytical relationship (referred to as SMAR) to infer the root zone state from surface data. The method is derived from a simplified two-layer soil water balance and involves a limited number of physically consistent parameters. It provides a solution that may be considered particularly reliable in dry areas, as a linear soil water loss function is assumed in the bottom layer to represent both evapotranspiration and percolation, which implies some limitations in the use in humid environments.

The relative saturation in the deeper layer at time t_j , given by the ratio between soil water content and porosity and indicated as $S_{2,j}$, is obtained from its previous estimate $S_{2,j-1}$ and the current relative saturation in the surface layer, $S_{1,j}$. For this purpose, SMAR use the following seven parameters that can be related to physical characteristics of the soil profile:

- n_1 and n_2 are the soil porosity of the top and bottom layers, respectively;
- Z_{R1} and Z_{R2} are the depth of the top and bottom layers, respectively;
- V_2 is the soil water loss coefficient of the bottom layer, accounting for both evapotranspiration and percolation losses;
- S_{C1} is the relative saturation at field capacity of the top layer;
- S_{W2} is the relative saturation at wilting point of the bottom layer.

The soil water balance of the deeper soil layer is controlled by infiltration from the first layer and soil water losses (other processes such as horizontal lateral flows are assumed negligible).

The water flux from the top layer can be considered significant only when the soil moisture exceeds field capacity:

$$y_j = \begin{cases} S_{1,j} - S_{C1} & S_{1,j} \geq S_{C1} \\ 0 & S_{1,j} < S_{C1} \end{cases} \quad (4.4)$$

where y_j is the fraction of soil saturation infiltrating in the lower layer at time t_j .

As mentioned, a linear function is assumed to describe soil water losses, varying from the maximum value V_2 under well-watered conditions to 0 at the wilting point.

Introducing the normalized coefficients:

$$a = \frac{V_2}{(1-S_{W2})n_2Z_{R2}} \quad (4.5)$$

$$b = \frac{n_1Z_{R1}}{(1-S_{W2})n_2Z_{R2}} \quad (4.6)$$

the analytical relationship can be written as:

$$S_{2,j} = S_{W2} + (S_{2,j-1} - S_{W2})e^{-a(t_j - t_{j-1})} + b(1 - S_{W2})y_j(t_j - t_{j-1}) \quad (4.7)$$

having reduced the number of parameters from seven to four (a , b , S_{C1} , S_{W2}). All these parameters may be estimated from the soil texture, the soil depth, and the soil water losses, or also calibrated using reference root-zone data. The relative saturation at field capacity can be set as initial value for the soil moisture in the second soil layer. It should be noted that SMAR may produce values higher than 1 and that these must automatically set equal to 1.

A modified version of SMAR was recently proposed (Faridani et al., 2017), by including a non-linear soil water loss function explicitly describing deep percolation and evapotranspiration dynamics.

4.5 Bias correction methods

Soil moisture observations are integrated in prediction models through the commonly used data assimilation methods, which provide optimal statistical solutions in terms of error minimization having made specific assumptions on the observation error structure, such as the independence between error and true state realizations, the absence of any other type of systematic error, observations and model estimates that on average agree

with the true states. In reality, however, biases are unavoidable and it is difficult to attribute the bias to the model or the observations.

In this sense, a more general linear error structure is often used for soil moisture datasets:

$$X_i = \alpha + \beta T_i + \varepsilon_i \quad (4.8)$$

where X_i and T_i are the estimate and the true state value at time i , respectively. The other terms, on the right-hand side, represent three separate types of errors, with the first two jointly quantify the systematic error. The term α represents a constant deviation, while the term β generates a scale error when deviates from unit. Finally, the third term, ε_i , indicates the independent random error, typically assumed as Gaussian distributed, with zero mean value and a standard deviation of σ_ε . In terms of statistical moments, X is biased in both mean and variance with respect to T :

$$\mu_X = \alpha + \beta \mu_T \quad (4.9)$$

$$\sigma_X^2 = \beta^2 \sigma_T^2 + \sigma_\varepsilon^2 \quad (4.10)$$

The proper treatment of systematic errors is critical for the success of data assimilation systems (Dee & da Silva, 1998). Without bias correction, it is not possible to conduct optimal data assimilation, and neither meaningful comparisons between datasets for evaluation purposes. Many studies have been generally devoted to addressing the discrepancies between observed and modelled soil moisture, focusing on the removal of the bias in observations or model states before or during data assimilation. In this sense, Dee (2005) characterizes the data assimilation systems as either “bias blind” or “bias aware” on the basis of their treatment of systematic errors.

As discussed in the previous chapter, dynamically bias-aware assimilation systems incorporate specific assumptions about the nature of biases and are specifically built to estimate and correct them during data assimilation procedures; these strategies typically attribute the bias to either the model or the observations.

The bias-blind DA systems assumes the absence of bias in both datasets and are designed to only correct random, zero-mean errors, so they require an a priori bias correction strategy.

Implicitly assuming no biases in the observation data, a first alternative is to calibrate the model parameters against the satellite-based measurements, in order to minimize the bias between model forecasts and observations, and then perform data assimilation for

the calibrated model (e.g. Kumar et al., 2012). However, this a priori calibration approach has not been widely implemented.

The second, most adopted solution is to perform bias-correction of the remote sensing data to conform them to the model, with the implicit outcome that all bias is attributed to the observations. Several a priori scaling methods have been proposed, often based on the match between datasets statistical moments. Many of them have been extensively used for the purpose of homogenization between datasets before comparison, for example in validation activities against a reference. In this sense, methods have been addressed to remove systematic differences such biases in both mean and amplitude of variations (as expressed in the statistical variance) between datasets, without any particular assumption on the error structure. The most widely used rescaling technique is the cumulative distribution function (CDF) matching approach (Reichle & Koster, 2004), that allows matching the complete CDF of satellite and model estimates by applying a nonlinear operator. Several studies employed simple linear techniques (e.g. Scipal et al., 2008a; Jackson et al., 2010) that are expected to be more robust as they require a lower number of parameters. A comprehensive overview of rescaling methods can be found in Afshar & Yilmaz (2017), who compared 31 different linear and nonlinear approaches, highlighting how the nonlinear ones perform better in reducing differences between rescaled and reference data.

However, the goal of the rescaling efforts can be different depending on the applications and their underlying assumptions. In the framework of data assimilation, both model estimate and observation are considered containing an ‘useful’ part related to the true state information (also called signal) and an error component (also called noise); in this sense the rescaling should be addressed to match only the signal part between datasets. This contrasts, for example, with rescaling methods that match the total variance, as model estimates and observations rescaled in model-space are expected to have different total variances, being both characterized by additive random error with distinct error variances. In this sense, common techniques such as CDF matching can induce artificial biases in the signal component of the satellite data and thus become sub-optimal in order to remove the biases; more generally, systematic differences between model estimates and rescaled observations can persist and degrade DA performances. For example, Kornelsen & Coulibaly (2015) highlighted how the common rescaling techniques are not effective in correcting multiplicative bias between datasets, which

involves that assimilation tends to systematically increase or decrease the background estimates depending on state value, instead of only reducing random errors. Rescaling techniques proper for DA have to take in account the error structure of datasets; for example the triple collocation analysis (Stoffelen, 1998) considers the error structure defined in (4.8) and provides linear rescaling solutions that allow to match the signal components of the datasets. Yilmaz & Crow (2013), showing that DA accuracy depends on the degree to which the signal component of observations are rescaled to the signal component of the model, demonstrated that triple collocation analysis is an optimal solution in an assimilation framework and a better rescaling method than other widely used linear techniques. Finally, triple collocation should be effective in reducing the multiplicative bias between datasets, as this problem can be intuitively associated to differences between the respective β terms in (4.8), persisting when common techniques are used and instead corrected with the signal matching approach. The triple collocation is a widely used method for random error characterization, and as such discussed in detail in Sect. 4.6.2, both in assumptions and formulations.

Four different rescaling techniques are illustrated in the following, three linear methods (linear regression, mean and variance matching, triple collocation analysis, then referred to as REG, VAR and TCA, respectively) and the (nonlinear) CDF matching approach. REG, VAR and CDF matching were widely implemented in satellite-based soil moisture rescaling prior to hydrological DA; about which technique performs better, contrasting results are reported in literature when they are used in a comparative way (e.g. Alvarez-Garreton et al., 2014; Lievens et al., 2015; Massari et al., 2015; Loizu et al., 2018). There are instead few studies in which the assimilated observations were rescaled with the triple collocation analysis (e.g. Alvarez-Garreton et al., 2015), although the method is extensively used prior to DA for the observation error characterization.

4.5.1 Linear methods

Linear approaches can be described with the common form:

$$X^* = \mu_Y + c(X - \mu_X) \quad (4.11)$$

where μ is the dataset mean, Y denotes the model dataset, c is a scalar rescaling factor, while X and X^* are the unscaled and rescaled versions of the observation dataset, respectively.

The linear regression method (e.g. Jackson et al., 2010) is based on the application of a regression equation between the satellite and model soil moisture time series, i.e. $X^* = mX + q$ with regression parameters that minimizes the squared-differences between Y and X^* . Referring to (4.11), this provides:

$$c_{REG} = \rho_{XY} \frac{\sigma_Y}{\sigma_X} \quad (4.12)$$

where ρ and σ denote the correlation and the standard deviation, respectively. The linear regression method only matches the mean of the two datasets.

The mean and variance matching (e.g. Scipal et al., 2008a), often reported as linear scaling, allows to match the first two moments of datasets. Referring to (4.11), this gives:

$$c_{VAR} = \frac{\sigma_Y}{\sigma_X} \quad (4.13)$$

According to Scipal et al. (2008a), the impact of ignoring differences in higher-order moments is compensated by the robustness of the method, as it requires a lower parameterization.

Finally, the triple collocation analysis requires a third independent dataset, Z , which works as instrumental variable. It is assumed that all three datasets are related to the true state value through an additive and a multiplicative bias components and a random error term, as in (4.8). According to this error structure, the total variance of a dataset has the two additive components showed in (4.10), related to true state and random error variances respectively. A rescaling can be made in order to isolate differences between datasets due to only random errors. In order to match the systematic component of Y , X can be rescaled using the (4.11) with the following covariance ratio:

$$c_{TCA} = \frac{\sigma_{YZ}}{\sigma_{XZ}} \quad (4.14)$$

Triple Collocation approach allows X^* to match Y in the mean and in the variance component related to true values (signal variance), i.e. the first term on the right-hand side in (4.10). Random errors in X are rescaled in the Y -space, and X^* and Y generally have different total variances, due to the general difference between their random error variances. Details on triple collocation approach, also as scaling method, can be found in Sect. 4.6.2.

4.5.2 CDF matching

The CDF matching (Reichle & Koster, 2004) is a non-linear method that allows to match the CDF of the satellite-derived data to the CDF of the model estimates. Conceptually, the observational dataset X is converted to X^* according to:

$$CDF_Y(X^*) = CDF_X(X) \quad (4.15)$$

where Y denotes the model dataset, as also schematically showed by the arrows in Fig. 4.1. CDF matching generally corrects all moments of the distribution function regardless of its shape.

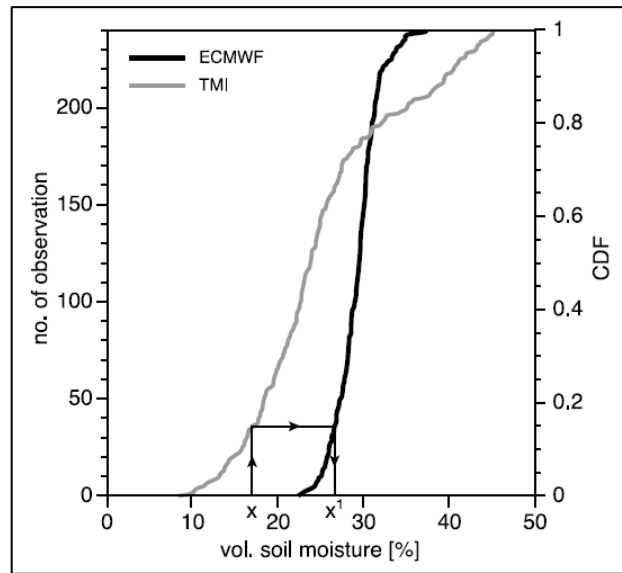


Figure 4.1. Figure taken from Drusch et al. (2005), representing CDF matching schematic. The arrows illustrate how the satellite SM data (here obtained from TMI sensor) are scaled to match the distribution of model data (here obtained from ECMWF simulations).

Technically, a common way to apply this method was proposed by Drusch et al. (2005). The two datasets (modelled soil moisture and observations) have to be ranked; consecutively, the differences in soil moisture between the corresponding elements of each ranked data set have to be calculated; finally, a polynomial function f is computed to fit the relationship between the ranked differences and satellite data. The polynomial function f is then used to calculate the CDF-corrected soil moisture datasets:

$$X^* = X + f(X) \quad (4.16)$$

Drusch et al. (2005) proposed a third-order polynomial for f ; in Brocca et al. (2013) the use of a fifth-order polynomial function was found more suitable.

4.6 Error characterization methods for evaluation and data assimilation of satellite soil moisture

As already discussed, bias-blind DA methods are used in the most of applications, implicitly assumed to target only the random errors, by optimally reducing their error variance with analysis schemes; systematic differences between model and observations are typically reduced prior to DA with bias correction techniques. In DA, random errors are typically modelled as additive, temporally-uncorrelated and Gaussian-distributed, with zero-mean and usually time-invariant known error variance. Then, an effective characterization of observation random error variance is fundamental in order to obtain optimal data assimilation results, i.e. analysis estimates actually characterized by the minimum error variance. However, in real cases the observation error structure is likely to be different from that typically assumed, in terms of both random errors modelling and systematic errors absence; an example is the error structure showed in (4.8).

More generally, error characterization constitutes a fundamental step to assess the suitability of observations for a specific application or also a decision-making process. Therefore, error assessment was one of the primary objectives of satellite-based soil moisture evaluation activities, even more so recently when dedicated missions and products have been developed with specific target accuracy. For remotely sensed soil moisture observations, validation activities have been finalized to error assessment through comparison with other soil moisture datasets, which are in turn affected by errors; furthermore, it would be better to use the term “evaluation” rather than “validation”, and to refer to “differences” rather than to “errors”.

Several approaches are available for error characterization, also provided by the widely used evaluation practices, which in this case well combine with soil moisture DA purposes. In the following it will be described how the error characterization is carried out by the use of standard statistical measurements and by the triple collocation and the error propagation methods.

The first, extensively used approach to evaluate remote sensing soil moisture is based on comparison with an independent dataset, considered more reliable and used as a

reference, by computing a common set of performance metrics that summarize the similarity and difference between the two time series. The most commonly used performance metrics are the correlation coefficient (r), the root-mean-square difference ($RMSD$), the bias ($Bias$), and the unbiased root-mean-square difference ($ubRMSD$) (Albergel et al., 2013); they are related and provide complementary information (Entekhabi et al., 2010a). In particular, r measures the relative agreement, while the others focus on absolute differences, with $RMSD$ being the most comprehensive one.

In situ measurements are generally considered the most reliable to act like reference dataset (e.g. Chan et al., 2016), as they might be highly accurate on a point scale, provided that a proper calibration has been performed to contain instrumental errors. However, given the strong difference in spatial support, ground sensor measurements could be affected by significant errors in the footprint scale soil moisture representation. In situ representativeness errors could then generate misleading results when sparse ground networks are considered in evaluation. The footprint scale soil moisture variability can instead be better characterized in dense networks (also referred to as core validation sites, CVS), with multiple calibrated stations within a satellite pixel: in this case it is possible to choose the most representative site (e.g. Vachaud et al., 1985) or to aggregate the local in situ measurements at the coarse scale (e.g. Teuling et al., 2006). However, core validation sites are very limited in number on a global scale. Model estimates are also used as reference dataset (e.g. Brocca et al., 2011a), especially in areas where no suitable ground data are available; models can be tuned to match the scale of the satellite data, but, of course, the interpretation of the results will always be hampered by the modelled soil moisture accuracy, influenced for example by model imperfections and errors in input data (precipitation, soil properties, etc.).

$RMSD$ generally takes in account the effect of random and systematic errors in the two datasets. Considering the case of the error structure described in (4.8), if there are no systematic differences (i.e. the two datasets have the same values of α and β) and random errors are mutually independent, then the mean square difference (MSD) ideally corresponds to the sum of random error variances of satellite and reference. Systematic differences between satellite and reference dataset can be reduced in preprocessing. While differences in mean (i.e. the $Bias$) can be easily corrected or taken in account (focusing on $ubRMSD$ rather than $RMSD$), other systematic deviations, such as the state-dependent ones, cannot be completely removed by the common bias correction

techniques based on total variance matching (such as linear scaling or CDF-matching). Then, residual systematic differences can also contribute in computed *RMSD*.

Assuming that random errors in chosen reference data are negligible (approximation usually accepted for CVS) and systematic differences with satellite measurements were previously effectively corrected, then the root-mean-square error (*RMSE*), here used rather than *RMSD* and coincident with the *ubRMSE* value, provides an estimate of the variance of the additive zero-mean random errors in remotely sensed soil moisture observations.

In this sense, the *RMSE* evaluated using in situ data from CVS (or the *ubRMSE* if no bias correction was performed on satellite data) is considered the fundamental benchmark in error characterization of many satellite soil moisture datasets, such as retrievals from AMSR-E (Jackson et al., 2010), SMOS (Jackson et al., 2012; Kerr et al., 2016) and SMAP (Chan et al., 2016; Colliander et al., 2017) missions.

For example, the dedicated SMAP mission target accuracy of $0.04 \text{ m}^3/\text{m}^3$ volumetric soil moisture is evaluated against in situ surface observations according to two approaches (Chan et al., 2016), with the first one being to compute *ubRMSE* on selected CVS (Colliander et al., 2017). In this case, point-scale measurements were aggregate to satellite pixel; the time-variant uncertainty in representing the spatial distribution of soil moisture was also computed, with information used only to verify that a threshold has not been exceeded during the CVS candidates selection (i.e. errors in ground estimate are limited, so differences between datasets are mainly due to retrieval errors). The second approach for evaluation of SMAP target accuracy considers instead sparse ground networks applying the Triple Collocation method (Chen et al., 2017), which will be discussed later.

The *RMSE* is not only the metric specified for remote sensing target accuracies: it is also useful for specifying observation error variances for data assimilation. In many DA applications it was used a *RMSE* value (or more likely *RMSD*, depending on the assessment methods), locally computed using a reference dataset or taken from literature for the considered satellite-based soil moisture product, although estimated elsewhere (e.g. Draper et al., 2011; Laiolo et al., 2016).

An alternative to the a priori assignment of a plausible value is the direct optimization of the observation error variance in order to obtain the best DA performance. When the

DA procedure is iteratively used to assess the observation error, then the ‘optimal’ result is essentially case-specific.

Making an example in which the application goal is to improve streamflow predictions, for the same catchment and assimilated satellite SM product the ‘calibrated’ error variance could be very variable depending on the considered observation preprocessing procedure, hydrological model structure and performance, DA implementation (e.g. Loizu et al., 2018); similarly, with equal these last factors, adjacent and similar catchments could have very different optimal observation errors for the same satellite product (e.g. Massari et al., 2015).

One approach which aims to overcome some of the difficulties encountered when comparing only two data sets is the so-called Triple Collocation (TC) technique, introduced by Stoffelen (1998) and firstly applied in soil moisture datasets evaluation by Scipal et al. (2008b). TC is based on the assumption that a SM estimate is related to the hypothetical but unknown real SM content through an additive and a multiplicative bias components and a random error term, as in (4.8); considering three independent collocated datasets, TC allows an estimation of their random error variances. The method assumes independent error structures, which means that the errors must not have the same origin. This is true when using any combination of in situ measurements, active or passive satellite observations, and land surface model estimates, provided that the model is not driven by one of the other dataset. TC can provide realistic error variance estimates and is able to successfully distinguish spatial error trends, not being limited by the availability of CVS (more generally TC does not require the availability of a high-quality reference dataset). The method has been widely recognized as a powerful tool for estimating random error variances of coarse-resolution SM datasets and was extensively used in DA applications (e.g. Chen et al., 2014; Alvarez-Garreton et al., 2015; Massari et al., 2018). Nevertheless, studies also showed that the result is highly sensitive to the input configuration (Loew & Schlenz, 2011; Zwieback et al., 2012), including different scales and represented physical quantities of the sources, the use of SM absolute values or anomalies, the time span under observation, and the available number of measurement triplets. About the last point, a sufficient long data record is needed for robust estimations (100 triplets is the minimum boundary, 500 is advised). Zwieback et al. (2012) showed that the robustness of the TC is closely related to the strict application of its basic assumptions, while Gruber et al. (2016a) provide a

comprehensive discussion of these assumptions, highlighting how on one hand some of them are not always met, while on the other hand the same assumptions are also implicitly made in the application of conventional performance metrics, e.g. temporally uncorrelated (white) random errors and stationarity of the random error variance.

Recently, several studies proposed improvements or variations to the TC method, originally developed for solving the random error variances of three independent datasets. McColl et al. (2014) introduced Extended Triple Collocation (ETC): using the same assumptions of TC, they derive the correlation coefficients of the datasets with respect to the unknown true states, that provides a complementary perspective compared to the random error variances. Pierdicca et al. (2015) proposed to extend TC analysis with a fourth dataset and to solve this Quadruple Collocation (QC) problem as an overdetermined system through a least squares minimization, in order to provide a more robust solution for the error variance estimates; QC still requires uncorrelated errors between all four data sets. Following Zwieback et al. (2012), where three systems were used to estimate the error variance, and a fourth one was exploited to estimate the cross-correlation among system errors, Gruber et al. (2016b) proposed a method, referred to as Extended Collocation (EC), for estimating error cross-correlations by generalizing the TC method to an arbitrary number (>3) of datasets and relaxing the therein made assumption of zero error cross-correlation for a limited number of dataset combinations, depending on the number of datasets used and their assumed underlying error structure; this implies that the systems with correlated errors must be known a priori. Pierdicca et al. (2017) presented an Extended Quadruple Collocation (E-QC) method, actually applicable to a generic number of datasets, that considers the possibility of an error cross-correlation between two products; E-QC allows an automatic detection of the couples of error cross-correlated systems, provided that one system is known to be independent from the others. Dong & Crow (2017) proposed a Generalized Triple Collocation analysis algorithm (GTC) which statistically decomposes the total random error variance into its autocorrelated and white error components. Nearing et al. (2017) presented a nonparametric triple collocation that alleviates the need for some of the original assumptions, such as the linear form presented in (4.8) and the independence of true state and random error; indeed no particular form of the error structure was assumed, it is only required that there be a

joint probability distribution between the truth and all three measurements. TC and all its variants are in any case based on error stationarity assumption.

Another technique used to assess the soil moisture error is the error propagation (EP), which does not involve a comparison with other datasets. The method is based on the propagation of random error variances in input variables and model parameters through the model equation (e.g. the retrieval algorithm), in order to calculate the error variances related to each individual model output. The main advantage of error propagation technique is that it allows for calculating an error variance estimate for each individual observation, unlike the previous methods which assume error stationarity. The method requires the error characterization of inputs and soil moisture model parameters, generally in the form of an error covariance matrix; the propagation to the model output is then carried out through analytical solutions or methods such as Monte Carlo simulations, mainly depending on the model complexity. EP do not consider the model structure uncertainty (Dorigo et al., 2010), which also contributes to the output error budget; this can lead to unrealistic error estimates, especially in the case of increasingly accurate inputs and model parameters. Studies comparing the suitability of EP e TC methods in satellite SM assessment indicate that, given their peculiarities, the two techniques should be seen as complementary (Dorigo et al., 2010; Draper et al., 2013).

Error propagation schemes have been developed for both passive and active remote sensing SSM retrieval algorithms. Parinussa et al. (2011) introduced an analytical solution for estimating the error variance of soil moisture retrievals from radiometers based on the Land Parameter Retrieval Model (Owe et al., 2008), obtaining results similar to those of a Monte Carlo approach in an application on AMSR-E measurements. Naeimi et al. (2009) and Pathe et al. (2009) set error propagation models in parallel to the development of the TU-Wien change detection method (Wagner et al., 1999) versions for ERS scatterometer and ASCAT and for Advanced Synthetic Aperture Radar (ASAR) on-board Environmental Satellite (EnviSat), respectively.

Few studies analysed the reliability, both in magnitude and in spatial patterns, of SSM error variances estimated by EP. Naeimi et al. (2009) observed high values of averaged ASCAT SSM error variance in correspondence of globe areas where the retrieval algorithm is not expected to work properly, e.g. areas with dense vegetation cover. Doubková et al. (2012) analysed the capability of ASAR SSM error variances to predict the difference observed between satellite and modelled SSM estimates over Australia,

finding a strong correspondence in spatial patterns as well as a good quantitative agreement. By comparing ASCAT and AMSR-E SSM anomalies from the mean seasonal cycle over North America using both TC method and error propagation estimates, Draper et al. (2013) highlighted that EP can accurately detect the large scale variability in soil moisture errors; however the magnitude of the EP output could be not generally informative: in their study ASCAT estimated error variances appear to be approximately correct, while the AMSR-E ones are unrealistically large.

SSM error variances estimated by EP were used to perform quality check of SSM retrievals (e.g. Scipal et al., 2008a; Mahfouf, 2010; Dharssi et al., 2011; Draper et al., 2011, 2013; Paulik et al., 2014) i.e. to detect areas where the retrieval method does not work well or to mask single spurious SSM observations.

There are few applications of EP estimates as observation error variances in DA. This is due to the fact that the error stationarity is usually assumed in soil moisture DA applications, although several studies (e.g. Loew & Schlenz, 2011; Su et al., 2014b) highlighted that this assumption seldom hold, mainly due to imperfections in the treatment of seasonalities of contributing processes (e.g. vegetation growth) in the retrieval model (Gruber et al., 2016a). In Draper et al. (2011) the time-variant EP estimates provided in ASCAT product were used to characterize the observation errors during the SSM assimilation into an operational hydrological model. Time-variant observation error variance was also assumed by Lievens et al. (2015, 2016) during the assimilation of SMOS soil moisture data in both hydrological and land surface models: in these studies the error estimates were derived from a linear combination of product flags, including the time-variant Data Quality index (DQX). DQX represents the SSM error standard deviation (provided in volumetric soil moisture) obtained with a more comprehensive approach, that includes the propagation of the errors in inputs and algorithm parameters but also embed other sources of errors (such as the Radio Frequency Interference, RFI).

4.6.1 Standard statistical measurements

A common approach to evaluate the accuracy of a dataset (that will be here referred to as the “test”) is by comparison with a second one, considered more reliable (the “reference”), quantifying how closely the test resembles the reference. Evaluation is

made by computing some metrics which statistically quantifies pattern similarity and differences.

The statistic most often used to quantify pattern similarity is the Pearson's correlation coefficient r , which well describes the ability of the test dataset to capture temporal soil moisture changes in the reference (Entekhabi et al., 2010a). Considering the two datasets x_{TEST} and x_{REF} defined at N discrete times, r is expressed by:

$$r = \frac{\frac{1}{N} \sum_{i=1}^N (x_{TEST,i} - \mu_{TEST})(x_{REF,i} - \mu_{REF})}{\sigma_{TEST} \sigma_{REF}} \quad (4.17)$$

where σ_{TEST} and σ_{REF} are the standard deviation and μ_{TEST} and μ_{REF} the mean values of the two datasets. The correlation coefficient, which ranges from -1 to 1 , is an index of the degree of linear relationship between the two datasets. If $r = 0$, no linear relationship exists; if $r = 1$ or -1 , a perfect positive or negative linear relationship exists. The correlation coefficient captures the correspondence in phase between the two datasets and does not take in account the additive and proportional differences, i.e. r does not allow determining whether x_{TEST} and x_{REF} have the same amplitude of fluctuations (as expressed in the statistical variance) and/or the same average values.

The statistic most often used to quantify differences in two datasets is the root-mean-square difference ($RMSD$), expressed as:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_{TEST,i} - x_{REF,i})^2} \quad (4.18)$$

It gives the average magnitude of differences, weighted according to their squares. $RMSD$ value is in the same unit of the two datasets, and varies on the interval $[0 \text{ to } \text{inf}]$, where values of 0 indicate a perfect fit.

Biases in the mean and/or differences in the amplitude of fluctuations (i.e. biases in the variance) between datasets will lead to high $RMSD$ values. In order to isolate the differences in the means of the two datasets, $RMSD$ can be resolved into two components, the bias ($Bias$) and the unbiased root-mean-square difference ($ubRMSD$).

The bias can be expressed as:

$$Bias = \mu_{TEST} - \mu_{REF} \quad (4.19)$$

With this convention, a negative bias indicates that the test underestimates the reference on average and a positive bias the other way around.

The unbiased root-mean-square difference is expressed by:

$$ubRMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N \left((x_{TEST,i} - \mu_{TEST}) - (x_{REF,i} - \mu_{REF}) \right)^2} \quad (4.20)$$

The two components add quadratically to yield the full mean square difference:

$$RMSD^2 = Bias^2 + ubRMSD^2 \quad (4.21)$$

For a given value of *ubRMSD* it is impossible to determine how much of the error is due to a difference in structure and phase and how much is simply due to a difference in the amplitude of the variations.

The correlation coefficient and the root-mean-square difference provide complementary information quantifying the correspondence between two patterns and are also linked to each other (e.g. Murphy, 1988; Gupta et al., 2009).

In particular, the following relationship exists:

$$ubRMSD^2 = \sigma_{TEST}^2 + \sigma_{REF}^2 - 2\sigma_{TEST}\sigma_{REF}r \quad (4.22)$$

so the *RMSD* can be easily decomposed into a bias-in-mean, bias-in-variance, and a correlation-dependent component:

$$RMSD^2 = (\mu_{TEST} - \mu_{REF})^2 + (\sigma_{TEST} - \sigma_{REF})^2 + 2\sigma_{TEST}\sigma_{REF}(1 - r) \quad (4.23)$$

With specific reference to soil moisture, temporal variations are often dominated by seasonal patterns of alternating dry and wet periods, which to a high degree determine the *r* values; therefore anomaly correlations are also computed, to exclude the seasonal cycle and evaluate the relative agreement between short-term fluctuations proceeding with a much weaker magnitude (Scipal et al., 2008a). Soil moisture anomalies are computed in two ways. In a first approach, anomalies are defined as the deviations from a simple moving average, considering a centred time window of appropriate extension, usually 35 days (Albergel et al., 2009). When multiple years coverage are available for both datasets, anomaly can be defined as the deviation from the long term climatological expectation; the latter is computed for a given day of the year (DOY) through averaging of multiyear data within a centred window of appropriate extension, usually 31 days (e.g. Dorigo et al., 2010).

4.6.2 Triple collocation

The Triple Collocation (TC) was originally proposed by Stoffelen (1998) in order to study the error characteristics of near-surface wind vector data derived from a model, buoy measurements and scatterometer observations, and is now widely used not only in

oceanography but also in hydrology, especially in the field of satellite-based soil moisture evaluation (McColl et al., 2014).

The TC method does not require the specification of a “true” reference dataset and is based on the availability of three spatially and temporally collocated, independent datasets (X, Y, Z) which can be assumed to represent the same geophysical variable. Each dataset is assumed to be related by a linear form equivalent to (4.8) with the unknown true state, here indicated as T :

$$X = \alpha_X + \beta_X T + \varepsilon_X \quad (4.24)$$

$$Y = \alpha_Y + \beta_Y T + \varepsilon_Y$$

$$Z = \alpha_Z + \beta_Z T + \varepsilon_Z$$

where α and β are additive and multiplicative bias terms (sometimes indicated as calibration constants), respectively, and ε is the zero-mean random error, with error variance σ_ε^2 .

The following underlying assumptions are required in TC method: 1) stationarity of signal and error statistics, 2) independency between the errors and the true state values (error orthogonality), and 3) independency between the errors of X, Y and Z (zero error cross-correlation).

There are two mathematically equivalent approaches to solve for the σ_ε^2 (Stoffelen, 1998). The first way is by cross-multiplying differences between the three a-priori rescaled data sets, while an alternative formulation is based on combinations of the covariances between the datasets; the former approach can be denoted as difference notation and the latter as covariance notation (Gruber et al., 2016a).

For the difference notation, one dataset is arbitrarily chosen as reference dataset (this will be dataset X for the following, as the choice of the reference not have impact on the error estimates), against which the other two datasets are linearly rescaled:

$$Y^X = \beta_Y^*(Y - \bar{Y}) + \bar{X} \quad (4.25)$$

$$Z^X = \beta_Z^*(Z - \bar{Z}) + \bar{X}$$

where the overbar denotes the mean value, while $\beta_Y^* = \beta_X/\beta_Y$ and $\beta_Z^* = \beta_X/\beta_Z$, which implies:

$$X = \alpha_X + \beta_X T + \varepsilon_X \quad (4.26)$$

$$Y^X = \alpha_X + \beta_X T + \beta_Y^* \varepsilon_Y$$

$$Z^X = \alpha_X + \beta_X T + \beta_Z^* \varepsilon_Z$$

that means the same systematic errors between dataset (so relative differences, expressed in a common data space, can be used to characterize random error variances). Given the underlying TC assumptions, the multiplicative rescaling parameters β_Y^* and β_Z^* can be derived by combining the three datasets in the following way:

$$\beta_Y^* = \frac{\beta_X}{\beta_Y} = \frac{\langle (X-\bar{X})(Z-\bar{Z}) \rangle}{\langle (Y-\bar{Y})(Z-\bar{Z}) \rangle} = \frac{\sigma_{XZ}}{\sigma_{YZ}} \quad (4.27)$$

$$\beta_Z^* = \frac{\beta_X}{\beta_Z} = \frac{\langle (X-\bar{X})(Y-\bar{Y}) \rangle}{\langle (Z-\bar{Z})(Y-\bar{Y}) \rangle} = \frac{\sigma_{XY}}{\sigma_{YZ}}$$

where $\langle \cdot \rangle$ indicates the temporal average and σ the covariance between two datasets, while the error variances can be estimated in the data space of the chosen scaling reference, i.e. X , as:

$$\sigma_{\varepsilon_X}^2 = \langle (X - Y^X)(X - Z^X) \rangle \quad (4.28)$$

$$\sigma_{\varepsilon_Y^X}^2 = \langle (Y^X - X)(Y^X - Z^X) \rangle$$

$$\sigma_{\varepsilon_Z^X}^2 = \langle (Z^X - X)(Z^X - Y^X) \rangle$$

and then converted back into their own data space as:

$$\sigma_{\varepsilon_Y}^2 = \frac{1}{(\beta_Y^*)^2} \sigma_{\varepsilon_Y^X}^2 \quad (4.29)$$

$$\sigma_{\varepsilon_Z}^2 = \frac{1}{(\beta_Z^*)^2} \sigma_{\varepsilon_Z^X}^2$$

In the covariance notation error variances can be directly estimated as:

$$\sigma_{\varepsilon_X}^2 = \sigma_X^2 - \frac{\sigma_{XY}\sigma_{XZ}}{\sigma_{YZ}} \quad (4.30)$$

$$\sigma_{\varepsilon_Y}^2 = \sigma_Y^2 - \frac{\sigma_{XY}\sigma_{YZ}}{\sigma_{XZ}}$$

$$\sigma_{\varepsilon_Z}^2 = \sigma_Z^2 - \frac{\sigma_{YZ}\sigma_{XZ}}{\sigma_{XY}}$$

The covariance notation does not require an a priori rescaling of the datasets, however it also allows for the direct estimation of the linear rescaling parameters using the relationships (4.27).

In summary, both the difference and the covariance notation can be used to estimate random error variances as well as rescaling parameters. In the difference notation, error variances are estimated within a common (arbitrarily chosen) reference data space, having the possibility of converting them back using the a priori estimated scaling parameters. The covariance notation, on the other hand, directly estimates unscaled error variances, which could be then scaled into a common data space using a posteriori (optionally) estimated scaling parameters.

An appropriate number of triplets must be available to obtain reliable estimates, regardless of the chosen approach; while more than 500 samples were recommended (Zwieback et al., 2012), several authors adopted a pragmatic threshold of 100 observations (e.g. Dorigo et al., 2010; Scipal et al., 2008b).

TC can be applied either to the original soil moisture absolute values (e.g. Scipal et al., 2008b) or to the anomalies from the long-term climatology (e.g. Dorigo et al., 2010). While using original values provides information on the capability of the soil moisture products in representing general temporal patterns of soil wetting and drying, the anomaly-based approach gives us more accurate information on the ability of the different datasets to capture single events of drying and wetting (e.g. due to rainfall). As a consequence, the anomaly-based approach tells us less about absolute deviations between datasets, e.g. like induced by a deviating seasonality.

TC can be regarded as a form of instrumental variable (IV) regression, where a third variable is used as an instrument to resolve the relationship between erroneous measurements of two variables. An alternative form of IV implementation is to use a lagged variable (LV) (i.e., a temporally shifted version) as third dataset in TC so that only two datasets are required (Su et al., 2014a). This particular implementation is possible when the geophysical variable of interest (e.g. soil moisture) had been sampled at time intervals shorter than their temporal correlation length, and under the condition of weakly autocorrelated errors in the lagged variable it can also provide consistent error variance estimates. The lag-based approach is suitable for practical circumstances where an adequate third collocated dataset is unavailable.

4.6.3 Error propagation

The error propagation method allows to evaluate the error variance of a model output through the propagation of error variances in input variables and model parameters.

Let output vector Y be determined through a model \mathcal{F} from the vector X , i.e. $Y = \mathcal{F}(X)$. X generally includes model inputs and parameters, with random errors that are fully characterized by the known error covariance matrix C_X , assuming a zero-mean Gaussian distribution and not excluding the possibilities of mutually correlated errors.

Random errors in X can be propagated through linear and nearly linear models by the analytic error propagation method, in order to characterize the output error covariance matrix, C_Y .

If \mathcal{F} is a linear model of the form $Y = AX + B$, then the output error covariance matrix is computed as:

$$C_Y = AC_X A^T \quad (4.31)$$

If, on the other hand, \mathcal{F} is a non-linear model, a linearization can be made by considering its first order Taylor approximation locally evaluated, that is:

$$Y = \mathcal{F}(X_0) + J(X - X_0) \quad (4.32)$$

having introduced the Jacobian matrix J , where $J_{ij} = (\partial f_i / \partial x_j)$ quantifies the influence that the j -th input in X has on the i -th output in Y , with f_i coming from \mathcal{F} in order to compute y_i . The first-order approximation can be considered valid if errors in X are sufficiently small that \mathcal{F} is linear over their likely ranges.

The output error covariance matrix can then be obtained as:

$$C_Y = JC_X J^T \quad (4.33)$$

For a scalar output variable $y = f(x_1, x_2, \dots, x_n)$ it gives the commonly encountered formula:

$$\sigma_\varepsilon^2(y) = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_\varepsilon^2(x_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \left(\frac{\partial f}{\partial x_i} \right) \left(\frac{\partial f}{\partial x_j} \right) \sigma_\varepsilon(x_i, x_j) \quad (4.34)$$

where σ_ε indicates the error covariance.

Error propagation is a general, conceptually simple and widely used technique for obtaining error characterisations, as in the conventional analytical form it only requires that J can be computed and C_X is known. For non-Gaussian error distributions and/or models that are significantly non-linear over the input error ranges, as well as when J is not easy to obtain due to model complexity and/or error correlations exist (non-diagonal terms in C_X) and are difficult to characterize, then numerical techniques such as the Monte Carlo approach could be employed.

5. ERROR PROPAGATION IN EXPONENTIAL FILTER APPROACH

5.1 Introduction

This chapter describes the first of two applications related to the assessment of satellite soil moisture preprocessing and error characterization practices.

The current application is made up of two studies subsequently carried out, having in common the topic of error propagation in the exponential filter approach.

The exponential filter is a widespread method to obtain a root-zone soil water index (SWI) from remotely-sensed surface soil moisture observations; however, the impact of some factors involved in SWI formulation, that can introduce inaccuracies in the outputs, have not been adequately detailed up to now.

In the following, some critical aspects of the exponential filter approach are showed, as well as the two mentioned studies are introduced. Then, the error propagation equations (expressed also in a practical recursive form) are proposed to analytically compute the SWI error variances. The study area, the materials and the experimental setup are described, being essentially common in the two studies. Finally, the specific results are analysed and discussed, highlighting the major outcomes.

5.2 Factors involved in SWI errors

The exponential filter proposed by Wagner et al. (1999) was previously described, here some main features are however briefly reported. The approach involves only one parameter, T , representing a characteristic time length, and implicitly taking into account the different factors that affect the temporal scale of soil moisture (layer depth, soil hydraulic properties, evapotranspiration etc.). The exponential filter smooths the SSM series to reproduce the soil moisture trend in deeper layers, according to the T value. The method has a greater ability in capturing the seasonal soil moisture behaviour rather than short time-scale fluctuations, and suffers from prolonged gaps in input data, up to be reinitialized.

The SWI found use in many applications where an estimate of the root-zone SM state is required, and even inside data assimilation frameworks, where the observation error

characterization is a critical factor. The great value of the exponential filter lies in the simplicity of its mathematical structure, which also implies the possibility to analytically propagate inputs and parameter error variances to the output SWI estimates. However, the uncertainty of SWI estimates has been poorly analysed.

SWI accuracy can be related to three factors. The first is the time distribution of input SSM data, i.e. the availability, for the single SWI estimate, of a sufficient number of measurements relative to the filter time scale. The exponential filter basically makes a weighted average of SSM inputs within a certain time interval, related to T , so the reliability of SWI output increases with an adequate SSM temporal distribution, and decreases when few inputs are available or in presence of SSM data gaps, due to revisit time but also to data masking procedures, e.g. for frozen or snow covered surface conditions. In literature some empirical criteria were proposed which address the problem. For example, referring to the original formulation (4.1), Wagner et al. (1999) suggested to calculate SWI value if there is at least one SSM measurement in the time interval $[t-T, t]$ and at least three measurements in the interval $[t-5T, t]$, while in Pellarin et al. (2006) all measurements taken within a period $[t-3T, t]$ are considered if at least four measurements have been recorded within the most recent time period $[t-T, t]$. These kind of criteria cannot be applied with the widely used recursive formulation (4.2), proposed by Albergel et al. (2008); however, information about the availability of SSM measurements used for SWI estimate is implicitly contained in the gain value, K_n , initialized to 1 and in general tends toward unity in the presence of extensive temporal data gaps (relative to the T value). So a relatively low value of the gain K_n indicates the presence of an adequate number of SSM observations in the temporal range of interest for the SWI computation. For example, in the Copernicus Global Land Service SWI daily products a quality indicator (*qflag*) is introduced to take in account the number of available SSM measurements really used for SWI calculation; this quality flag is updated with a formulation similar to (4.3) but resulting inversely proportional to K_n .

As a second factor, SWI estimates are also influenced by random errors in SSM inputs. As showed in the previous chapter, an estimate of SSM error variance can be obtained in different ways, usually assuming its time stationarity. However, using an error propagation (EP) approach for SWI error characterization, it is also possible to exploit the time-variant SSM error variance estimates, provided in some satellite products and computed by EP through SSM retrieval algorithm, which are typically underemployed.

Finally, there is the uncertainty on T parameter, which determines the SWI dynamics for a given SSM input time series. T represents different static and dynamic processes that affect the temporal scale of soil moisture (layer depth, soil hydraulic properties, evapotranspiration, infiltration etc.) but is typically approximated as a constant. The potential impact of T error on SWI estimates under different conditions has also been poorly studied.

The use of an EP approach allows to take in account the effects of these different factors on the single SWI estimates, so considered to be characterized by a time-variant error variance; the issue was addressed here through the two studies described below. However, EP do not consider the intrinsic model error, which also contributes to the output error budget; this can lead to unrealistic error estimates, especially in the case of increasingly accurate inputs and model parameters.

The first study involved the evaluation of a satellite SSM product from an ongoing mission, by local comparison with in situ measurements collected in Italy and hosted on the International Soil Moisture Network (ISMN). Given the limited availability of ground data, in term of both time coverage and observation depth, the ASCAT SSM retrievals were chosen as satellite dataset, with the exponential filter used to propagate the remotely sensed surface states to the in situ measurements layers. Ten sites were selected with a sufficient observation period, also used in previous satellite SM evaluation studies and assumed to be representative at coarse scale. The capability of ASCAT-derived SWI to reproduce observed SM values in root-zone (implicitly considered as truth) was here analysed through commonly used performance metrics. Then, the effect of different data masking procedures were considered, based on controls on additional data fields available in SSM product. ASCAT retrievals also include standard error estimates, indicated as ‘soil moisture noise’, whose information content was rarely considered.

A first, immediate use of ‘soil moisture noise’ (from now referred to as SSM noise) data is in a quality check framework, applying a threshold value to mask less reliable data: this can be a proper choice when SSM is used to update continuous model SM estimates in a data assimilation system. However, using the exponential filter approach, the attempt to propagate the SSM standard error information to SWI estimates appears more appropriate.

The first approach (i.e. threshold on SSM noise) implies the loss of SWI updates and the introduction of temporal gaps in input series that as seen can be critical for the successive SWI estimates. Then, masked SSM data anyway contain some informative content that in this way will end up lost; moreover, due to its mathematical structure, exponential filter tends to suppress high random errors, proportionally to the value assumed by T , so limiting the consequence of using less reliable data.

In alternative, by propagating surface errors to root zone estimates without applying a priori masking, the impact of SSM standard error on SWI outputs can still be monitored. For this purpose, a simplified error propagation scheme (as it do not consider the uncertainty in the model parameter T) was proposed, aimed at estimating an index (referred to as SWI noise) which takes in account both standard errors and availability of the input SSM data actually used for each single SWI estimate. The usefulness of this approach, compared to the a priori SSM noise thresholds, has been verified by applying some comparable a posteriori SWI noise thresholds, in order to investigate the relative capabilities to remove SWI data that do not fits well in situ observations. In this sense, the use of SWI noise information outperforms the a priori SSM noise thresholds. However, the results of this first study were influenced by seasonal patterns in SSM error variance estimates which characterize the employed ASCAT product release (DR2015), limiting the analysis to summer periods where a good agreement exists between SWI and ground observations, mainly due to the limited rainfall events. According to Draper et al. (2013), these patterns were associated with the seasonal cycle in the sensitivity of SSM retrieval model parameters to various errors, rather than with the high noise level of the input backscatter measurements.

In the second study the successive ASCAT product release was used (DR2016), where seasonal patterns are considerably reduced; moreover, SWI error variances are estimated by error propagation equations including the effect of exponential filter parameter uncertainty. A preliminarily assessment of the computed error estimates was carried out similarly to the first study, by data masking in the comparison between satellite-derived SWI and root-zone in situ measurements. In this case the EP scheme showed skills in detecting potentially less reliable SWI values in the study sites, improving the agreement with reference data with respect to the configuration without quality controls. The proposed approach not only provided simultaneous estimates of

time-variant SWI error variance, but also allowed a better understanding of the exponential filter shortcomings.

When T is low, SWI outputs are more sensitive to SSM temporal gaps and random errors (this is properly taken in account by the terms included in the SWI noise index). High T values can instead led to excessive smoothed SWI time series, with short-scale SSM variations being suppressed during the weighted average that generally involves many observations, and this can be associated to the known limitations of exponential filter in modelling the short-time fluctuations observed in root zone SM. With regard to the corresponding SWI error variance estimates obtained by EP, the filter structure tends to average and compensate random errors in the underlying SSM data, resulting in an inputs contribution to error budget that appears to be inversely proportional to T (or more appropriately to the number of SSM measurements actually used for the SWI calculation). On the other hand, the error contribution due to uncertainty considerations on T parameter can be remarkable during SM transitions, also for high T values. The observed root-zone short-time fluctuations are usually associated to variations in surface SM data, that is the circumstance where the SWI output is more sensitive to T value, which in turn determines how much to weigh these recent inputs that differ from the previous ones. Approximation in the constant T estimate can explain the short time-scale deviations between the reference data and the smoothed SWI. Thus, limitations due to the simple model structure of exponential filter can be partially taken in account through considerations on its parameter uncertainty effect.

5.3 Soil water index error variance estimates

Referring to Equation (4.1) and assuming that errors in SSM measurements and in parameter T are normally distributed and uncorrelated, the random error variance of SWI at time t_n , $\sigma^2(\text{SWI}_n)$, can be estimated by the error propagation approach as:

$$\sigma^2(\text{SWI}_n) = \sum_{i=1}^n \left(\frac{\partial \text{SWI}_n}{\partial \text{SSM}_i} \right)^2 \sigma^2(\text{SSM}_i) + \left(\frac{\partial \text{SWI}_n}{\partial T} \right)^2 \sigma^2(T) \quad (5.1)$$

where:

$$\frac{\partial \text{SWI}_n}{\partial \text{SSM}_i} = \frac{e^{-\frac{t_n-t_i}{T}}}{\sum_i^n e^{-\frac{t_n-t_i}{T}}} \quad (5.2)$$

$$\frac{\partial \text{SWI}_n}{\partial T} = \frac{1}{T^2} \cdot \frac{\sum_i^n \text{SSM}_i \cdot e^{-\frac{t_n-t_i}{T}} \left(\sum_j^n e^{-\frac{t_n-t_j}{T}} (t_j-t_i) \right)}{\left(\sum_i^n e^{-\frac{t_n-t_i}{T}} \right)^2} \quad (5.3)$$

Equation (5.1) can be written in recursive form as:

$$\sigma^2(\text{SWI}_n) = \Delta_n^2 + \left(\frac{\partial \text{SWI}_n}{\partial T} \right)^2 \sigma^2(T) \quad (5.4)$$

where:

$$\Delta_n^2 = K_n^2 \sigma^2(\text{SSM}_n) + (1 - K_n)^2 \Delta_{n-1}^2 \quad (5.5)$$

$$\frac{\partial \text{SWI}_n}{\partial T} = \frac{K_n}{T} \left[G_n (\text{SWI}_{n-1} - \text{SWI}_n) + e^{-\frac{t_n-t_{n-1}}{T}} \frac{T}{K_{n-1}} \frac{\partial \text{SWI}_{n-1}}{\partial T} \right] \quad (5.6)$$

having defined G_n as:

$$G_n = e^{-\frac{t_n-t_{n-1}}{T}} \left(G_{n-1} + \frac{1}{K_{n-1}} \cdot \frac{t_n-t_{n-1}}{T} \right) \quad (5.7)$$

while K_n comes from Equation (4.3). Initial value Δ_0 is set to $\sigma(\text{SSM}_0)$, while $(\partial \text{SWI}_0 / \partial T)$ and G_0 are set to 0.

In the first study, the error variance of the exponential filter parameter, $\sigma^2(T)$, was not considered; thus, $\sigma(\text{SWI})$ was coincident with the term Δ (and referred to as SWI noise), that considers only the effect of SSM uncertainties and also takes into account the effect of possible prolonged temporal gaps in input SSM data, related on the T value, as this information is implicitly contained in the gain value K_n .

In the second study the term $\sigma(T)$ was included in the analysis, and so the jacobian term $(\partial \text{SWI} / \partial T)$ that assumes high values proportionally to the latest SSM inputs variability on a time-scale related to the T parameter, which reflects in SWI value significant changes, e.g. state transition from dry to wet conditions.

5.4 Case study

5.4.1 Study areas and in situ observations

The in situ soil moisture dataset used in the following investigations comes from ten sites located in 3 Regions (Calabria, Campania and Umbria) across Italy (Fig. 5.1). The selected sites are included in the International Soil Moisture Network (ISMN); measurements collected in the ISMN database are distributed with a temporal resolution of one hour. Selected sites have at least three years of observation during the period 2007-2014, at depths ranging from 20 to 60 cm.

The main characteristics for each measurement location are given in Table 5.1. Volumetric soil moisture (m^3m^{-3}) is measured through electromagnetic sensors, specifically ML2X Theta probes, with the exception of site 10 where a TRASE-BE device is installed. Nine out of ten sites are part of the Regional hydro-meteorological monitoring networks for Civil Protection activities aimed at flood and landslide risk mitigation managed by the “Centro Funzionale Decentrato” of Calabria (sites 1-5), Campania (site 6) and Umbria (sites 7-9) respectively. Site 10 corresponds to one (Perugia - Field 1) of the two fields which constitute the HYDROL-NET network, located in Umbria and settled up by the University of Perugia (Morbidelli et al., 2014). All sites mainly experience a Mediterranean semi-humid climate, with rainy periods coinciding with fall and winter months, while summers are hot and dry, strongly affecting the seasonal runoff regime of streams.

Soil texture and organic carbon contents from Harmonized World Soil Database v1.1 are reported by ISMN. According to the soil characteristics and USDA (United States Department of Agriculture) classification, the texture class of soil for almost all sites is loam, with the exception of sites 4 and 5 where soil is classified as sandy loam and clay loam, respectively.

ISMN datasets are subjected to an automated quality control to flag spurious observations (Dorigo et al., 2013). The selected dataset was found to be substantially affected by: 1) values exceeding the expected saturation water content; 2) spikes, typically defined as unanticipated significant rises or drops lasting only one time step, due for example to temporary sensor failure or reduced current supply; 3) positive jumps, which typify sudden increases in the registered soil moisture value, usually from one time unit to the next, and 4) constant high values usually corresponding to situations when the soil moisture content exceeds the upper limit of the sensor sensitivity. Each flagged observation was visually examined and compared to neighbouring observations (in depth and time) and, when necessary, to available rainfall time series, to remove the most probable spurious observations. Only spikes at sites 1 and 2, which did not show a link with rainfall patterns, were removed from the dataset. Other typologies of flagged data were retained for the subsequent investigations, as they were considered sufficiently reliable. On the other hand, although not flagged, 30 cm depth measurements for the second half-year of 2012 for site 1 and 20 cm depth

measurements for the first half-year of 2010 for site 9, were removed as they are generally characterized by high implausible values.

Data from selected sites have already been partially used in previous studies aimed at evaluate the satellite-based SWI (Brocca et al., 2010b, 2011a), where an analysis of in situ observations was performed to establish the representativeness at footprint scales. As part of the ISMN, measurements from the four considered Italian networks were also used in global studies aimed to validate the ASCAT SWI (Paulik et al., 2014) and to analyse the coarse scale representativeness of in situ data (Gruber et al., 2013). In the last study, random errors in coarse-scale representation are characterized by triple collocation, comparing in situ point-scale data with two independent coarse-scale datasets; as showed in Fig 5.2, “Calabria” sites show lower error values than stations in other considered networks. Thus, the selected study sites are considered sufficiently suitable for comparison with coarse-scale SWI, although with different levels of quality in representing soil moisture at footprint scales.

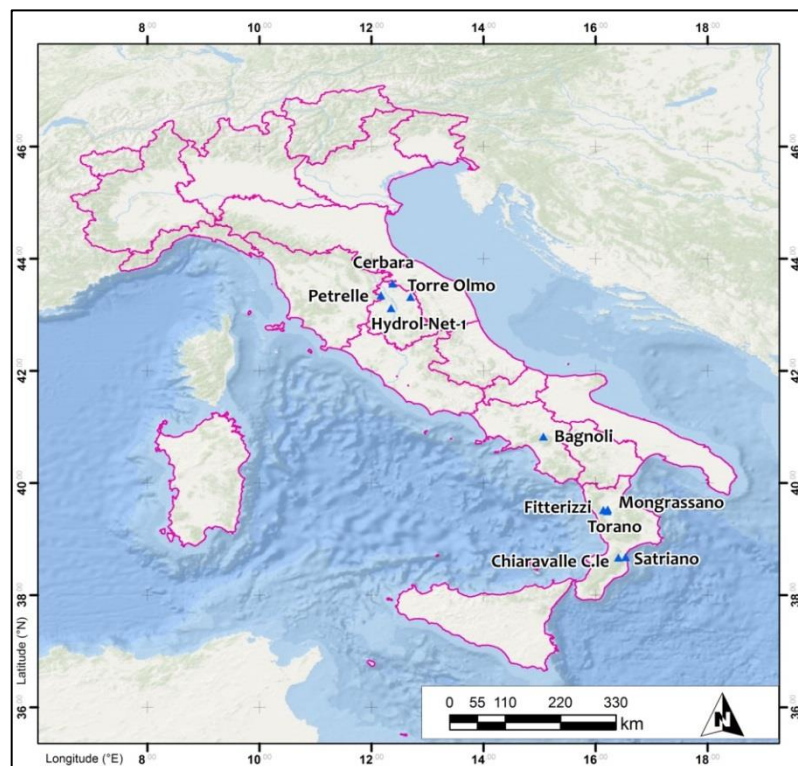


Figure 5.1. Location of the study sites in Italy.

Table 5.1. Main characteristics of study sites.

Code	Station name	Network	Latitude (°)	Longitude (°)	Depth (cm)	Observation period
1	Fitterizzi	Calabria	39.52	16.14	30, 60	1 Jan 2001 - 31 Dec 2012
2	Mongrassano	Calabria	39.53	16.22	30, 60	1 Jan 2001 - 16 May 2011
3	Torano	Calabria	39.50	16.21	30, 60	1 Jan 2001 - 31 Dec 2012
4	Chiaravalle C.le	Calabria	38.67	16.41	30, 60	1 Jan 2001 - 31 Dec 2012
5	Satriano	Calabria	38.68	16.54	30, 60	1 Jan 2001 - 31 Dec 2012
6	Bagnoli	Campania	40.83	15.07	30	1 Dec 2000 - 21 Nov 2012
7	Cerbara	Umbria	43.56	12.38	20, 40	30 Oct 2009 - 31 Jul 2014
8	Petrelle	Umbria	43.35	12.17	20, 40	30 Oct 2009 - 31 Jul 2014
9	Torre Olmo	Umbria	43.32	12.70	20, 40	23 Sep 2009 - 31 Jul 2014
10	Perugia (Field1)	HYDROL-NET	43.12	12.35	25, 35	1 Jan 2010 - 31 Dec 2013

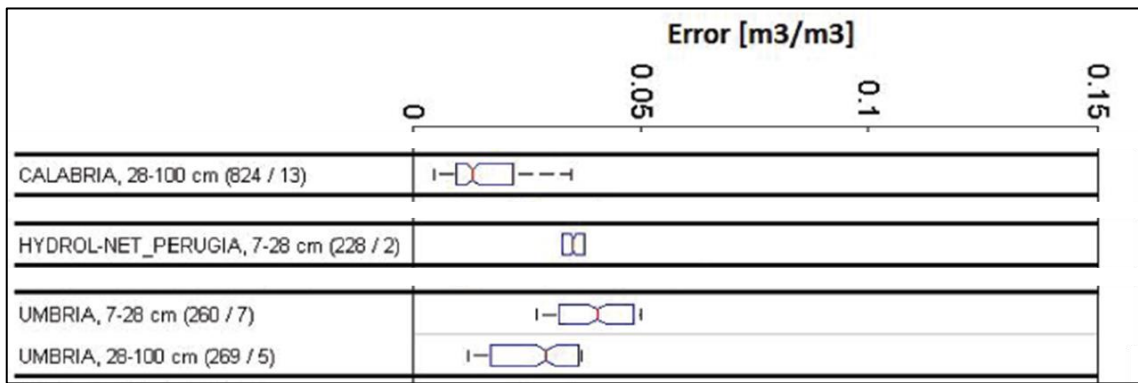


Figure 5.2. Figure adapted from Gruber et al. (2013). Box-Whiskers plots summarizing the TC results in the different networks. Single standard error estimates were obtained for each in situ time series (i.e. for a given station at a given depth). Values in brackets show the average number of triplets used for the error estimate/the total number of error estimates represented in the boxplot.

5.4.2 Remote sensing observations and quality flags

The SSM time series derived on WARP5 discrete global grid (grid spacing of 12.5 km) from ASCAT backscatter measurements via the offline WARP processing chain (Naeimi et al., 2009) were employed. ASCAT SSM unit is degree of saturation, and several additional data fields are provided, including standard error estimates computed by an EP approach and denoted as ‘soil moisture noise’. Two releases of the product were used, both distributed within the framework of the H-SAF project, and specifically the DR2015 version (distributed as H25 product) in the first study and the DR2016

version (H109 product) in the second one. Only data from the ASCAT sensor on board the MetOp-A satellite were here considered.

The nearest neighbour technique was used to obtain a correspondence between ASCAT grid points and in situ stations (Table 5.2), and the hourly in situ measurements associated with the closest temporal match to the acquisition time of SSM data were selected.

Table 5.2 - Selection of ASCAT time series datasets, identified by WARP5 Grid Point Index (GPI).

In situ stations		Corresponding ASCAT time series datasets				
Code	Station name	GPI	Latitude (°)	Longitude (°)	Topographic complexity (%)	Wetland fraction (%)
1	Fitterizzi	2069259	39.51	16.13	20	0
2	Mongrassano	2069263	39.51	16.28	20	0
3	Torano	2069263	39.51	16.28	20	0
4	Chiaravalle C.le	2034369	38.72	16.39	16	0
5	Satriano	2034373	38.72	16.53	15	34
6	Bagnoli	2128143	40.86	15.12	19	0
7	Cerbara	2242295	43.56	12.38	10	0
8	Petrelle	2232959	43.34	12.18	10	0
9	Torre Olmo	2232971	43.34	12.64	12	0
10	Perugia (Field1)	2223595	43.11	12.29	7	5

ASCAT SSM data can have different levels of intrinsic quality, due to various factors involved in the processing that can affect the accuracy of retrievals. Thus, SSM products include several additional data fields, containing attributes and flags which can be useful in judging the reliability of remotely sensed estimates and should be taken into account depending on the specific application. The attributes considered in this study are briefly described below.

ASCAT datasets include wetland fraction and topographic complexity advisory flags. These two static indicators are derived from external datasets and are useful in detecting areas where the retrieval method does not work properly because of the influence from open water or orographic effects in the satellite footprint. In several global data masking applications, observations at satellite grid points with a wetland fraction larger than 15% or topographic complexity larger than 20% were considered as inaccurate (Scipal et al., 2008a; Dharssi et al., 2011; Paulik et al., 2014).

The surface state flag (SSF) returned by the algorithm developed by Naeimi et al. (2012) indicates the temporary surface conditions and detects any backscatter measurements from snow covered or frozen land surfaces which can lead to an incorrect determination of soil moisture. Soil moisture estimates occurring when soil is not unfrozen are as a rule rejected.

The processing flag comprises information on internal quality checks and specific processing details (H-SAF, 2014). In some cases, the extrapolated backscatter at the reference incidence angle can exceed the dry or the wet backscatter reference. As a result, the surface soil moisture is outside the nominal range of 0 - 100%. However, when it is still within certain limits, it is set to the extremes of 0% or 100% during the processing. In these cases, the processing flag is not set to its default value, namely 0, thus highlighting restrictions in data interpretation and use.

The orbit direction attribute provides information on the satellite passing time. The use of morning surface observations is considered more appropriate to calculate the SWI in deeper layers, because the soil is most likely to be in hydrologic equilibrium conditions, avoiding daytime decoupling due to evapotranspiration (Jackson, 1980, as quoted in Albergel et al., 2008; Wanders et al., 2012). Thus, the highest correlations between ground-based measurements and SWI have been experimentally obtained in several studies using only morning satellite measurements (e.g., Wagner et al., 1999). In ASCAT data, morning passes are associated with a descending orbit.

The field 'soil moisture noise', here referred to as SSM noise to avoid ambiguity, contains an estimate of the standard error in soil moisture retrievals, expressed in degrees of saturation [%]. Errors from the scatterometer observations, as well as the uncertainties associated with model parameters are taken into account and propagated through the retrieval algorithm to determine SSM noise. High values of average SSM noise were observed in areas where the retrieval algorithm is not expected to work

properly, e.g. areas covered by dense vegetation (Naeimi et al., 2009). In the literature, depending on the specific application, threshold values on SSM noise usually ranging from 5% to 30% are fixed for different ASCAT soil moisture products (near real time or offline time series) and processor versions, to detect areas where retrieval methods do not work well or to mask single spurious SSM observations (e.g. Mahfouf, 2010; Dharssi et al., 2011; Draper et al., 2011, 2012, 2013; Paulik et al., 2014).

5.4.3 Experimental setup

SWI time series were computed from SSM datasets, estimating the optimal T value that maximizes the correlation coefficient r between in situ and satellite-based soil moisture observations, where r is generally dominated by the strong seasonal pattern in time series; after linearly rescaling SWI with the mean and variance matching technique, the root-mean-square-difference (RMSD) was calculated.

In the first study, the correlation coefficient r_{anom} was also computed by considering in situ and SWI anomalies, in order to evaluate the capability of ASCAT-derived SWI to capture not only the seasonal trend but also the short-term variability in soil moisture. The dimensionless soil moisture anomalies, $\theta_{\text{anom}}(t)$, were evaluated as suggested by Albergel et al. (2009):

$$\theta_{\text{anom}}(t) = \frac{\theta(t) - \text{mean}(\theta(t-17:t+17))}{\text{std}(\theta(t-17:t+17))} \quad (5.8)$$

where $\theta(t)$ is the in situ observation or SWI value at time t , while $\text{mean}(\theta(t-17:t+17))$ and $\text{std}(\theta(t-17:t+17))$ are the mean and the standard deviation of $\theta(t)$, respectively, for a time window of 35 days (corresponding to five weeks) centred on time t and defined by $t \pm 17$ days. The $\theta_{\text{anom}}(t)$ is computed if there are at least five measurements in the sliding window.

5.5 Application of EP equations neglecting the error in SWI parameter

In the first study, a simplified SWI standard error estimate was used, referred to as SWI noise, which implicitly takes into account both the noise of SSM data actually used for the SWI calculation and their availability in a relatively sufficient number. SWI noise

was addressed to be used in a quality check framework, in the context of a more general evaluation of the impacts of different data masking procedures.

For each station, in situ measurements (indicated as OBS plus corresponding depth in cm, and as OBS_{mean} when referring to the mean of more depths) were compared to SWI estimated under different configurations. The comparison results are first illustrated (Sect. 5.5.1) by considering standard controls based on available ASCAT attributes. The outcomes derived from considering a posteriori control on the SWI noise are then discussed in Sect. 5.5.2.

Regarding the SSM static advisory flags, according to the previously mentioned thresholds (15% for the wetland fraction and 20% for the topographic complexity), as can be inferred from Table 5.2, only the grid point associated with site 5 (which is located near the Ionian Sea coast) is characterized by a wetland fraction over the assumed threshold. However, site 5 was retained in the following analysis, considering that the corresponding SSM time series also have relatively low values of SSM noise. In addition, the closest grid point with appropriate values on both advisory flags could not be considered representative as it is too far from the site location.

5.5.1 Data masking on ASCAT product attributes

Four different data masking configurations on remote sensing data were here considered. Specifically, the effects of using the processing flag and different orbit directions to mask SSM observations before the SWI computation were evaluated in terms of the agreement between satellite-derived and in situ soil moisture time series. The adopted data masking configurations are identified as follows:

DM0 - is a reference basic configuration where no controls are carried out;

DM1 - only SSM values corresponding to the default value for the processing flag (i.e. processing flag value equal to 0) were retained in the SWI computation;

DM2 - considers only SSM observed in the morning (satellite descending orbit);

DM3 - considers only SSM observed in the evening (satellite ascending orbit).

In each of these configurations, SSM retrievals with invalid SSF values have been preliminarily masked. Regarding no-default processing flag values, in Calabria and Campania datasets we found almost exclusively situations of surface soil moisture below the dry reference value and artificially set to 0%, that were mainly concentrated

during the summer months and particularly in July and August. In the grid point close to station 5 there are twice as many occurrences than at the other remote sensing datasets and they are significantly distributed over a longer period (from April to October). In Umbria region, instead, there are also situations of SSM retrievals above the wet reference, albeit to a lesser extent, that are concentrated during winter months; the frequency of soil moisture estimates below the dry reference is similar to abovementioned sites, but distributed over a longer summer-centred period, especially in site 10 which has a similar behaviour to site 5.

The results obtained under the assumed DM configurations at the most superficial observation depths are illustrated for each site in Table 5.3. The rescaled SWI is indicated as SWI^* . Figures 5.3 and 5.4 report in situ and satellite derived SWI^* time series and scatterplots for six of the selected sites, referring to DM0 configuration at the most superficial observation depth. The data generally show a good ability of the ASCAT derived SWI to reproduce temporal patterns and absolute values of ground observations. Results regarding the four remaining stations substantially confirm the same outcomes.

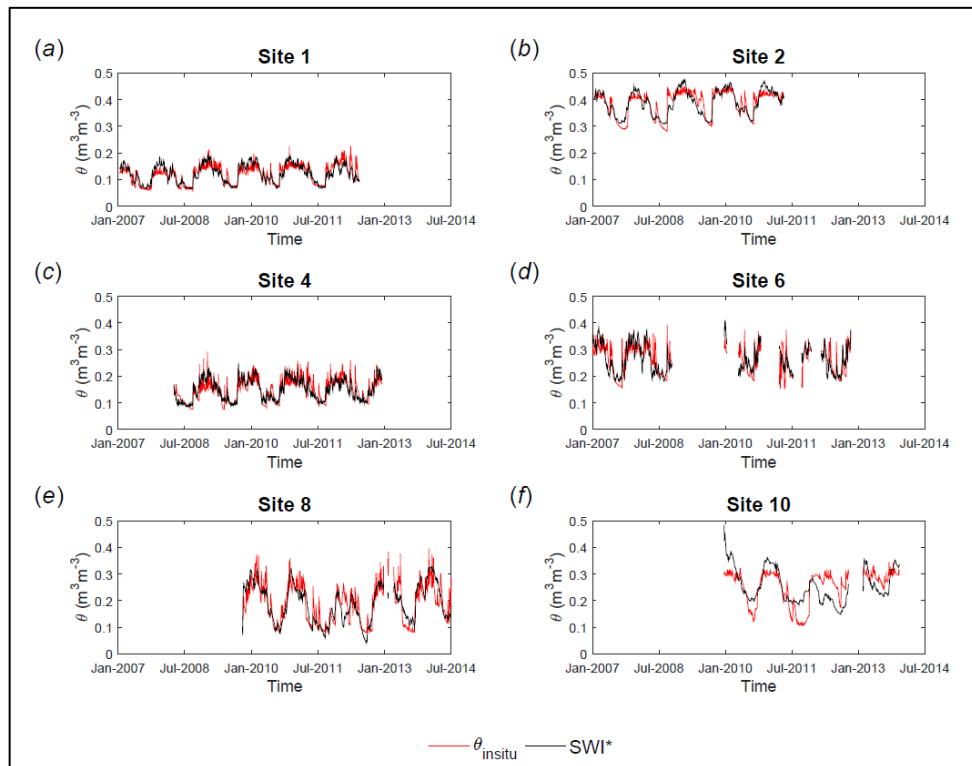


Figure 5.3. Time series of satellite-based SWI^* and in situ measurements at the most superficial observation depth, for a selection of six sites in DM0 configuration.

Table 5.3. Results of the different “a priori” data masking procedures, in terms of exponential filter T value and performance scores. N_{obs} and N_{anom} refer to the number of observations and anomalies respectively.

		T (days)	r	RMSD (m^3m^{-3})	r_{anom}	N_{obs}	N_{anom}			T (days)	r	RMSD (m^3m^{-3})	r_{anom}	N_{obs}	N_{anom}
Site 1	DM0	6	0.851	0.019	0.522	1565	1539	Site 6	DM0	5	0.745	0.038	0.539	933	897
OBS ₃₀	DM1	7	0.815	0.020	0.536	1428	1402	OBS ₃₀	DM1	5	0.727	0.039	0.529	874	838
	DM2	7	0.855	0.019	0.554	779	765		DM2	9	0.757	0.038	0.463	447	428
	DM3	9	0.836	0.020	0.498	786	772		DM3	5	0.723	0.040	0.562	486	468
	DM0	15	0.864	0.025	0.522	1266	1239		Site 7	DM0	4	0.718	0.061	0.497	1601
OBS ₃₀	DM1	17	0.823	0.025	0.524	1145	1118	OBS ₂₀	DM1	4	0.712	0.060	0.484	1523	1485
	DM2	16	0.865	0.025	0.554	621	607		DM2	5	0.712	0.062	0.464	799	780
	DM3	17	0.855	0.026	0.494	645	629		DM3	5	0.710	0.062	0.482	802	781
	DM0	10	0.884	0.023	0.464	1768	1740		Site 8	DM0	18	0.823	0.042	0.415	1536
OBS ₃₀	DM1	12	0.870	0.024	0.483	1620	1592	OBS ₂₀	DM1	18	0.820	0.041	0.417	1498	1460
	DM2	11	0.884	0.023	0.476	874	859		DM2	21	0.815	0.043	0.418	805	785
	DM3	13	0.870	0.025	0.443	894	879		DM3	17	0.824	0.041	0.399	731	711
	DM0	4	0.831	0.023	0.452	1473	1451		Site 9	DM0	20	0.670	0.017	0.302	1390
OBS ₃₀	DM1	3	0.813	0.024	0.464	1390	1368	OBS ₂₀	DM1	18	0.656	0.017	0.323	1298	1263
	DM2	4	0.827	0.024	0.438	728	715		DM2	19	0.681	0.017	0.356	688	667
	DM3	7	0.822	0.024	0.388	745	729		DM3	21	0.658	0.017	0.312	702	685
	DM0	6	0.841	0.024	0.304	1768	1740		Site 10	DM0	38	0.585	0.057	0.419	1048
OBS ₃₀	DM1	8	0.807	0.026	0.313	1506	1478	OBS ₂₅	DM1	36	0.622	0.055	0.407	957	930
	DM2	7	0.833	0.025	0.337	878	863		DM2	39	0.591	0.057	0.443	526	512
	DM3	10	0.826	0.025	0.227	890	875		DM3	39	0.577	0.057	0.394	522	508

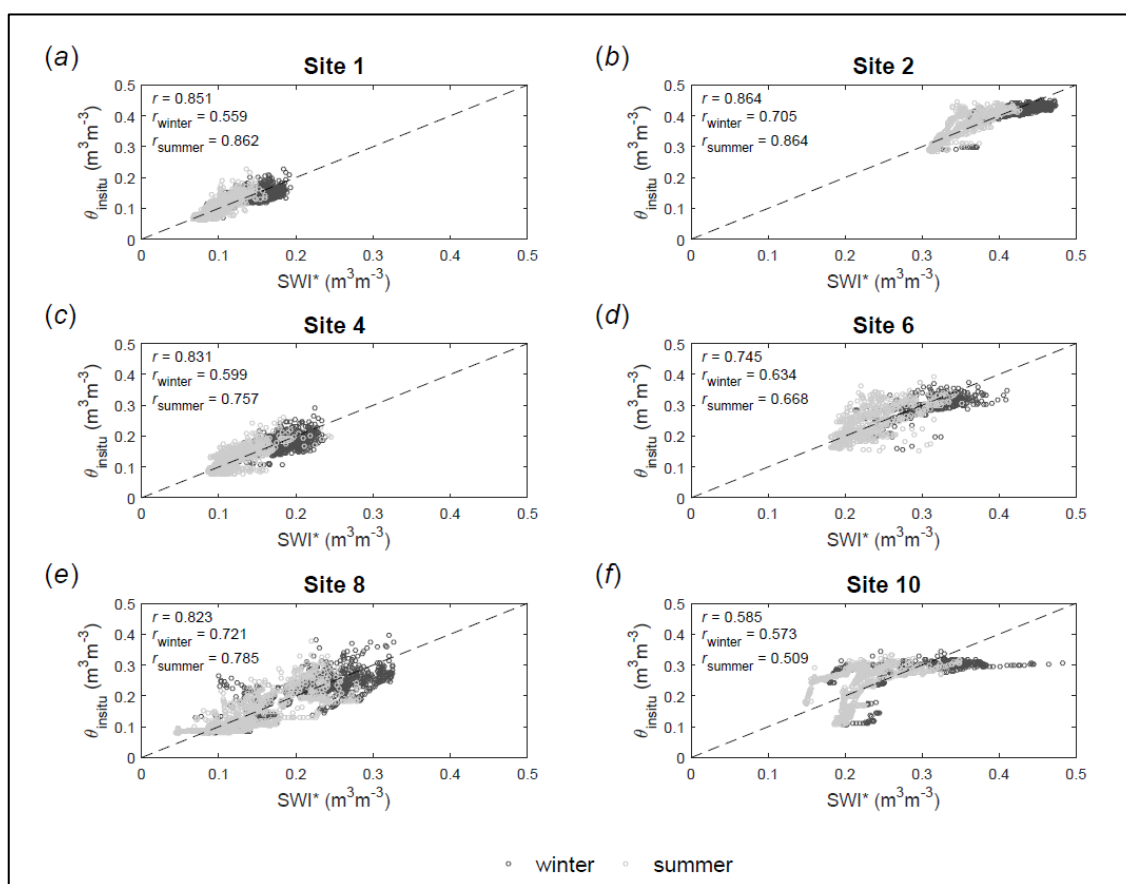


Figure 5.4. Scatterplots of satellite-based SWI* and in situ measurements for a selection of six sites in DM0 configuration. Correlation coefficients r , r_{winter} and r_{summer} are referred to measurements occurred in the entire observation period or in winter (from October to March) or summer (from April to September) respectively.

Optimal T values in Table 5.3 are consistent with results obtained in the literature for similar soil thicknesses (e.g. Wagner et al., 1999; Albergel et al., 2008, 2009; Brocca et al., 2010a, 2010b, 2011a; Ford et al., 2014) and show relatively small variations under different masking criteria.

About absolute values performance scores, the obtained r and RMSD values generally indicate a good agreement between SWI and ground data in the DM0 configuration. In sites 1-8, r is always over 0.7, while RMSD ranges between 0.019 and 0.061 $m^3 m^{-3}$. Sites 9 and 10 show lower correlation coefficient values (0.670 and 0.585 respectively). In Umbria and HYDROL-NET networks the performance scores are generally worse than in the other sites: this could be attributed to higher in situ errors in coarse-scale representation, rather than a local inadequacy of satellite-derived soil moisture, as

evaluated in Gruber et al. (2013) introducing a third independent large scale dataset (Fig. 5.2).

Based on the anomalies of in situ and estimated SWI, r_{anom} was also computed. The r_{anom} values are lower than the absolute time series correlations (varying between 0.302 and 0.539 for the DM0 configuration at the most superficial depth), similarly to other studies (e.g., Brocca et al., 2010b, 2011; Paulik et al., 2014) and do not show any particular geographical pattern. However, as for absolute time series, p -value (measuring the probability of null correlation) indicate that all the correlations obtained for anomalies time series are statistically significant at the 0.05 level of significance. The r_{anom} is found to sensibly decrease with depth in all sites (exception being sites 9 and 10 where it slightly improves), in line with the outcomes of Paulik et al. (2014). In this sense, the decrease in the anomaly correlation with depth is due to the effect of the exponential filter with higher T values, which tends to smooth short-time variations observed in SSM series.

A calibration of the exponential filter parameter was also performed (not shown here for the sake of brevity) by maximizing r_{anom} : generally, significantly lower T values are obtained; nevertheless, r_{anom} do not show noteworthy variations compared to values in Table 5.3, implying a low ability of the exponential filter approach to mimic the less prominent short-term fluctuations rather than seasonal variability. As an example of this evidence, the results obtained in site 1 for DM0, show that optimizing r_{anom} instead of r implies variations in T from 6 to 3 days, while r_{anom} increases from 0.522 to 0.531 and r decreases from 0.851 to 0.841.

It is worth noting that the goodness of agreement between satellite-derived and in situ measurements in maximizing r strongly depends on the seasons of the year. Splitting the sample into a summer period, April - September, and a winter period, October - March, a better correlation is generally obtained considering the warm months (indicated as r_{summer} in Figure 5.4) compared to the one calculated considering the rest of the year (i.e., r_{winter}). The correlation coefficients r_{summer} and r_{winter} are in most cases lower than r .

By performing an analogue analysis on the anomalies time series, it is not possible to identify a similar common behaviour; moreover, the best 'seasonal' correlation coefficient is frequently higher than r_{anom} calculated over the entire observation period.

The effects of different masking criteria were evaluated taking the DM0 configuration as a reference. Masking SSM observations with the no-default processing flag value (DM1) prior to the SWI estimation has a negative effect in terms of performance metrics over all the analysed sites with the exception of site 10. In fact, discarded observations are almost all concentrated in summer periods, where although probably not quantitatively accurate, neglected data provide a qualitative indication of dry conditions, and their informational content has been found to improve SWI estimates. Concerning the orbit direction, using only satellite morning passes (DM2) for some sites (sites 1, 2, 3, 6, 9 and 10 in Table 5.3) slightly improves the performance scores, but always at the cost of halving the number of obtained SWI estimates and degrading the temporal coverage of the product. The DM3 control, considering only ascending passes, as expected, results in lower performance compared to DM0 and DM2, with limited exceptions, as some scores in sites 6 and 8.

5.5.2 Data masking based on SSM noise and SWI noise

The effect of removing the SWI estimates characterized by low accuracy through a threshold on SWI noise was examined. The analysis was conducted by comparing three different approaches:

- SSM-1) Setting a threshold on SSM noise and then identifying the optimal T according to the best r performance of SWI in reproducing the corresponding subset of in situ measurements. Different SSM noise thresholds are considered, varying from its [median value+1] to the [maximum value-1] encountered in the time series, that correspond to a specific number of discarded data with higher noises that are not employed in the SWI estimation;
- SWI-2) Generating a SWI noise time series according to Equation (5.5), having assumed the same optimal T value obtained in SSM-1. SWI noise thresholds were ‘a posteriori’ assumed, corresponding to SWI series with the same amount of discarded data resulting in the previous approach from the application of SSM noise thresholds;
- SWI-3) Recalibrating the optimal exponential filter parameter T according to a SWI time series obtained by setting a threshold on SWI noise to remove the same number of SWI data as the corresponding ‘a priori’ control on SSM. The T

parameter is optimized considering only SWI values characterized by noise below the threshold on the basis of the correlation coefficient with in situ measurements.

To illustrate the smoothing effect of the exponential filter and its dependency on the optimized T value, Figure 5.5 shows for site 4, the time series of SSM, SWI, SSM noise and SWI noise for two different T values. The SWI noise seasonal pattern reflects the seasonality in SSM noise whose highest values occur in summer months.

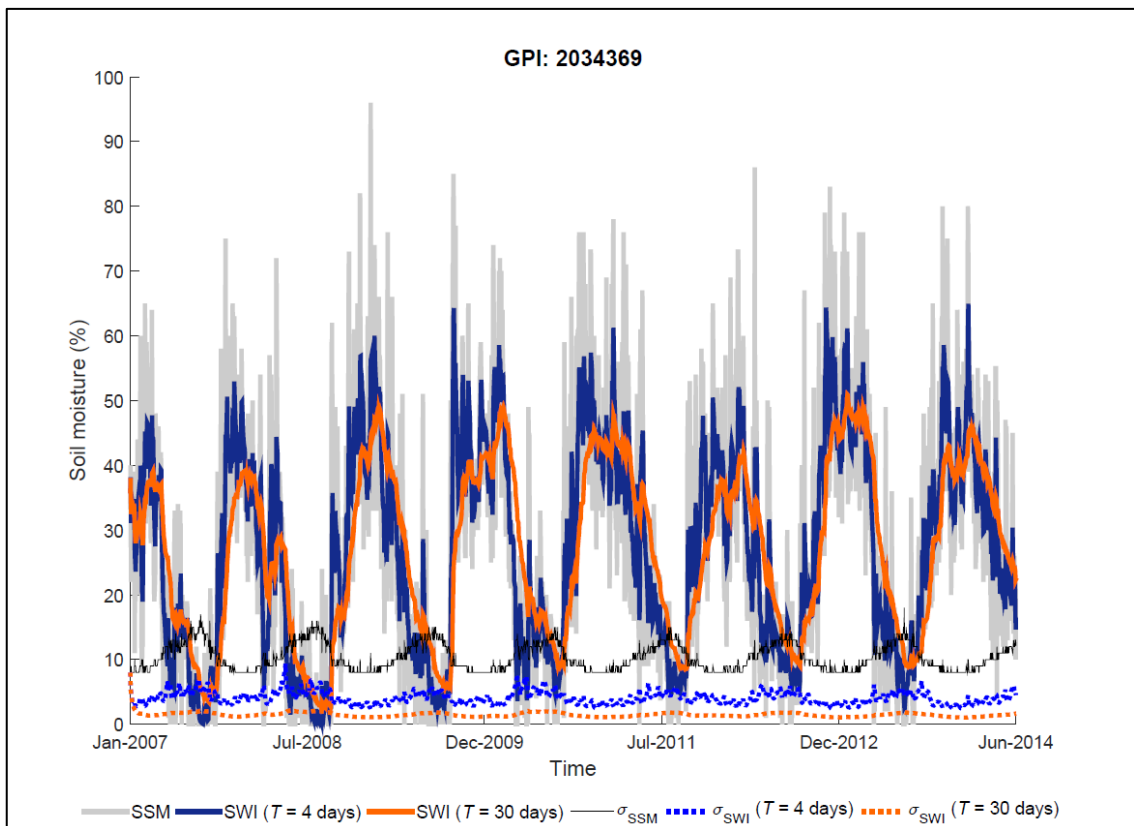


Figure 5.5. Time series of SSM and SWI (calculated for $T=4$ and $T=30$ days) and their corresponding noises σ , estimated for the ASCAT grid point nearest to site 4.

For the sake of brevity, only the results obtained for site 4 are shown in detail in Table 5.4, where, to facilitate the comparison, performance scores of the DM0 configuration are also reported. Five different thresholds, based on SSM or SWI noise values are assumed and referred to as ‘Threshold 1’ - ‘Threshold 5’, where increasing indexes correspond to lower critical noise values and consequently more discarded data. Figure 5.6 shows the results in terms of the ratio between the obtained r and the one

corresponding to the DM0 configuration, for different noise thresholds expressed in terms of the percentage of removed data, either SSM or SWI, for six of the selected sites. Observed ranges for r values obtained from the application of different thresholds for the three approaches, are summarized in Figure 5.7a; Figures 5.7b, 5.7c and 5.7d illustrates the same analysis in terms of RMSD, r_{anom} and T .

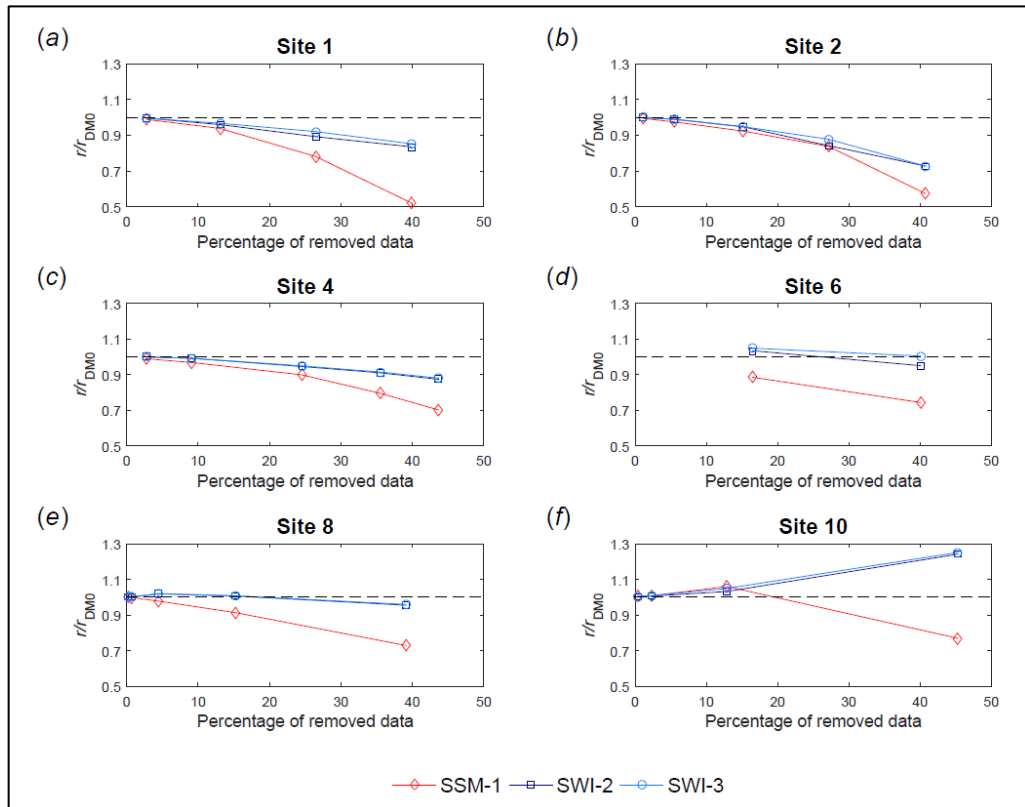


Figure 5.6. Comparison between effects of assuming different SSM and SWI noise thresholds on correlation coefficients.

Table 5.4. Results of application of the different thresholds on SSM and SWI noise for site 4, compared to DM0 configuration (no quality check). T expressed in days, RMSD in $m^3 m^{-3}$.

		DM0	Threshold 1			Threshold 2			Threshold 3			Threshold 4			Threshold 5		
			SSM-1	SWI-2	SWI-3	SSM-1	SWI-2	SWI-3	SSM-1	SWI-2	SWI-3	SSM-1	SWI-2	SWI-3	SSM-1	SWI-2	SWI-3
Site 4	$\sigma_{\text{threshold}}$	-	15	6.34	6.49	14	5.68	5.72	13	5.23	5.70	12	5.95	6.05	11	5.94	6.61
OBS ₃₀	T	3.6	3.6	3.6	3.4	3.5	3.5	3.4	3.1	3.1	2.6	2.0	2.0	1.9	1.8	1.8	1.4
	r	0.831	0.823	0.831	0.831	0.805	0.825	0.825	0.748	0.787	0.790	0.661	0.757	0.761	0.584	0.727	0.732
	RMSD	0.023	0.024	0.023	0.023	0.024	0.024	0.024	0.026	0.024	0.024	0.027	0.024	0.024	0.027	0.025	0.025
	r_{anom}	0.452	0.457	0.462	0.464	0.480	0.479	0.480	0.507	0.500	0.515	0.543	0.535	0.535	0.538	0.551	0.565
	N_{obs}	1473	1433	1433	1433	1339	1339	1339	1112	1112	1112	949	949	949	830	830	830
Site 4	$\sigma_{\text{threshold}}$	-	15	5.11	5.74	14	4.75	4.25	13	5.07	5.08	12	5.99	5.46	11	5.78	6.07
OBS ₆₀	T	5.5	5.2	5.2	4.3	4.9	4.9	6.1	3.3	3.3	3.3	2.0	2.0	2.4	1.9	1.9	1.7
	r	0.816	0.805	0.816	0.816	0.783	0.808	0.810	0.715	0.767	0.767	0.628	0.734	0.737	0.562	0.710	0.718
	RMSD	0.021	0.022	0.021	0.021	0.022	0.022	0.022	0.024	0.023	0.023	0.026	0.023	0.023	0.025	0.023	0.023
	r_{anom}	0.364	0.369	0.376	0.380	0.394	0.380	0.380	0.425	0.420	0.419	0.473	0.446	0.455	0.463	0.485	0.488
	N_{obs}	1475	1435	1435	1435	1341	1341	1341	1114	1114	1114	951	951	951	832	832	832
Site 4	$\sigma_{\text{threshold}}$	-	15	5.79	6.53	14	5.28	5.60	13	5.16	5.62	12	5.98	6.05	11	5.88	6.06
OBS _{mean}	T	4.3	4.2	4.2	3.4	4.0	4.0	3.6	3.2	3.2	2.7	2.0	2.0	1.9	1.8	1.8	1.7
	r	0.840	0.831	0.840	0.840	0.813	0.833	0.834	0.754	0.796	0.798	0.669	0.768	0.771	0.596	0.741	0.746
	RMSD	0.021	0.021	0.021	0.021	0.022	0.021	0.021	0.023	0.022	0.022	0.024	0.022	0.022	0.024	0.022	0.022
	r_{anom}	0.454	0.459	0.466	0.469	0.480	0.478	0.481	0.499	0.499	0.507	0.540	0.518	0.520	0.535	0.548	0.554
	N_{obs}	1473	1433	1433	1433	1339	1339	1339	1112	1112	1112	949	949	949	830	830	830

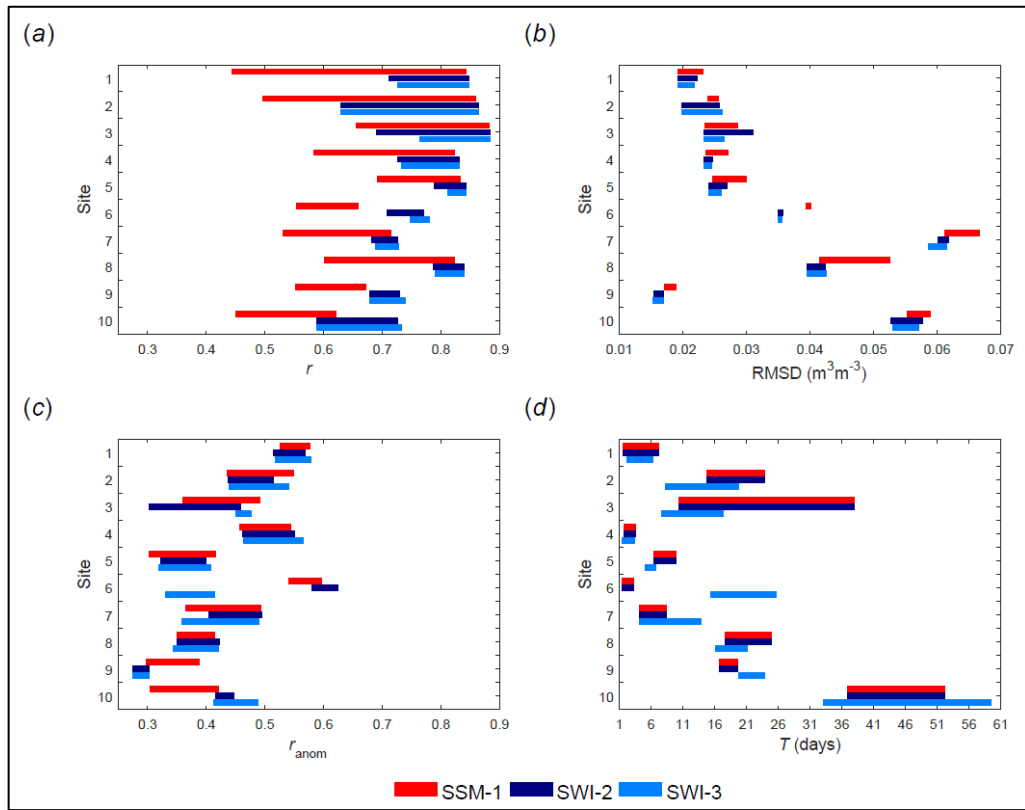


Figure 5.7 – Effects of the different SSM and SWI noise thresholds, for all sites at the most superficial observation depth, on ranges of: a) r ; b) RMSD; c) r_{anom} ; d) T parameter.

Discarding ASCAT data based on SSM noise (SSM-1) generally lead to the lower performances of SWI in reproducing in situ measurements, in terms of r and RMSD, as shown in Figures 5.7a and 5.7b. On the other hand, optimizing T without masking underlying SSM input (SWI-3), always outperforms the other tested approach.

In every case, p -value confirm that the correlations obtained are statistically significant for both absolute and anomalies time series at the 0.05 level of significance. Differences between correlation coefficients obtained with the three approaches are also checked performing two significance tests based on the well-known Fisher's Z-transformation and on Zou's (2007) confidence interval respectively. As expected, significant differences are obtained comparing the SSM-1 approach to SWI-2 or SWI-3, both implying the use of a threshold on the proposed SWI noise and differing only for T parameter estimation. Moreover, the differences between correlation coefficients become significant for decreasing threshold values that remove larger amount of data.

Focusing on the SWI-3 approach, the use of more restrictive thresholds, i.e. lower values, on SWI noise generally leads to performance worsening compared to DM0 configuration, with the exceptions of sites 9 and 10 characterized by lower reference values r_{DM0} . For example, referring to OBS_{30} at site 4 (Table 5.4), r decreases from 0.831 to 0.790, 0.761 and 0.732, while RMSD increases from 0.023 to 0.024 and 0.025 m^3m^{-3} , for Thresholds 3, 4 and 5 respectively. A possible explanation of this deterioration is that restrictive thresholds tend to remove SWI estimates characterized by high noise in underlying SSM measurements rather than by insufficient input availability; as evidenced by Figure 5.5, according to the observed SSM noise seasonal pattern, the removal of SWI values is mainly concentrated in the summer months, generally characterized by a better relative agreement with in situ measurements.

This evidence does not emerge in the analysis of the anomalies: the use of more restrictive thresholds on SSM noise and SWI noise, actually, can either results in better or worse performance in terms of r_{anom} . The dependence of r_{anom} on the seasonality of the root-zone soil moisture obtained for the DM0 configuration is less evident as already stated. Furthermore, comparing the r_{anom} results obtained for the investigated sites, the three approaches show similar performances; SWI-3 approach clearly outperforms the SSM-1 one only in site 10, as shown in Figure 5.7c.

Finally, Figure 5.7d highlights the key role played in this analysis by the exponential filter parameter. Sensible differences in T values estimated from SSM and SWI noise thresholds obviously affect the performance metrics too. As expected, lower T , associated to more irregular SWI series, correspond to marked improvement in the skill of the anomaly correlation. This evidence is carried to extreme in site 6: sensibly higher T values estimated with the SWI-3 methodology, although associated to valuable performance in terms of r for the absolute time series, similar to those obtained with the SWI-2 approach, correspond to r_{anom} significantly worse than the other two approaches.

5.6 Application of EP equations considering the error in SWI parameter

In the second study an updated ASCAT data records release was used, in which the SSM error estimates show much lower magnitude and considerably reduced seasonal patterns. The latest product version benefits from several updates, mostly in instrument

calibration, model parameters, and algorithms; consequently, no prominent seasonal patterns are reflected in estimated SWI standard error time series, as it happened instead in the previous study.

The comparison between SWI and in situ measurements generally confirms the results obtained using the previous ASCAT SSM product version (e.g. optimal T values reported in Table 5.5).

Table 5.5. Estimated T values at different depths for the considered study sites.

Code, depth (cm)	T (days)	Code, depth (cm)	T (days)	Code, depth (cm)	T (days)	Code, depth (cm)	T (days)	Code, depth (cm)	T (days)
#1, 30	6	#3, 30	10	#5, 30	7	#7, 20	4	#9, 20	23
—, 60	17	—, 60	25	—, 60	19	—, 40	5	—, 40	31
#2, 30	13	#4, 30	4	#6, 30	4	#8, 20	18	#10, 25	38
—, 60	34	—, 60	6			—, 40	19	—, 35	40

In parallel SWI error variance time series were estimated with the approach described in Sect. 5.3, assuming $\sigma(T)$ equal to the 10% of the estimated T value: although it is rather arbitrary, this is a commonly used value for error analysis when parameters do not have well-defined accuracy (Parinussa et al. 2011, Pathe et al. 2009). As shown in Figure 5.8, the differences in computed uncertainty due to the introduction of the term related to the parameter T can be remarkable.

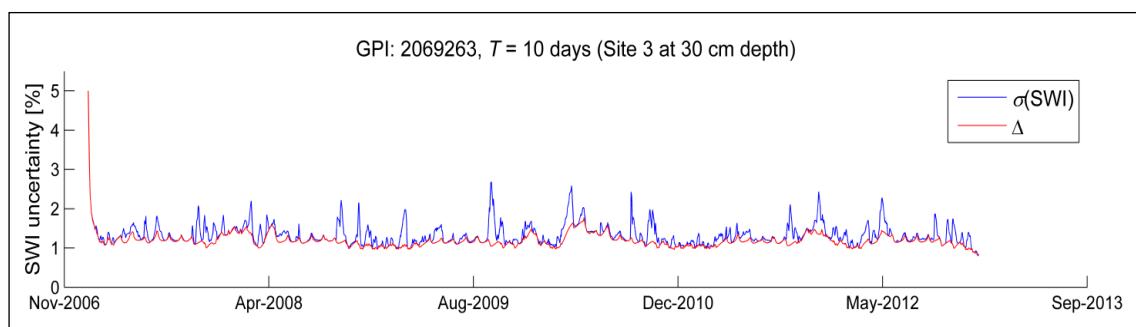


Figure 5.8. Example of estimated SWI uncertainties time series (in red the values calculated by considering only input uncertainties).

In this study, a preliminary assessment of estimated SWI uncertainties was carried out, by focusing only on the highest values in the time series, investigating the

correspondence with potentially less reliable SWI data, i.e. with data that do not fit well in situ observations. The correspondence between observed deviations and SWI random error realizations theoretically holds well as long as the latter is the main error term. Ground measurements, used as reference for the evaluation of coarse-scale SWI, also contain random errors (instrumental and representativeness); furthermore, residual systematic differences are to be expected, as mean and variance matching technique constitutes a suboptimal rescaling solution (Yilmaz & Crow, 2013).

The more uncertain SWI values are firstly masked by setting four $\sigma(\text{SWI})$ thresholds to remove fixed percentage of data (5, 10, 15 and 20%), and then a linear scaling was performed to remaining SWI data; finally, performance metrics r and RMSD are recomputed to check whether the discarded data contributed positively or not to the overall agreement between the in situ and satellite-based soil moisture time series.

The masked data are deemed to be mainly those characterized by high $(\partial\text{SWI}/\partial T)$ values and correspond to significant changes in the SWI value, which occurring during seasonal or sub-seasonal (short-time) fluctuations due to high or low T parameter respectively (Figure 5.9). When SWI well describes the observed soil moisture dynamics (representative sites, good r and RMSD values) just showing the known exponential filter limitations in modelling sub-seasonal fluctuations, it is expected that $\sigma(\text{SWI})$ thresholds improve performance scores as they identify the situations of actual lower agreement between SWI and ground data. About the role of exponential filter parameter, as stated above the smoothed SWI time series obtained for high T values show well localized uncertainty seasonal patterns (Fig. 5.9 for $T = 25$ days).

To examine the role played by parameter uncertainty, the described procedure was also performed by applying thresholds on Δ , whose value depends on SSM error variance patterns and temporal availability on a time-scale related to the T value. The obtained r and RMSD values are shown in Tables 5.6 and 5.7 respectively, while Figure 5.10 summarizes the results for 9 sites at the most superficial observation depth.

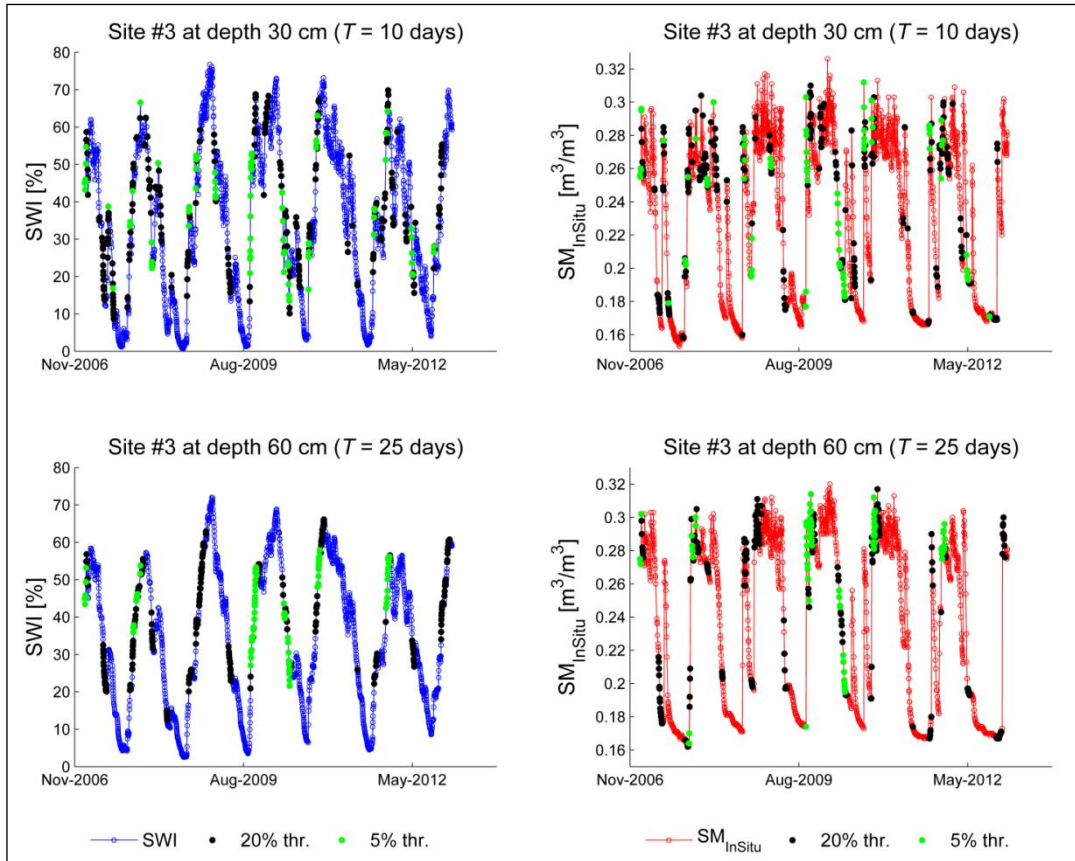


Figure 5.9. Example of SWI estimates over $\sigma(\text{SWI})$ thresholds (on the left) and corresponding in situ measurements (on the right), for different T values, before mean and variance matching.

Table 5.6. r values for the different error thresholds.

Code, depth (cm)	No thr.	Data removed by $\sigma(\text{SWI})$ threshold:				Data removed by Δ threshold:			
		5%	10%	15%	20%	5%	10%	15%	20%
#1, 30	0.858	0.863	0.866	0.869	0.873	0.862	0.866	0.865	0.861
—, 60	0.876	0.878	0.883	0.889	0.894	0.876	0.875	0.879	0.881
#2, 30	0.864	0.872	0.881	0.885	0.889	0.864	0.866	0.864	0.854
—, 60	0.899	0.899	0.900	0.905	0.910	0.898	0.897	0.895	0.895
#3, 30	0.885	0.896	0.908	0.914	0.918	0.886	0.888	0.887	0.883
—, 60	0.854	0.857	0.867	0.875	0.889	0.853	0.852	0.851	0.849
#4, 30	0.831	0.834	0.837	0.837	0.834	0.832	0.833	0.829	0.825
—, 60	0.816	0.825	0.832	0.840	0.845	0.818	0.818	0.812	0.808
#5, 30	0.840	0.842	0.845	0.850	0.849	0.843	0.844	0.843	0.841
—, 60	0.853	0.860	0.862	0.866	0.872	0.853	0.854	0.855	0.857
#6, 30	0.743	0.763	0.776	0.777	0.785	0.760	0.773	0.773	0.779
#7, 20	0.721	0.731	0.741	0.749	0.765	0.732	0.740	0.747	0.749
—, 40	0.721	0.735	0.748	0.754	0.765	0.738	0.747	0.755	0.760
#8, 20	0.824	0.827	0.831	0.833	0.832	0.842	0.849	0.853	0.853
—, 40	0.817	0.819	0.823	0.826	0.822	0.826	0.828	0.830	0.832
#9, 20	0.624	0.637	0.647	0.653	0.664	0.683	0.687	0.697	0.694
—, 40	0.802	0.803	0.802	0.809	0.809	0.795	0.800	0.805	0.807
#10, 25	0.589	0.577	0.567	0.555	0.549	0.591	0.592	0.586	0.583
—, 35	0.625	0.611	0.599	0.592	0.588	0.623	0.623	0.617	0.611

Table 5.7 RMSD (m^3/m^3) values for the different error thresholds.

Code, depth (cm)	No thr.	Data removed by $\sigma(\text{SWI})$ threshold:				Data removed by Δ threshold:			
		5%	10%	15%	20%	5%	10%	15%	20%
#1, 30	0.0185	0.0184	0.0184	0.0184	0.0182	0.0185	0.0185	0.0186	0.0187
—, 60	0.0275	0.0275	0.0272	0.0268	0.0263	0.0276	0.0279	0.0275	0.0274
#2, 30	0.0248	0.0241	0.0235	0.0234	0.0233	0.0253	0.0256	0.0257	0.0261
—, 60	0.0267	0.0268	0.0267	0.0263	0.0258	0.0272	0.0277	0.0283	0.0289
#3, 30	0.0231	0.0221	0.0209	0.0203	0.0199	0.0233	0.0233	0.0234	0.0237
—, 60	0.0279	0.0275	0.0265	0.0257	0.0243	0.0283	0.0286	0.0290	0.0294
#4, 30	0.0233	0.0231	0.0227	0.0227	0.0229	0.0233	0.0232	0.0234	0.0235
—, 60	0.0211	0.0203	0.0200	0.0194	0.0191	0.0212	0.0214	0.0217	0.0218
#5, 30	0.0241	0.0237	0.0235	0.0234	0.0232	0.0240	0.0241	0.0243	0.0246
—, 60	0.0230	0.0225	0.0225	0.0224	0.0221	0.0234	0.0235	0.0237	0.0235
#6, 30	0.0386	0.0369	0.0359	0.0359	0.0352	0.0373	0.0365	0.0363	0.0361
#7, 20	0.0607	0.0600	0.0596	0.0592	0.0575	0.0603	0.0599	0.0592	0.0591
—, 40	0.0575	0.0565	0.0555	0.0554	0.0543	0.0565	0.0559	0.0554	0.0551
#8, 20	0.0414	0.0409	0.0409	0.0412	0.0418	0.0394	0.0385	0.0381	0.0384
—, 40	0.0297	0.0297	0.0296	0.0298	0.0302	0.0292	0.0289	0.0288	0.0287
#9, 20	0.0184	0.0177	0.0175	0.0175	0.0174	0.0170	0.0170	0.0166	0.0167
—, 40	0.0727	0.0731	0.0736	0.0718	0.0716	0.0744	0.0741	0.0737	0.0735
#10, 25	0.0569	0.0580	0.0591	0.0605	0.0617	0.0571	0.0577	0.0589	0.0597
—, 35	0.0592	0.0604	0.0616	0.0625	0.0635	0.0596	0.0602	0.0614	0.0622

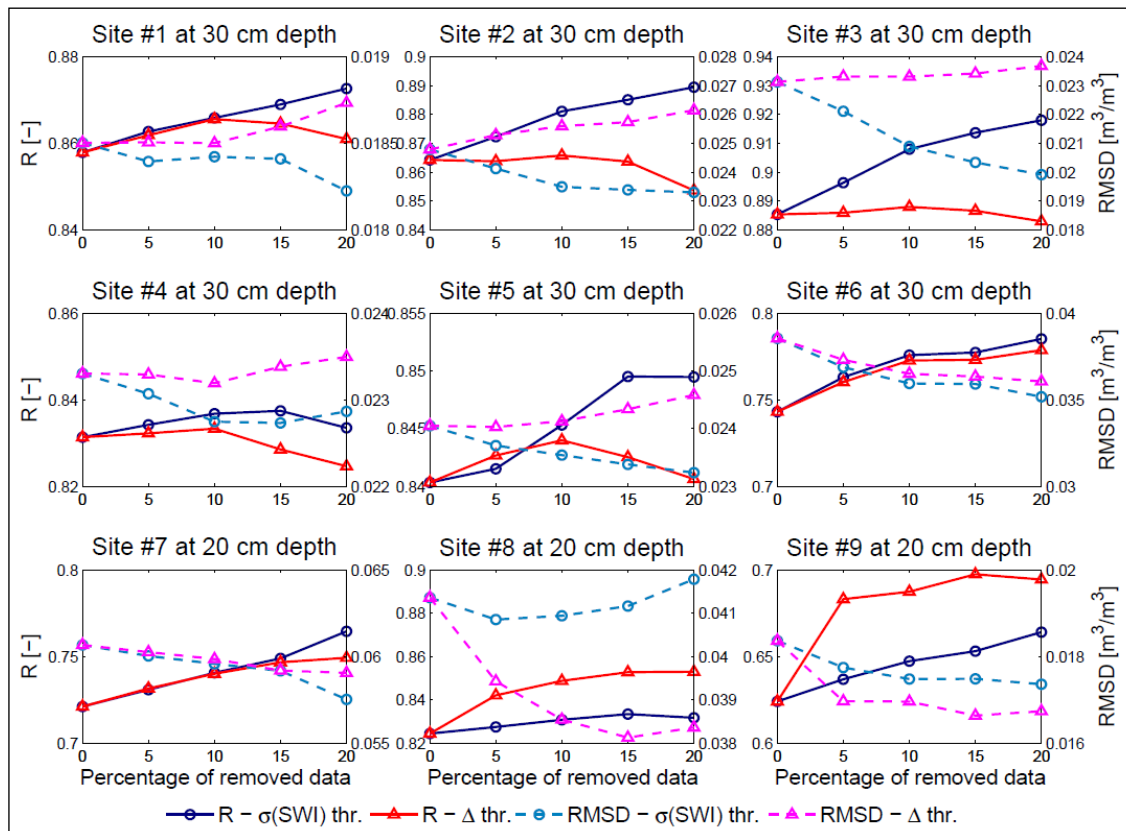


Figure 5.10. Effects of the different error thresholds on correlation coefficient R and RMSD values.

The obtained correspondence between the removing more uncertain SWI values and the improving the observed performance metrics is generally in agreement with the expected. Increasing thresholds on $\sigma(\text{SWI})$ seem to lead to a systematic better correlation between remaining SWI data and reference ground measurements. The only exception is the site 10 (HYDROL-NET network), characterized by the lowest r starting values, probably indicating that also in situ errors contribute significantly on observed deviations between time series; in this site the highest optimal T values are also obtained. Marked performance improvements are observed in sites 1-5 (Calabria) and also in 6 (Campania), both in terms of r and RMSD. In Umbria network, where the starting performance scores are generally lower, the procedure gave good results on site 7, characterized by the lowest T values, with $\sigma(\text{SWI})$ thresholds detecting short-time soil moisture variations substantially distributed in all seasons of the year. In the remaining sites (8 and 9), where T values are higher and masked observations show the expected seasonal patterns, the performance improvements are less evident (it is worth to note that RMSD is also influenced by the trend in the variance of remaining reference ground data).

Regarding Δ , data masking generally shows reduced performance improvements, often limited to the removal of a moderate percentage of SWI values (that include those corresponding to exponential filter initializations); however, in sites 8-10, where $\sigma(\text{SWI})$ thresholds gave less positive results, the control on Δ performs better, due to different temporal patterns in masked data, confirming that in this sites SWI seasonal transitions are not the most critical periods in reproducing ground data.

The performance improvements obtained in applying $\sigma(\text{SWI})$ thresholds can be therefore explained by the removal of data that testify the limited skills of exponential filter in detecting short time-scale fluctuations, including those that occur in correspondence of seasonal state variation. According to results obtained in some representative sites, the seasonal state variation proved to be situation where exponential filter shows modelling limits. The proposed EP scheme thus succeeds in identifying SWI estimates that potentially do not capture properly soil moisture state transition and deviate considerably from reference ground measurements.

5.7 Conclusions

An error propagation approach was proposed to characterize the time-variant SWI error variance.

In a first study, a simplified scheme was used, taking into account the errors as well as the availability of the underlying SSM data in SWI estimate, but neglecting the uncertainty in T parameter. In this study, ASCAT-derived SWI showed a good agreement with reference root zone ground data, even when no quality control was performed, with the exponential filter generally showing a greater ability in capturing the seasonal soil moisture behaviour rather than short time-scale fluctuations. Different SSM data masking procedures, based on several indicators available for the ASCAT product, were analysed; however, they did not necessarily led to a more accurate description of the root-zone water content in terms of agreement with in situ observations. Regarding the EP scheme, the use of thresholds on SWI noise led to performance worsening compared to the configuration without quality controls, due to SSM noise seasonal pattern that implies the removal of SWI values mainly concentrated in a season characterized by a better relative agreement with in situ measurements. However, removing data according to ‘a posteriori’ control on SWI noise generally improved the performance metrics with respect to ‘a priori’ SSM noise threshold application.

In the second study, error in T was taken in account, and a successive release of ASCAT SSM dataset was employed, with no prominent seasonal patterns in SSM noise. In this case, the proposed EP scheme has shown capabilities to identify potentially less reliable SWI values in the selected study sites, improving performance metrics respect to the configuration without quality controls. The preliminary results suggest the utility of EP approach in the SWI evaluation, in the comprehension of the exponential filter shortcomings, and in the time-variant SWI error characterization. Due to the exponential filter mathematical structure, error estimates could be not reliable in magnitude; however, they can help to discriminate more and less certain data. The information provided by analytical error propagation equations can be suitable not only for a preliminary screening analysis, but also to reproduce the trend, rather than the actual values, of satellite-based SM error variance. However, with regard to data assimilation applications, the role of observation error temporal structure on analysis

performances is still under research: there are few studies in which the temporal variability in error variance is not neglected, and the experiments generally did not exhibit significant variations in assimilation results with respect to the stationary case (e.g. Alvarez-Garreton et al., 2013).

6. EFFECTS OF DIFFERENT RESCALING AND ERROR CHARACTERIZATION SCHEMES IN AN EXTENSIVE DATA ASSIMILATION EXPERIMENT

6.1 Introduction

This chapter describes a second application related to the assessment of satellite soil moisture preprocessing and error characterization practices, carried out during a research period at CNR-IRPI with the Hydrology group, that provided datasets and support.

In this study, the effects of remotely sensed soil moisture assimilation on hydrological model performances have been extensively explored, over a large number of catchments located across Europe. The benefit from integrating satellite SSM in rainfall-runoff modelling is potentially very high, however several studies have shown ambiguous results in terms of actual improvements in runoff predictions (see Sect. 3.3). This can be attributed to two main factors: 1) the actual quality of the several satellite products under different climatic and physiographic conditions; 2) some methodological issues like those related to the observations preprocessing and the error characterization phases (see Sect. 4.2). In this application these aspects are investigated in the following way. Different ESA CCI SM products are used, that merge several available active and passive measurements, in order to evaluate the role of the sensors on DA performances. In the preprocessing phase, satellite datasets are rescaled to the reference according to two alternative approaches, i.e. CDF-matching and Triple Collocation analysis (TC), that imply the matching of total variance and only signal component, respectively. Then, TC is also used for observation error characterization, considering distinct triplet configurations also in order to test the impact of different observation weights in data assimilation system performances. These different procedure setups do not involve any DA performance optimizations on the individual catchments.

A previous DA experiment on the same dataset was made by CNR-IRPI Hydrology group, using mean and variance matching for rescaling and then a nudging scheme with gain calibrated by optimizing the DA system performance; the experiment showed only limited model performance improvements after assimilation on a moderate percentage of catchments, with ‘optimal’ gains less than 0.1 in most cases (that means high errors

attributed to satellite with respect to the model, small weight to observation in state updating, and finally limited corrections and analysis error variance improvements).

Here we use the Ensemble Kalman Filter (EnKF), which properly takes in account background error evolution. The EnKF has been widely used for SM DA given its computational efficiency and the strongly nonlinear dynamics of hydrological processes. Kalman Filter and its variants provide an optimal solution in minimizing the error variance of analysis state variable, relying on the assumption of background and observation errors strictly random and correctly characterized in statistical terms, as well as mutually uncorrelated and white (uncorrelated in time).

The role of the rescaling approach has been analysed also considering its effect in terms of multiplicative bias between rescaled satellite-based and reference model SM estimates (i.e. differences between datasets that show a systematic relationship with the estimate values), as the existence of multiplicative bias components between satellite-based and model data may provide non-optimal assimilation results. In this respect, CDF-matching is not targeted to correct multiplicative bias (Kornelsen & Coulibaly, 2015), although it exhibits good performances in fitting the reference dataset. Triple Collocation analysis was proposed as optimal rescaling solution for data assimilation framework (Yilmaz & Crow, 2013), as it takes into account the existence of multiplicative error terms in dataset error structures; the resulting rescaled dataset should show less evidence of multiplicative bias, although it reproduces less well the model reference.

In the following paragraphs are firstly described the study catchments and the provided datasets, then the MISDc-2L model (Brocca et al., 2012a) used for the simulation of hydrological response, the methodologies adopted for observation preprocessing and error characterization, the assumptions made for the representation of model errors, the EnKF implementation, and finally the data assimilation results.

6.2 Study catchments and datasets

6.2.1 Study catchments and hydrological data

This study makes use of a hydrological dataset originally including 881 catchments located across Europe; however, only a subset of 729 is actually employed for the data assimilation experiments (Fig. 6.1), selected considering a threshold on hydrological

model calibration performance. Data refer to the study period 2002-2011, chosen by intersecting the temporal availability of runoff and satellite SM observations. Daily discharge data and basin characteristics have been obtained from the Global Runoff Data Centre (GRDC, <https://www.bafg.de/GRDC>). Catchment areas range from 150 to 150000 km². The list of study catchments is in Annex 1.

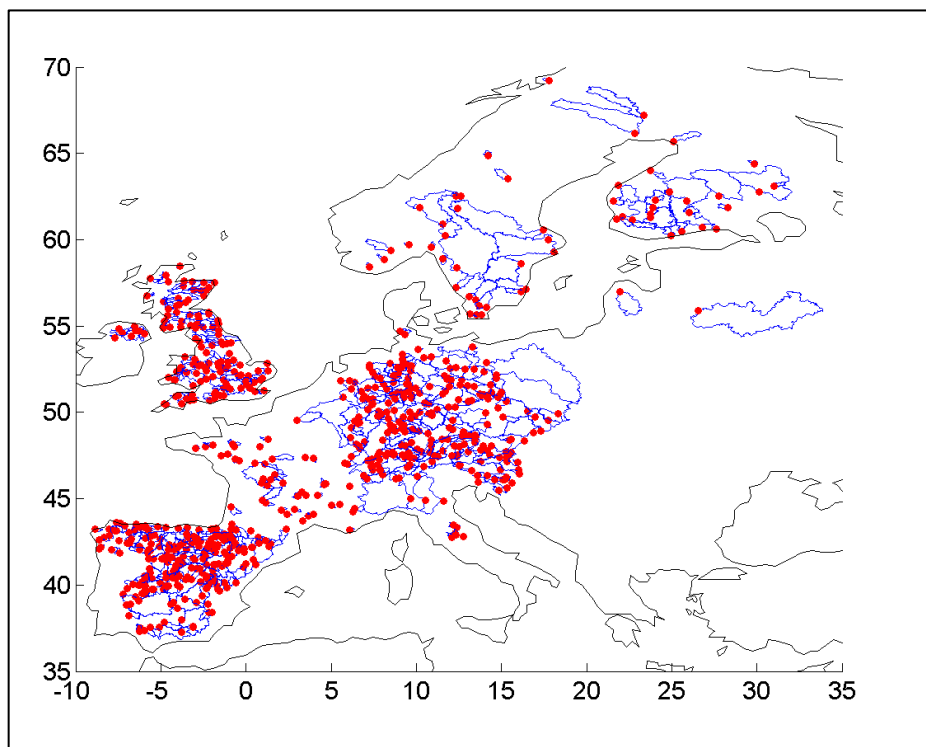


Figure 6.1. Study catchments (red points indicate the river gauge).

6.2.2 Climatic data

Ground-based daily rainfall and mean temperature data have been collected from the European Climate Assessment & Dataset E-OBS (<https://www.ecad.eu/>) (Haylock et al., 2008). E-OBS gridded products are delivered for the European domain on four spatial resolutions, including 0.25° by 0.25° on a regular latitude-longitude grid. The gridded data are obtained from daily station observations by using different appropriate methods for the spatial interpolation of the several climate variables (Hofstra et al., 2008), in order to provide the best estimate of grid box averages. Additionally, uncertainty estimates are available, provided as standard errors, that only represent interpolation errors and are shown to be largely dependent on the number of contributing observations (Haylock et al., 2008). According to Hofstra et al. (2009), the

provided standard errors significantly underestimate the true interpolation errors when cross-validated with ground data and therefore will similarly underestimate the interpolation errors in the gridded E-OBS data. E-OBS gridded daily data for the study period have been averaged over the catchment areas; rainfall standard error estimates are instead used only to visually assess the presence of spatial uncertainty patterns within the study catchments (Fig. 6.2).

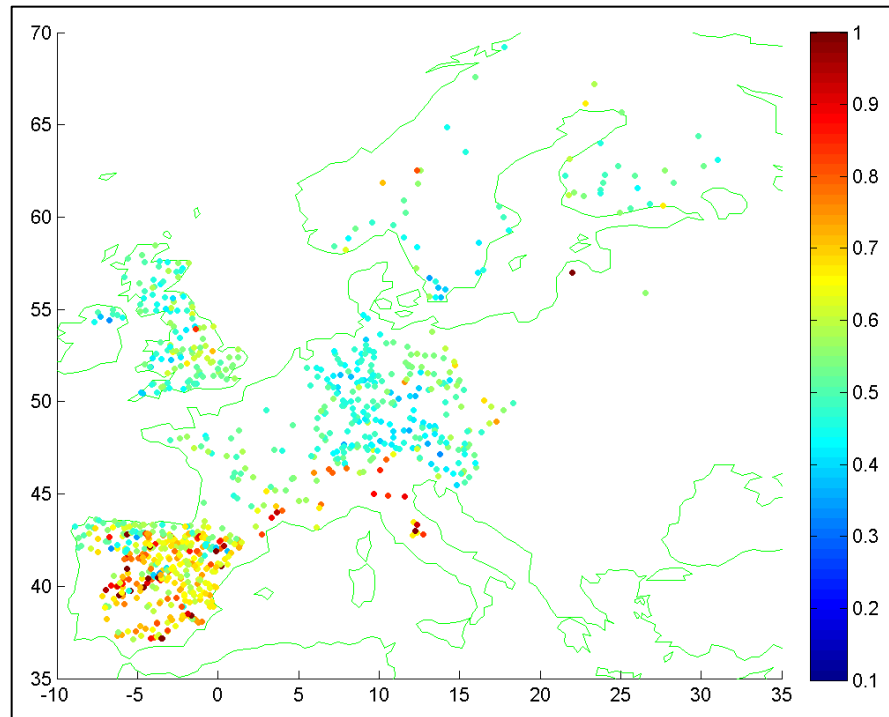


Figure 6.2. E-OBS rainfall standard error averaged over selected study catchments. Provided errors show a clear spatial pattern, with higher values in Mediterranean area than at northern latitudes.

6.2.3 Remotely-sensed soil moisture data

The satellite surface soil moisture observations have been derived from the European Space Agency (ESA) Climate Change Initiative (CCI), which provides active, passive and combined merging products with a spatial resolution of 0.25° and a daily temporal sampling (Liu et al., 2012a; Dorigo et al., 2017). The version v03.2 has been here employed. The quality indicators provided within the SSM dataset have been used to mask unreliable estimates (e.g. from snow cover or frozen soil), and then the mean daily SSM values have been computed for each study catchment. Regarding the merged sensors data, active product is based on backscatter measurements from the ERS scatterometers (characterized by a revisit time of several days) until 2006 and then from

ASCAT on board MetOp-A satellite (with an almost daily resolution). In the selected study period, passive product consists in measurements of AMSR-E sensor on Aqua satellite, WindSat on Coriolis, and from the dedicated SMOS mission launched in November 2009. For a more objective comparison between the different satellite products performance, some DA results will be presented referring to two separate eras, i.e. 2003-2006 and 2007-2011 periods, due to the different temporal availability of observations in ESA-CCI datasets.

6.3 Hydrological model and calibration results

MISDc (Modello Idrologico Semi-Distribuito in continuo) is a continuous semi-distributed rainfall-runoff model developed by Brocca et al. (2011b). A two layers version, MISDc-2L, was proposed in Brocca et al. (2012a), which incorporates a thin surface soil layer and is specifically addressed to the assimilation of remotely sensed SSM (Fig. 6.3). Here the MISDc-2L variant used in Massari et al. (2018) is employed, that also includes a snow module.

The model is applied in a lumped mode, with a daily time-step. It uses rainfall and temperature data as inputs and simulates the temporal evolution of the soil water states in layer 1 (i.e. the surface one) and 2, W_1 and W_2 , expressed in water depth and limited to $W_{\max,1}$ and $W_{\max,2}$, respectively. Processes that determine water balance in layer 1 are rainfall infiltration, percolation to layer 2, and actual evapotranspiration. The saturation degree in the first layer, $W_{1\%} = W_1 / W_{\max,1}$, determines the partitioning of rainfall in infiltration and surface runoff contribute (infiltration excess) and the percolation rate, both through power law functions, as well as the actual evapotranspiration, through a linear relationship with the potential one, that is calculated in function of the temperature using a modified version of the Blaney & Criddle method (Blaney & Criddle, 1950). Similarly, percolation from layer 2 is related to saturation degree in the second layer and constitutes the subsurface runoff component. Saturation excesses in layers 1 and 2 occur when water content W exceeds the water capacity W_{\max} . Surface runoff generated from infiltration and saturation excess is convoluted through a Geomorphological Instantaneous Unit Hydrograph (GIUH) while the subsurface runoff is transferred to the outlet section by a linear reservoir approach. For both routing schemes, the lag time is evaluated by the relationship proposed by Melone et al. (2002).

$W_{\max,1}$ is fixed at 150 mm (which roughly corresponds to a minimum of 30 cm soil depth by assuming a porosity value of 0.5), while the following parameters are calibrated for each study catchment:

- $W_{\max,2}$ (total water capacity in 2nd layer);
- m_1 (exponent of percolation in 1st layer);
- m_2 (exponent of percolation in 2nd layer);
- $K_{s,1}$ (hydraulic conductivity in 1st layer);
- $K_{s,2}$ (hydraulic conductivity in 2nd layer);
- η (coefficient lag time - area relationship);
- K_c (potential evapotranspiration parameter);
- δ (exponent of infiltration relationship);
- C_m (snow module parameter).

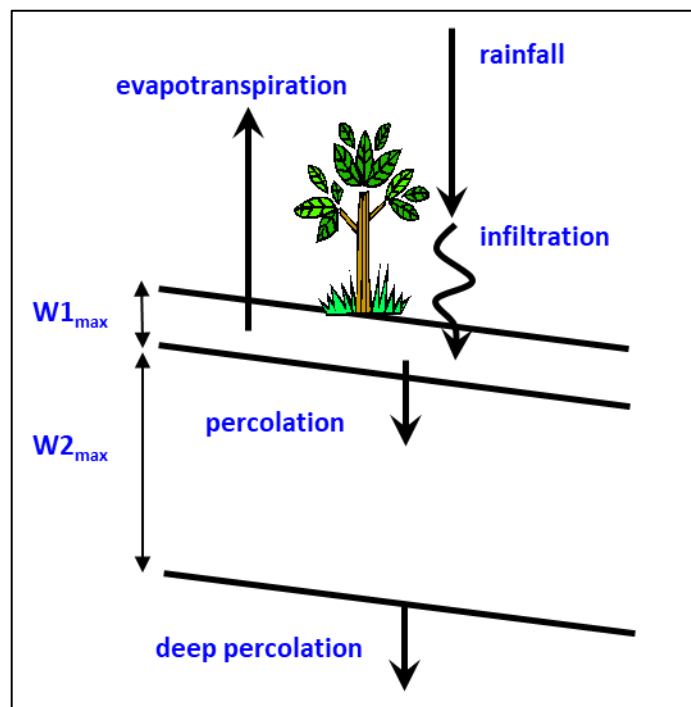


Figure 6.3. Conceptual scheme of MISDc-2L model (figure adapted from Brocca et al., 2012a).

An initial observation period (i.e. the year 2002) was used for model “warmup”, while the years 2003-2011 were considered for parameters calibration. The calibration was performed by optimizing the Kling-Gupta Efficiency index, KGE (Gupta et al., 2009). KGE is formulated by computing the Euclidian distance (ED) from the ideal point of three components, i.e. correlation and differences in mean and variability:

$$KGE = 1 - ED \quad (6.1)$$

with:

$$ED = \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (6.2)$$

$$\alpha = \sigma_s / \sigma_o \quad (6.3)$$

$$\beta = \mu_s / \mu_o \quad (6.4)$$

where r is the correlation coefficient, σ_s and σ_o the standard deviation and μ_s and μ_o the mean of the simulated and observed runoff, respectively. KGE ranges between $-\infty$ and 1 (optimal value).

The model calibration performances are reported in Fig. 6.4. All selected catchments have KGE over 0.6, with a median value of 0.84. In this respect, model performance can be considered relatively good, and this ensures the robustness of the model for streamflow simulations. It is worthy to note that patterns in KGE index show some similarities with those of interpolation errors in precipitation dataset shown in Fig. 6.2.

Time series of the degree of saturation in the first layer, $W_{1\%}$, obtained during the deterministic run with optimal parameters, are then used as reference data for the remotely sensed SSM preprocessing.

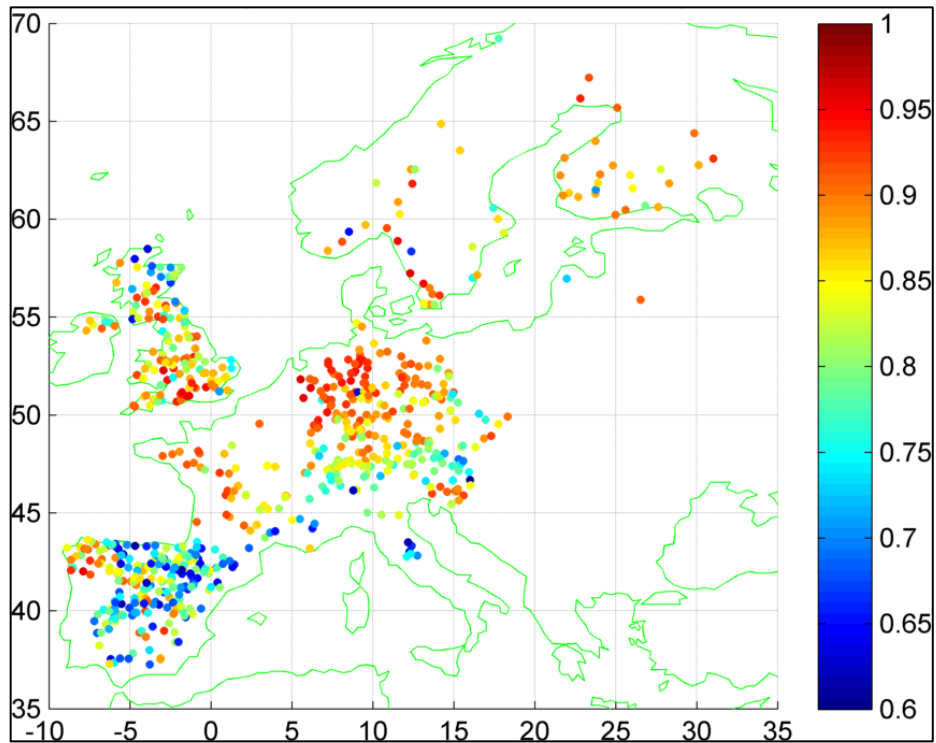


Figure 6.4. KGE values obtained during MISDc-2L calibration.

6.4 Observation preprocessing and error characterization

The ESA-CCI SM estimates, related to few centimetres below the soil surface, are mapped into the $W_{1\%}$ model variable space through two steps. Firstly, the exponential filter is used to address the depth mismatch, converting the SSM observations into SWI values; the parameter T is estimated by optimizing the correlation coefficient between SWI and the reference $W_{1\%}$ time series obtained from calibration. Then, two approaches are considered to address the systematic differences between SWI and $W_{1\%}$ data, i.e. the CDF-matching and the Triple Collocation (TC). The random error variance σ_{ϵ}^2 of the rescaled SWI, indicated as SWI*, is in both cases computed by TC. A graphical representation of the two methodologies followed for preprocessing and error characterization is reported in Fig. 6.5.

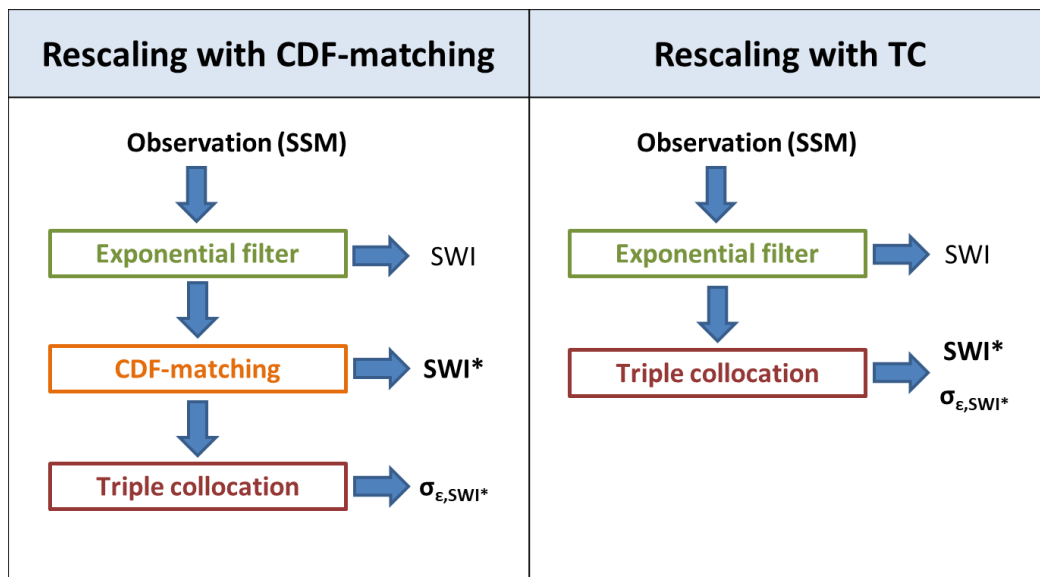


Figure 6.5. Schemes describing the two methodologies followed for preprocessing and error characterization.

In the first case, CDF-matching is used to obtain the SWI* that will be assimilated during the update of model states. The CDF-matching is a non-linear technique, that involves the matching of the total variances between reference and rescaled data. Following the method described in Sect. 4.5.2, a 4th order polynomial function is here used to correct the distribution differences. The SWI* random error variance will be then computed by TC with the Equation (4.30), including in the analysis three independent datasets, according to the configurations later described.

In the second case, Triple Collocation is used to simultaneously rescale SWI and compute SWI* random error variance. In TC analysis a linear rescaling is made, in

order to match only the variance part due to the signal. To obtain SWI*, the model dataset is chosen as reference dataset, against which the other two considered datasets are rescaled, according to Equations (4.25) and (4.27). The SWI* error variance is instead estimated (in the data space of the model reference) with the Equation (4.28).

With regard to the choice of the three independent datasets to be used in TC, two configuration types are considered. The first configuration includes model estimates and active and passive satellite-based observations, excluding the combined ESA-CCI product which is retrieved by merging the two others satellite datasets. Random standard error estimates obtained for this triplet configuration are showed in Figures 6.6 (CDF-matching rescaling) and 6.7 (TC rescaling).

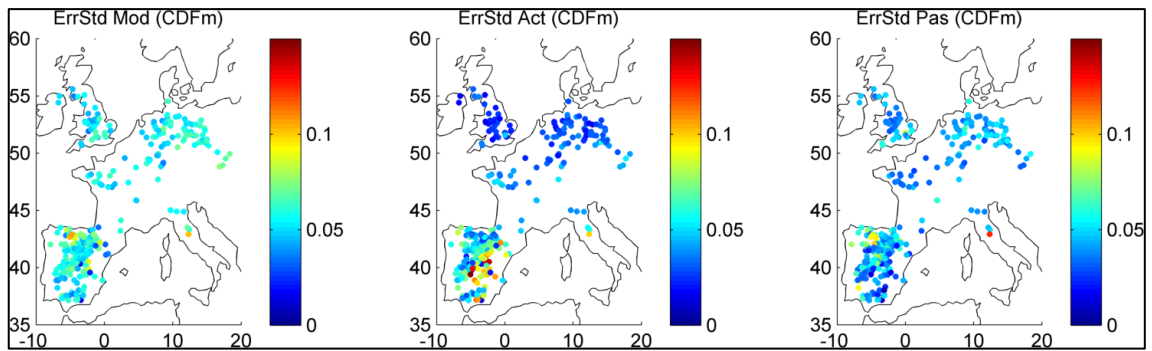


Figure 6.6. Random standard error estimates by TC, with the triplet consisting of model and active and passive based SWI* (from left to right), previously rescaled by CDF-matching.

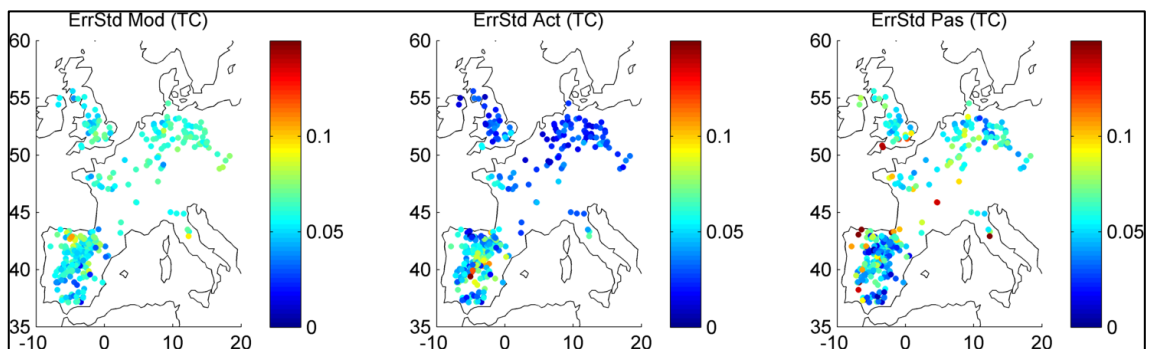


Figure 6.7. Random standard error estimates by TC, with the triplet consisting of model and active and passive based SWI* (from left to right), simultaneously rescaled by TC.

In alternative to this first TC configuration, the lagged variable (LV) approach proposed by Su et al. (2014a) is also used. In this case, a triplet is considered consisting of model, satellite observation (active or passive or combined), and 2 day-lagged model dataset as third variable. Results of LV approach for active and passive datasets are showed in Fig. 6.8, in order to make a comparison with those obtained with the first triplet

configuration. When lagged model is used as third variable, higher error variance are estimated for satellite-based datasets, that means lower gains in data assimilation analysis steps.

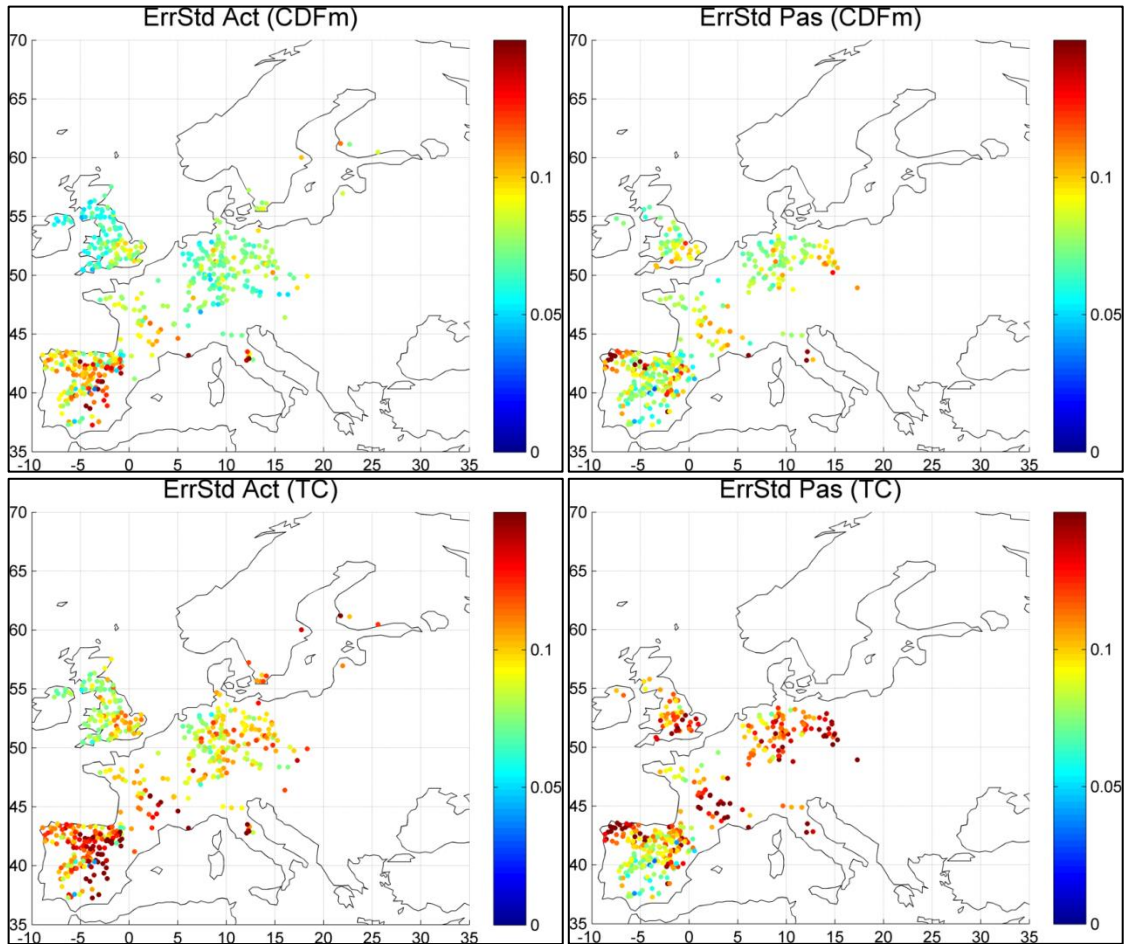


Figure 6.8. Random standard error estimates obtained by TC using the LV approach. The considered satellite data in the triplet is, from top left clockwise: active based SWI* rescaled by CDF-matching, passive based SWI* rescaled by CDF-matching, passive based SWI* rescaled by TC, and active based SWI* rescaled by TC.

In this study TC is always applied on climatological anomalies, after removing long-term 31 day moving average (e.g. Chen et al., 2014), as illustrated in Sect. 4.6.1. A minimum triplet size of 300 elements is imposed in order to obtain reliable error variance estimates, and a threshold is also set on correlation coefficient ($r > 0.5$) between SWI* and the first layer modelled soil moisture. For this reason, depending on rescaling approach and TC triplet configuration, the number of catchments actually used can be slightly different.

The obtained error variances will then be used in EnKF to compute Kalman gain and to perturb the observation (Burgers et al., 1998).

The different rescaling approaches (i.e. CDF-matching and Triple Collocation) are also evaluated in terms of ability to correct the multiplicative bias between SWI* and model estimates (see Sect. 4.5). Following Kornelsen & Coulibaly (2015), the slope of a linear fit between modelled soil moisture and its differences with SWI* is computed and used as indicator of multiplicative bias magnitude. As expected, TC analysis results to be more effective in reducing the multiplicative bias between satellite and reference datasets as quantified by this slope (Fig. 6.9); this is mainly because TC makes assumptions on (linear) error structure of datasets and aims to the matching of the signal part (that includes multiplicative bias terms). Positive slope values are always observed, indicating a tendency of SWI* to under- and overestimate the larger and smaller modelled SM value occurrences, respectively.

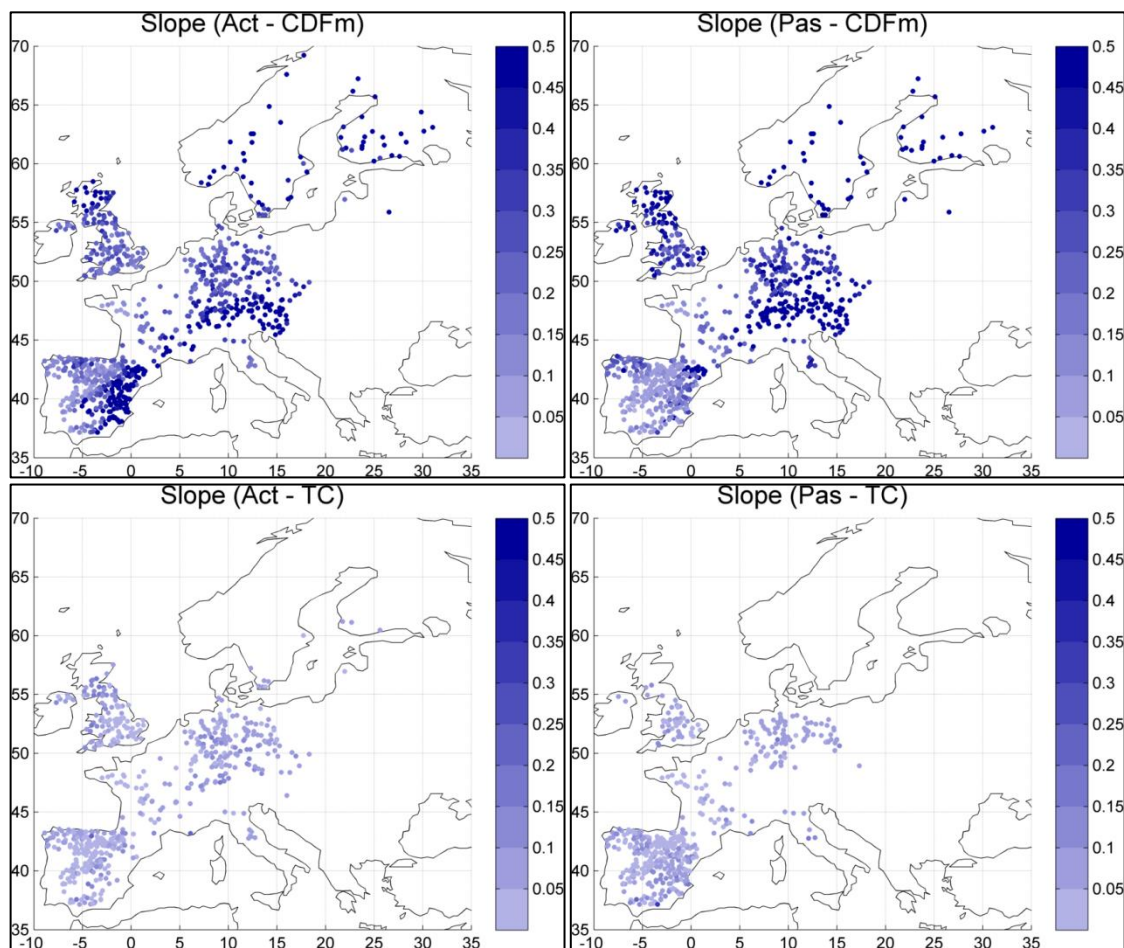


Figure 6.9. Slope of the linear fit between reference modelled soil moisture and its differences with SWI*. The considered satellite data is, from top left clockwise: active based SWI* rescaled by CDF-matching, passive based SWI* rescaled by CDF-matching, passive based SWI* rescaled by TC (LV approach), and active based SWI* rescaled by TC (LV approach).

6.5 Model error representation

As the Ensemble Kalman Filter technique was selected for the assimilation of satellite-based SWI*, a proper ensemble of model SM states is needed to represent the model background error covariance matrix. Errors in model representation of state variables are a consequence of both errors in forcing data and errors in parameters and structure of the model itself. A common approach is to represent both parameter and model structural errors by directly perturbing model states (e.g. Alvarez-Garreton et al., 2014; Massari et al., 2018); thus, in this study model errors are reproduced through perturbation of climate forcing (i.e. rainfall and temperature inputs) and soil moisture states (W_1 and W_2 model variables) according to appropriate error distributions. All generated errors are assumed temporally independent.

A multiplicative error is considered for precipitation (Tian et al., 2013), coming from a lognormal distribution with mean 1 and standard deviation σ_p . Inaccuracies in precipitation data are typically considered as the main error source in soil moisture modelling (e.g. Han et al., 2014; Massari et al., 2018). Given the high impact of rainfall on SM and runoff variability, the dimensionless σ_p value is calibrated for every catchment in order to satisfy the following ensemble test on discharges. If the ensemble spread is large enough, the temporal average of the ensemble skill, $\langle sk \rangle$, should be similar to the temporal average of the ensemble spread, $\langle sp \rangle$, i.e. $\langle sp \rangle / \langle sk \rangle \cong 1$ (De Lannoy et al., 2006), with:

$$\langle sp \rangle = \frac{1}{T_p} \sum_{k=1}^{T_p} \left\{ \frac{1}{N} \sum_{i=1}^N (Q_{i,k} - \langle Q \rangle_k)^2 \right\} \quad (6.5)$$

$$\langle sk \rangle = \frac{1}{T_p} \sum_{k=1}^{T_p} \left\{ (\langle Q \rangle_k - Q_{obs,k})^2 \right\} \quad (6.6)$$

where $Q_{i,k}$ is the simulated discharge of the i -th ensemble member at time t_k , $\langle Q \rangle$ the mean over the ensemble, Q_{obs} the observed discharge, N the ensemble size and T_p the number of time intervals. The ensemble size in this work was set to 100 members. The estimated σ_p value show spatial pattern similarities with Fig. 6.2, confirming differences in rainfall data accuracy within the study area.

For the other considered errors, an additive model is used, according to a zero-mean Gaussian distribution. With regard to temperature uncertainties, error standard deviation σ_T is fixed equal to 1° C. The state vector consisting of both W_1 and W_2 model variables, here expressed in saturation degree:

$$\mathbf{x} = [W_{1\%} \quad W_{2\%}] \quad (6.7)$$

is instead perturbed similarly to Han et al. (2014), according to the error covariance matrix Σ :

$$\Sigma = \begin{bmatrix} \Sigma & \gamma\Sigma r \\ \gamma\Sigma r & \gamma^2\Sigma \end{bmatrix} \quad (6.8)$$

where error variance Σ here corresponds to an error standard deviation of 1% in saturation degrees, r indicates the vertical correlation between perturbations applied to each soil layer and is assumed to be equal to 1 (as in Han et al., 2014) and γ reflects the ratio of 2nd to 1st layer error standard deviation. Here, γ is assumed equal to the ratio of covariance between $W_{1\%}$ and $W_{2\%}$ time series, to variance in $W_{1\%}$, both observed during the deterministic run. Assuming the presence of correlation between perturbations in the two soil layers implies an artificial increase in covariance between perturbed $W_{1\%}$ and $W_{2\%}$, and consequently an increase of the gain term during the update of 2nd layer SM variable.

The correction of unintended perturbation bias proposed by Ryu et al. (2009) is also integrated in model ensemble. The method consisted of running a single unperturbed model prediction in parallel with the perturbed model ensemble: at each time step, the bias between ensemble mean and unperturbed forecast is computed and then subtracted from each ensemble member (Fig. 6.10).

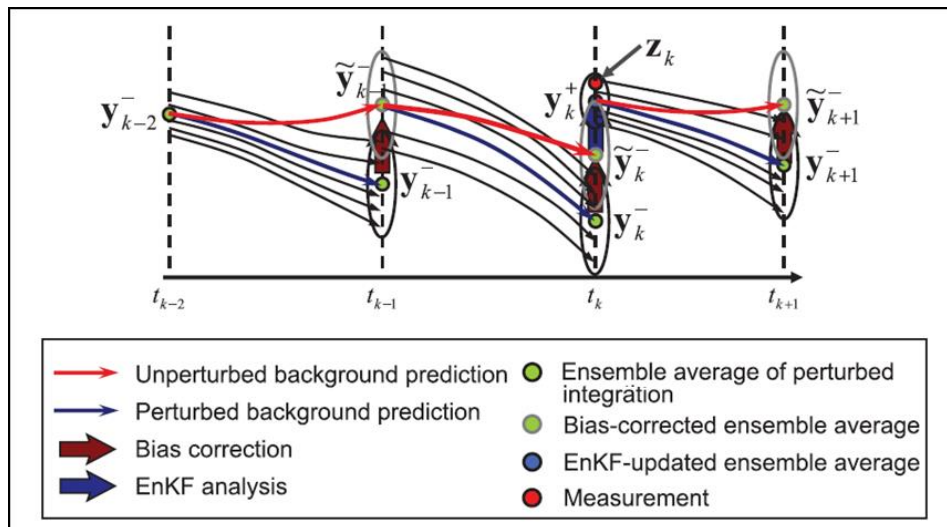


Figure 6.10. Figure adapted from Ryu et al. (2009), representing the procedure to correct the unintended perturbation bias.

6.6 EnKF implementation

The EnKF theory was described in Sect. 3.5.2.1, here its implementation for the current application is briefly showed.

Considering the following vector of states for the i -th ensemble member:

$$\mathbf{x}^i = [W_{1\%}^i \quad W_{2\%}^i]^T \quad (6.9)$$

if observation y (i.e. SWI*) is available at time t_k the state updating can be expressed as:

$$\mathbf{x}_k^{i+} = \mathbf{x}_k^{i-} + G_k(y_k + v_k^i - H\mathbf{x}_k^{i-}) \quad (6.10)$$

where:

- \mathbf{x}_k^{i-} and \mathbf{x}_k^{i+} refer to the background and analysis states, respectively;
- $H = [1 \quad 0]$ is the observation operator;
- v_k^i is the observation perturbation for the i -th ensemble member, coherently with the observation error variance;
- G_k is the 2x1 Kalman gain matrix.

The Kalman gain matrix is obtained as:

$$G_k = P_k^- H^T / (H P_k^- H^T + R) \quad (6.11)$$

where R is the observation error variance (in this case the scalar and stationary value previously estimated by TC) and P_k^- is the 2x2 background error covariance matrix, defined as:

$$P_k^- = \frac{1}{N-1} D_k D_k^T \quad (6.12)$$

with:

$$D_k = [x_k^{1-} - \bar{x}_k^-, \dots, x_k^{N-} - \bar{x}_k^-] \quad (6.13)$$

$$\bar{x}_k^- = \frac{1}{N} \sum_{i=1}^N x_k^{i-} \quad (6.14)$$

where N is the ensemble size.

6.7 Results and discussion

The impacts of assimilating SWI* obtained from different satellite products and rescaling procedures, with random error variance values depending on TC triplet configuration, are here showed. The effects of DA are evaluated by comparison with the open loop (OL) simulations, with model ensembles obtained as described in Sect. 6.5

which run without assimilation. The Root Mean Squared Error (RMSE) between simulated (in terms of ensemble mean) and observed discharges is chosen as reference metric, proposed in a standardized (and dimensionless) form to represent the results in a homogeneous way between the various basins, characterized by very different runoff regimes. The performance metric is indicated as fractional RMSE (Draper et al., 2013):

$$\text{fRMSE} = \text{RMSE}/\sigma_o \quad (6.15)$$

where σ_o is the standard deviation of observed runoff time series. The Nash-Sutcliffe efficiency index (NSE), which is one of the most common criteria to evaluate hydrological model performances, can be directly obtained from fRMSE, as:

$$\text{NSE} = 1 - \text{fRMSE}^2 \quad (6.16)$$

OL reference performances are showed in Fig. 6.11, computed by splitting the observation period in two eras, where ‘era 1’ and ‘era 2’ are 2003-2006 and 2007-2011 years respectively, to take in account the low temporal density of active observations in era 1, as mentioned in Sect. 6.2.

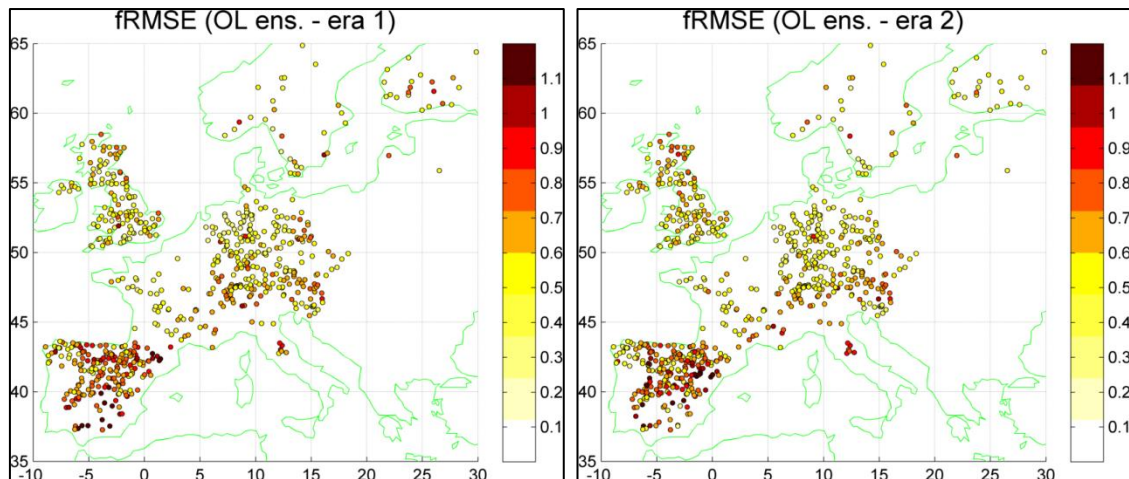


Figure 6.11. Model ensembles performances in open loop configuration, in terms of fractional RMSE, for the periods 2003-2006 (left) and 2007-2011 (right).

Figures 6.12-6.17 show variations of fRMSE due to DA, for active (only considering era 2) and passive products, according to both CDF-matching and TC rescaling methods, when LV approach is used for triplet configuration and error variance estimation. The use of combined ESA-CCI product (not showed) leads to similar results to those of active and passive datasets, at the same conditions of rescaling and error characterization approaches. In scatterplots, the points where the red lines cross refer to the medians while the red line edges represent the 25th and the 75th percentiles.

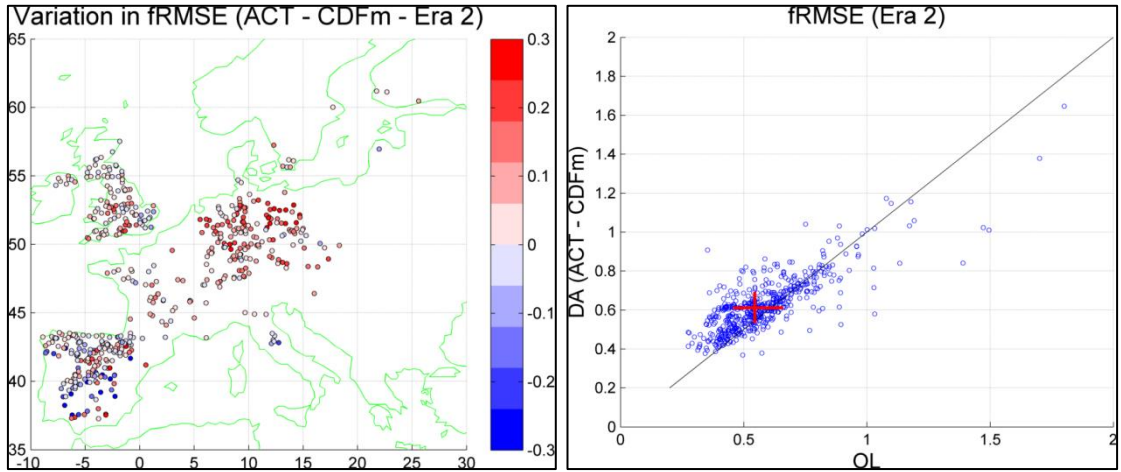


Figure 6.12. Effects of assimilating active-based SWI*, rescaled with CDF-matching approach, during era 2.

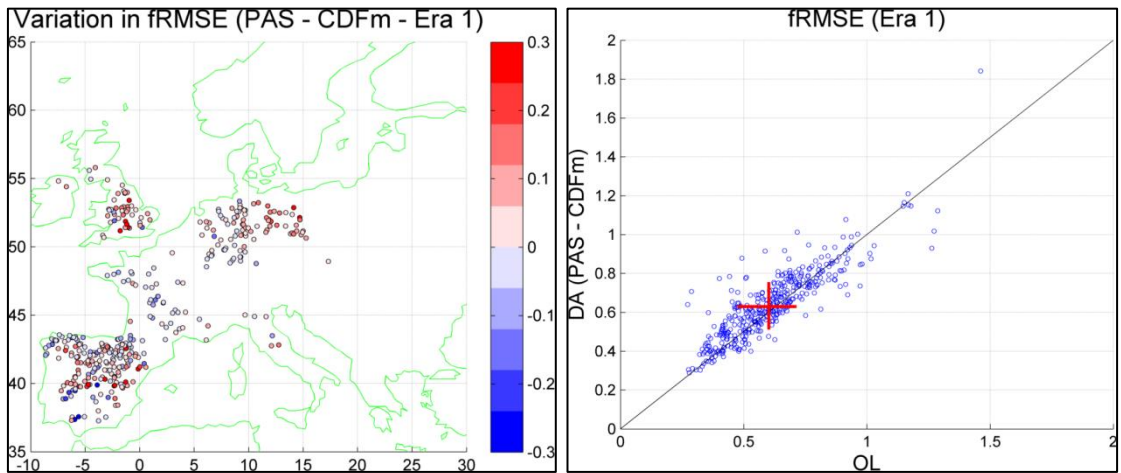


Figure 6.13. Effects of assimilating passive-based SWI*, rescaled with CDF-matching approach, during era 1.

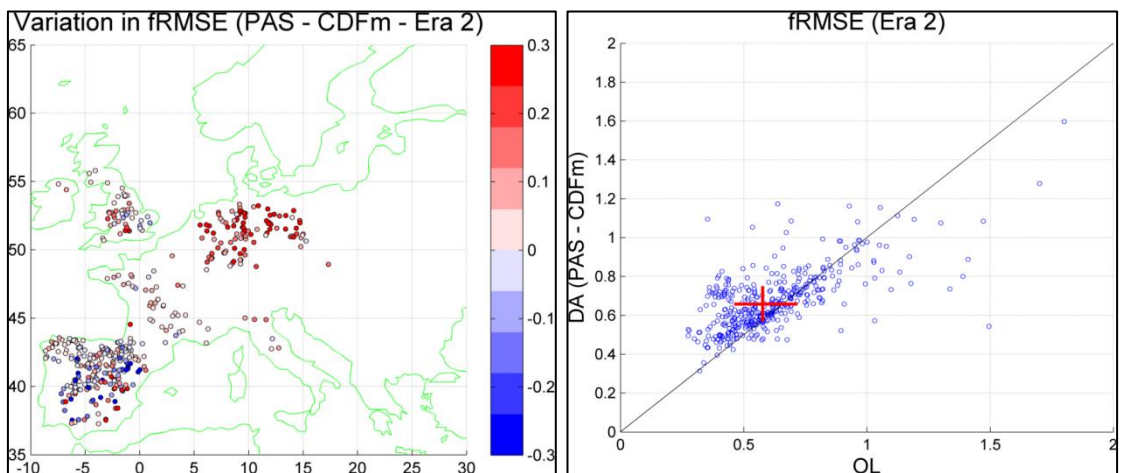


Figure 6.14. Effects of assimilating passive-based SWI*, rescaled with CDF-matching approach, during era 2.

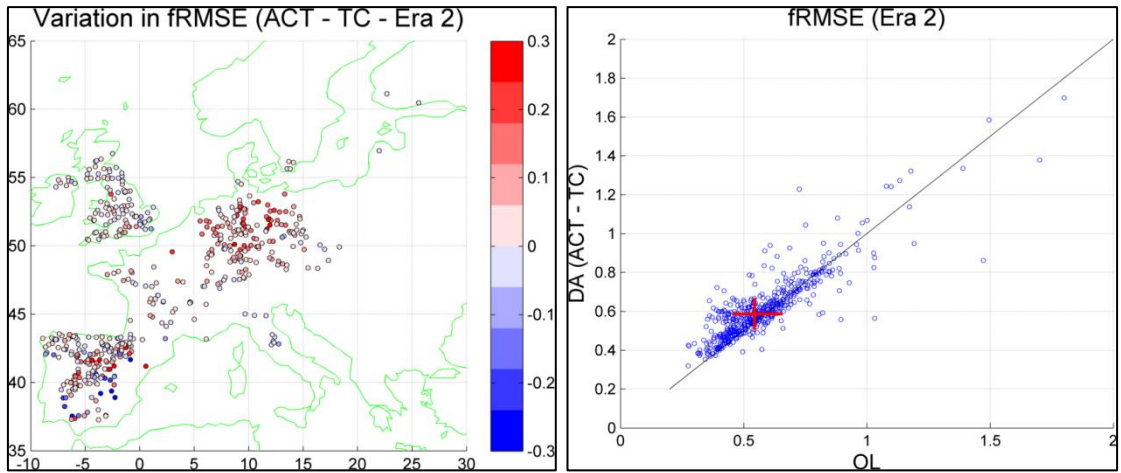


Figure 6.15. Effects of assimilating active-based SWI*, rescaled with TC approach, during era 2.

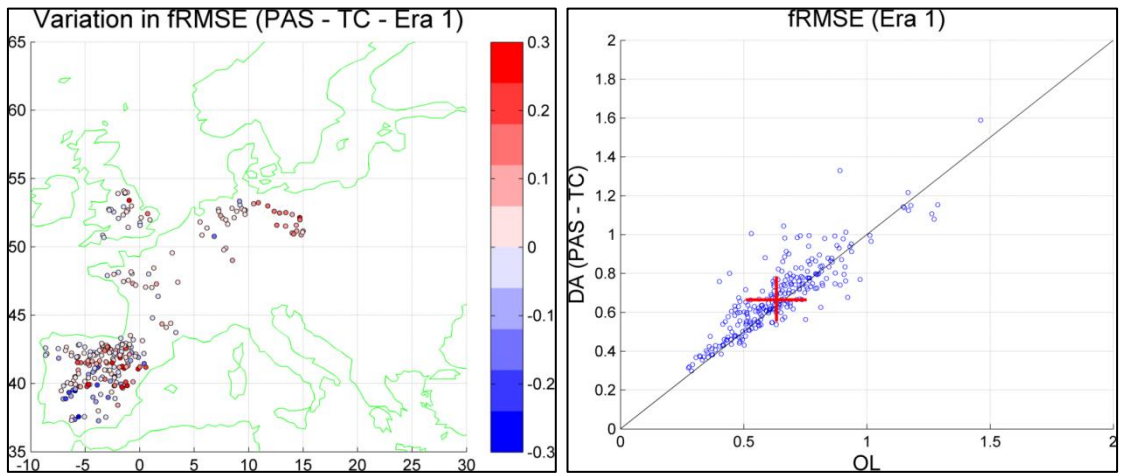


Figure 6.16. Effects of assimilating passive-based SWI*, rescaled with TC approach, during era 1.

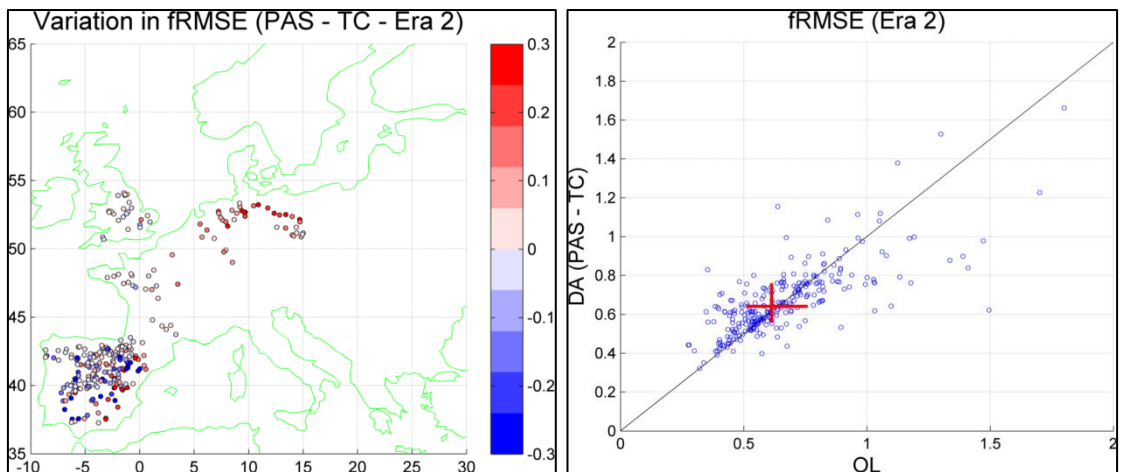


Figure 6.17. Effects of assimilating passive-based SWI*, rescaled with TC approach, during era 2.

The assimilation results generally show a degradation of OL model performance, with improvements limited to a moderate percentage of catchments, well localized around Mediterranean area (Iberian peninsula, Italy, south of France). Increment of fRMSE due to DA are instead systematically observed at northern latitudes. Spatial patterns in DA performances are consistent with those obtained during model calibration (in a reverse way, Fig. 6.4), as well as with those related to rainfall uncertainty, as it results from both E-OBS interpolation errors and ensemble test results described in Sect. 6.5. In this sense, assimilation of satellite SM confirms a greater potential in catchments where model does not work so well and/or precipitation data (which can be the main error source in runoff modelling) are not so accurate, with these two situations being often coupled in practice. On the other hand, when there is less evidence of errors in forcing data and in model parameters and structure, the added-value of remotely sensed SM observations appears to be very limited. However, the reasons why this happens must be deepened, identifying specific critical aspects and possible solutions, even in terms of preprocessing and error characterization procedures. The possibility of spatial variability in satellite SM products quality should also be considered, due to the different conditions which characterize the study catchments.

There are no remarkable differences in fRMSE performances attributable to the different ESA-CCI products or rescaling procedures (i.e. CDF-matching and TC). With regard to the triplet configuration in TC, the use of lagged variable approach lead to better results, as it assigns a lower weight to satellite observations in DA. In this sense, not only the quality of satellite-based data but also its proper characterization plays a key role during assimilation. With triplet made of model, active and passive dataset, lower satellite error variances are estimated with respect to the model in TC (Sect. 6.4), maybe due to similarities in errors generated during the processing of SSM datasets (e.g. introduced through exponential filter); if TC is performed prior to DA, the use of LV approach proved to be a valid alternative to characterize the relative uncertainty between model and observations.

Results obtained in era 2 are summarized in Fig. 6.18, where a performance comparison is made between different products (active and passive) and rescaling methods (CDF-matching and TC).

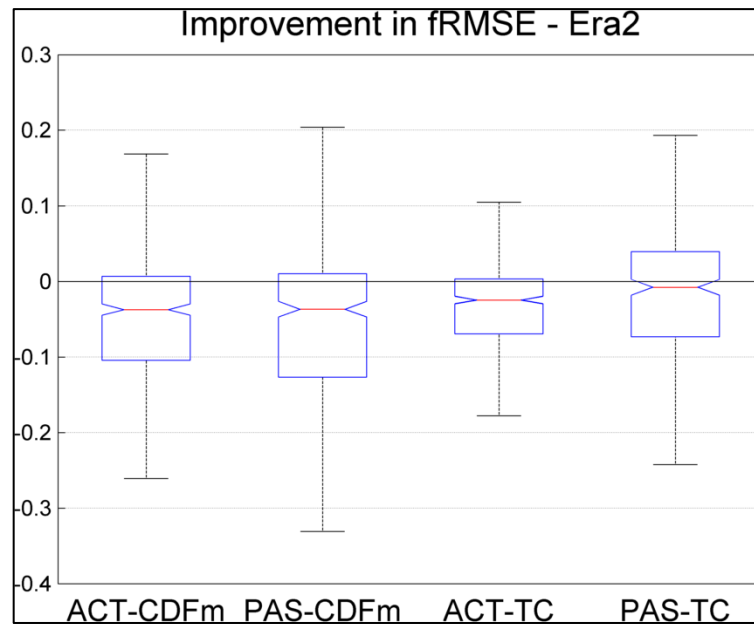


Figure 6.18. Comparison of DA performances in era 2 when different products (active, ACT, and passive, PAS) and rescaling methods (CDF-matching and TC) were used.

Finally, the two rescaling approaches are evaluated in terms of multiplicative bias in simulated discharge, by checking for the evidence of a state-dependent systematic error in runoff. With this regard, it is worth noting that efficiency indexes commonly used for the hydrologic models calibration (e.g. NSE, KGE) introduce a tendency to underestimate runoff peaks (Gupta et al., 2009). Then, DA techniques that correct for random errors in states also reduces the total variability of analysis estimates with respect to the background model predictions; this could reflect on model outputs, although a non-linear relationship exists between SM states and runoff in hydrological models. The impact of DA techniques on analysis values variability is a current research topic (Seo et al., 2018).

In Figures 6.19-6.21 simulated runoff is compared with the observed one, with discharges being normalized and a representative selection of fractiles being plotted; simulated DA discharges are those obtained through LV approach. As showed, adopting TC for rescaling led to better results in comparison with the use of the classical CDF-matching approach, and this can intuitively be traced back to a) the higher SWI* total variability generally obtained by matching only the signal part of variance with model estimates, b) the limited multiplicative biases between SWI* and model SM estimates as quantified by the magnitude of the positive slope values computed in Sect. 6.4.

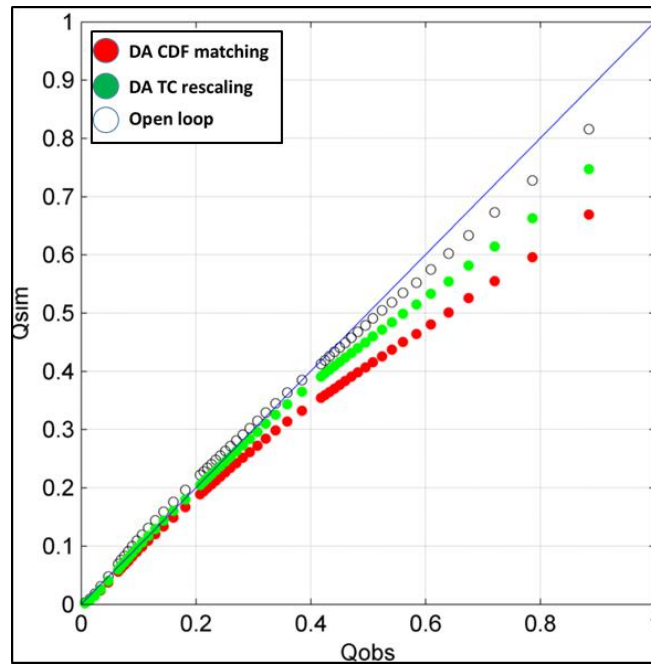


Figure 6.19. Quantile-quantile plot of simulated against observed runoff, obtained during era 2 by assimilating the active product.

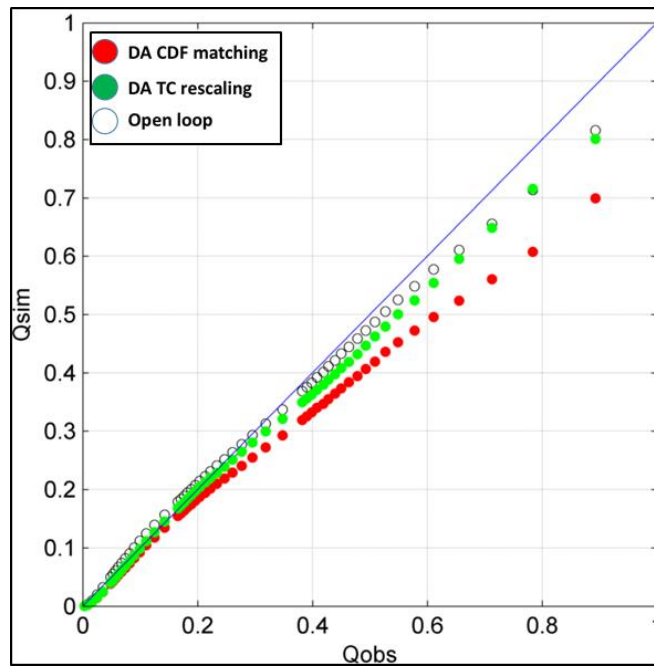


Figure 6.20. Quantile-quantile plot of simulated against observed runoff, obtained during era 1 by assimilating the passive product.

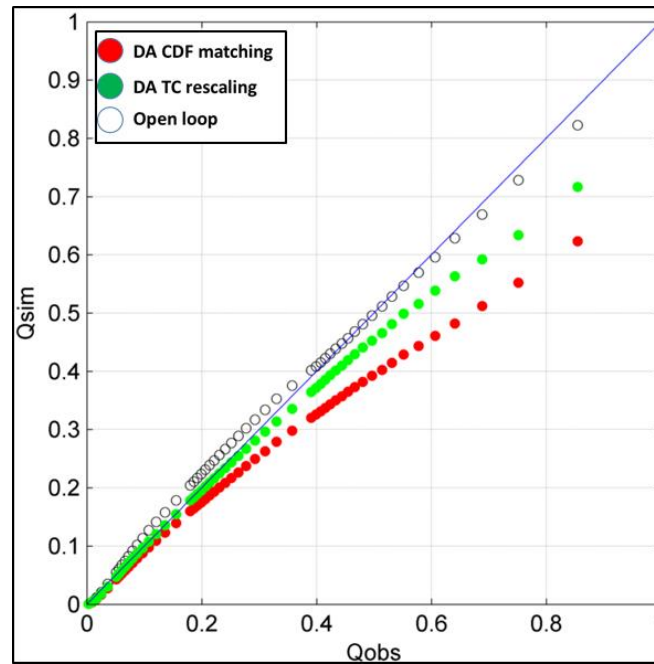


Figure 6.21. Quantile-quantile plot of simulated against observed runoff, obtained during era 2 by assimilating the passive product.

6.8 Conclusions

In this study, the effects of remotely sensed soil moisture assimilation in rainfall-runoff modelling have been extensively explored, through a large experiment involving numerous catchments located across Europe, and the use of different observations preprocessing and error characterization approaches.

The model performance in open loop can be considered generally good, while the effects of assimilation are contrasting. The improvements due to DA are substantially limited to catchments in Mediterranean area, while a degradation of model results is almost systematically observed at northern latitudes. Spatial patterns in DA performances are inversely related to those of model calibration performances and rainfall accuracies; in this sense assimilation of satellite SM shows skills where model does not work so well and/or higher errors in precipitation data could be expected. There are no remarkable differences in performances attributable to the different ESA-CCI products or rescaling procedures (i.e. CDF-matching and TC), as showed in Fig. 6.18. However, adopting TC for rescaling appears to be more effective in limiting multiplicative bias evidences in simulated discharges. The use of lagged model as third variable in error variance characterization by TC, that in this case implies higher

uncertainties attributed to satellite-based observations, lead to better DA results with respect to the integration of two satellite-based datasets in the triplet configuration.

Some critical aspects for DA performances are here seen to be the presence of bias between datasets and the incorrect observation error variance evaluation. Even the preprocessing phase (e.g. exponential filter, rescaling method) can introduce further errors in observation dataset (including systematic ones not addressed by the DA system) or worsen conflict situations between real cases and EnKF analysis scheme and assumptions. For example, here in all the preprocessing configurations the exponential filter was used, although it may have considerable limitations for DA applications, such as the loss of signal variability at short-time scale due to too high smoothing effects or the presence of time-correlated errors in SWI due to filter structure (while EnKF requires observation white noise).

In conclusion, this study confirms in some way the contrasting available results on satellite SM data assimilation in hydrological models. Here, the integration of remote sensing data seems suitable for specific areas, and shows a high potential to correct for uncertainties associated with rainfall estimates. However, there is a wide presence of negative DA performance occurrences (although often associated with already good OL model results that are difficult to improve) that must be deepened, identifying the factors that lead to non-optimal solutions, including local conditions and methodology issues.

7. CONCLUSIONS

Accurate knowledge of soil moisture (SM) is widely recognized as essential for a multitude of applications, including hydrological modelling and flood forecasting. In recent years, increasing availability of satellite data has brought great interest in the use of remotely sensed SM in data assimilation (DA) frameworks, in order to improve runoff simulations. However, contrasting results are obtained about added-value of satellite SM observations in hydrological modelling. This can be related to the reliability of satellite retrievals, as well as to some methodological issues.

In particular, this work concerns some of these issues that are related to the satellite observations, such as the low quality of soil moisture retrievals under certain conditions, the mismatch between soil depth simulated in the model and observed by the remote sensor, the bias between satellite data and model states, the assessment of the magnitude and the structure of the observation errors. Several approaches have been proposed to take these issues into account during observations preprocessing and error characterization phases that take place before DA in prediction models. However, although the significant impact on final DA performances, this general theme is still under-research.

After providing a theoretical background on soil moisture modelling and monitoring approaches, data assimilation techniques, satellite data preprocessing and error characterization methods, here the results of two applications are described, involving the use of different satellite products and an evaluation based on ground data of soil moisture and river discharge.

The first application is made up of two studies subsequently carried out, having in common the topic of error propagation in the exponential filter approach. The exponential filter is a widespread method to obtain a root-zone soil water index (SWI) from remotely-sensed surface soil moisture (SSM) observations; however, the impact of some factors involved in SWI formulation, that can introduce inaccuracies in the outputs, have not been adequately detailed up to now. An analytical error propagation (EP) scheme was proposed, aimed to assess the time-variant error variance of the exponential filter outputs taking due consideration of some shortcomings of the method. In the first study, a simplified scheme was used, taking into account the errors as well as

the availability of the underlying SSM data in SWI estimate, but neglecting the uncertainty in exponential filter parameter, to obtain a standard error estimate indicated as SWI noise. ASCAT SSM retrievals were employed to test the proposed EP scheme, with datasets that include time-variant standard error estimates (referred to as SSM noise). The SWI noise was addressed to be used in a quality check framework, in the context of a more general evaluation of the impacts of different data masking procedures. ASCAT-derived SWI showed a good agreement with reference root zone ground data, even when no quality control was performed, with the exponential filter generally showing a greater ability in capturing the seasonal soil moisture behaviour rather than short time-scale fluctuations. Different SSM data masking procedures, based on other indicators available for the ASCAT product, were analysed; however, they did not necessarily led to a more accurate description of the root-zone water content in terms of agreement with in situ observations. Regarding the EP scheme, the use of thresholds on SWI noise led to performance worsening compared to the configuration without quality controls, due to SSM noise seasonal pattern that implies the removal of SWI values mainly concentrated in a season characterized by a better relative agreement with in situ measurements. However, removing data according to ‘a posteriori’ control on SWI noise generally improved the performance metrics with respect to ‘a priori’ SSM noise threshold application. In the second study, error in exponential filter parameter was taken in account, and a successive release of ASCAT SSM dataset was employed, with no prominent seasonal patterns in SSM noise. In this case, the proposed EP scheme has shown capabilities to identify potentially less reliable SWI values in the selected study sites, improving performance metrics respect to the configuration without quality controls. The preliminary results suggest the utility of EP approach in the SWI evaluation, in the comprehension of the exponential filter shortcomings, and in the time-variant SWI error characterization. Due to the exponential filter mathematical structure, error estimates could be not reliable in magnitude; however, they can help to discriminate more and less certain data. The information provided by analytical error propagation equations can be suitable not only for a preliminary screening analysis, but also to reproduce the trend, rather than the actual values, of satellite-based SM error variance.

The second application was mainly carried out during a research period at CNR-IRPI with the Hydrology group, that provides datasets and support. In this study, a data

assimilation experiment was made on a large number of catchments located across Europe, in order to assess the effects of integrating remotely sensed SM on a rainfall-runoff model performances, and the use of different observations preprocessing and error characterization approaches has been tested. ESA-CCI SM products are employed, that merged several available active and passive measurements, in order to also evaluate the role of the sensors on DA performances. In the preprocessing phase, the exponential filter is used to address the depth mismatch, while two alternative approaches are considered for satellite data rescaling, i.e. CDF-matching and Triple Collocation analysis (TC), that imply the matching of total variance and only signal component, respectively, between reference and rescaled datasets. Then, TC is also used for observation error characterization, considering different triplet configurations also in order to test the impact of different observation weights in data assimilation system performances. Finally, the Ensemble Kalman Filter (EnKF) is employed to assimilate the rescaled satellite-based SWI into the MISDc-2L hydrological model. The model performance in open loop can be considered generally good, while the effects of assimilation are contrasting. The improvements due to DA are substantially limited to catchments in Mediterranean area, while a degradation of model results is almost systematically observed at northern latitudes. Spatial patterns in DA performances are inversely related to those of model calibration performances and rainfall accuracies; in this sense assimilation of satellite SM shows skills where model does not work so well and/or higher errors in precipitation data could be expected. There are no remarkable differences in performances attributable to the different ESA-CCI products or rescaling procedures (i.e. CDF-matching and TC). However, adopting TC for rescaling appears to be more effective in limiting multiplicative bias evidences in simulated discharges. The use of lagged model as third variable in error variance characterization by TC, that in this case implies higher uncertainties attributed to satellite-based observations, lead to better DA results with respect to the integration of two satellite-based datasets in the triplet configuration. In conclusion, the second application confirms in some way the contrasting available results on satellite SM data assimilation in hydrological models. Here, the integration of remote sensing data seems suitable for specific areas, and shows a high potential to correct for uncertainties associated with rainfall estimates. However, there is a wide presence of negative DA performance occurrences (although often associated with already good OL model results that are difficult to improve) that must

be deepened, identifying the factors that lead to non-optimal solutions, including local conditions and methodology issues. Some critical aspects for DA performances are here seen to be the presence of bias between datasets and the incorrect observation error variance evaluation.

The issues related to observation preprocessing and error characterization are confirmed to have a great impact on the performance of data assimilation systems, with even the same relevance of the considered observational soil moisture dataset or data assimilation algorithm. In this sense, it seems necessary to put a lot of efforts into these aspects to fully utilize the great potential offered by remote sensing data for specific applications such as runoff modelling. The results of this study indicate that a more aware and critical use of existing procedures is appropriate, as they have advantages and limitations also depending to the specific application (e.g. assimilation in rainfall-runoff for extreme predictions), without excluding the opportunity to propose new solutions also on the basis of the experiences provided by those currently used.

ANNEX 1

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
1	YVEL	FR_6118015	France	47.99	-2.37	315
2	EVEL	FR_6118060	France	47.90	-2.98	316
3	MEU	FR_6118210	France	48.13	-1.94	468
4	GAVE DE PAU	FR_6119010	France	43.51	-0.85	2575
5	GAVE D'OSSAU	FR_6119020	France	43.19	-0.60	488
6	SAISON	FR_6119040	France	43.25	-0.87	480
7	ARROS	FR_6119120	France	43.13	0.26	173
8	LEYRE	FR_6119200	France	44.55	-0.87	1650
9	SEINE	FR_6122102	France	47.41	3.49	371
10	OISE (TRIB. SEINE)	FR_6122141	France	49.56	2.99	4290
11	EURE	FR_6122150	France	48.45	1.29	330
12	OURCE	FR_6122200	France	47.35	3.96	248
13	ERDRE	FR_6123170	France	47.46	-1.48	472
14	ERDRE	FR_6123171	France	47.56	-1.05	169
15	LAYON	FR_6123180	France	47.32	-0.63	920
16	LAYON	FR_6123181	France	47.19	-0.37	250
17	JOUANNE	FR_6123190	France	48.03	-0.71	410
18	ERNEE	FR_6123200	France	48.17	-0.78	375
19	BRAYE	FR_6123220	France	48.01	0.82	270
20	INDRE	FR_6123250	France	47.02	1.13	1712
21	SAULDRE	FR_6123350	France	47.29	1.53	2254
22	VIENNE	FR_6123400	France	47.05	0.54	19920
23	GLANE	FR_6123420	France	45.91	0.92	288
24	BRIANCE	FR_6123430	France	45.76	1.24	597
25	CREUSE	FR_6123450	France	46.38	1.68	1235
26	CREUSE	FR_6123451	France	45.89	2.17	165
27	PETITE CREUSE	FR_6123460	France	46.39	1.69	850
28	ROZEILLE	FR_6123470	France	45.93	2.18	186
29	ALAGNON	FR_6123640	France	45.38	3.26	984
30	ALAGNON	FR_6123641	France	45.16	3.03	310
31	SENOUIRE	FR_6123650	France	45.20	3.52	155
32	DUNIERE	FR_6123770	France	45.21	4.21	228
33	VINCOU	FR_6123800	France	46.13	1.02	286
34	GARTEMPE	FR_6123820	France	46.11	1.43	570
35	CEOU	FR_6124300	France	44.79	1.17	603
36	VEZERE	FR_6124501	France	44.90	0.96	3736
37	RUISSEAU ALZOU	FR_6125610	France	44.36	2.05	199
38	AVEYRON	FR_6125620	France	44.40	2.85	270
39	GIFFOU	FR_6125720	France	44.11	2.43	175
40	AUDE	FR_6128100	France	43.21	2.36	1770
41	LERGUE	FR_6128350	France	43.73	3.32	228

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
42	ARRE	FR_6128380	France	44.00	3.66	159
43	REAL MARTIN	FR_6128630	France	43.19	6.11	277
44	SEILLE (TRIB. MOSELLE)	FR_6136100	France	49.10	6.19	1280
45	MADON	FR_6136150	France	48.54	6.13	940
46	MEURTHE	FR_6136160	France	48.28	6.96	374
47	MOSELLE RIVER	FR_6136200	France	48.16	6.45	1220
48	MOSELLE RIVER	FR_6136201	France	48.07	6.61	621
49	MOSELLE RIVER	FR_6136202	France	47.91	6.69	153
50	GARDON DE SAINT-JEAN	FR_6139060	France	44.07	3.96	263
51	BEZ	FR_6139130	France	44.69	5.49	227
52	ROUBION	FR_6139220	France	44.63	5.02	186
53	BES (TRIB. RHONE)	FR_6139255	France	44.22	6.28	165
54	UBAYE	FR_6139260	France	44.45	6.40	946
55	AZERGUES	FR_6139360	France	45.89	4.62	336
56	AZERGUES	FR_6139361	France	45.86	4.69	792
57	BREVENNE	FR_6139365	France	45.81	4.60	219
58	OGNON (TRIB. RHONE)	FR_6139681	France	47.41	6.18	1250
59	LOUE	FR_6139790	France	47.04	5.81	1380
60	LISON	FR_6139795	France	47.03	5.96	217
61	CHERAN	FR_6139825	France	45.72	6.10	249
62	BEROUNKA	CZ_6140250	Czech Rep.	49.96	14.09	8284
63	OTAVA	CZ_6140300	Czech Rep.	49.30	14.15	2913
64	ELBE RIVER	CZ_6140400	Czech Rep.	50.79	14.23	51123
65	SAZAVA	CZ_6140450	Czech Rep.	49.74	15.10	1421
66	JIZERA	CZ_6140480	Czech Rep.	50.24	14.78	2159
67	JIZERA	CZ_6140500	Czech Rep.	50.64	15.28	791
68	DIVOKA ORLICE	CZ_6140700	Czech Rep.	50.07	16.55	182
69	MORAVA	CZ_6142100	Czech Rep.	49.75	16.97	1559
70	THAYA	CZ_6142110	Czech Rep.	48.80	16.86	12283
71	MORAVA	CZ_6142120	Czech Rep.	48.93	17.32	9146
72	BECVA	CZ_6142300	Czech Rep.	49.53	17.75	1275
73	ODER RIVER	CZ_6157100	Czech Rep.	49.92	18.33	4665
74	ENNINGDALSÆLVEN	SE_6229100	Sweden	58.88	11.54	624.1
75	GOETA ÆLV	SE_6229500	Sweden	58.36	12.37	46885.5
76	VISKAN	SE_6233100	Sweden	57.24	12.31	2160.2
77	KLINGVAÆLSAN	SE_6233140	Sweden	55.64	13.54	191.6
78	FYLLEAN	SE_6233150	Sweden	56.72	13.12	259.7
79	KAÆVLINGEAN	SE_6233160	Sweden	55.65	13.82	262.3
80	LAGAN (SWEDEN)	SE_6233170	Sweden	56.49	13.51	5479.5
81	HOEJE A	SE_6233190	Sweden	55.70	13.15	237
82	GROETSJOEN	SE_6233200	Sweden	61.81	12.44	565
83	DALAÆLVEN	SE_6233201	Sweden	60.56	17.44	28920.5
84	TAENNAN	SE_6233220	Sweden	62.54	12.35	226.6
85	LJUSNAN	SE_6233227	Sweden	62.55	12.60	340.3
86	HOERLINGEAN	SE_6233230	Sweden	56.19	13.68	202.7

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
87	HELGE A	SE_6233250	Sweden	56.10	14.13	3664.5
88	MOTALA STROEM	SE_6233301	Sweden	58.59	16.17	15384
89	ALSTERAN	SE_6233350	Sweden	57.01	16.16	1332.7
90	EMAN	SE_6233360	Sweden	57.14	16.45	4446
91	FYRSJOEN	SE_6233400	Sweden	63.52	15.39	2428.4
92	MAELAREN	SE_6233410	Sweden	59.30	18.08	22638.8
93	ANKARVATTNET	SE_6233450	Sweden	64.86	14.21	427.8
94	VATTHOLMAAN	SE_6233600	Sweden	60.02	17.73	293.8
95	KALIXAELVEN	SE_6233850	Sweden	66.17	22.82	23102.9
96	TORNEAELVEN, TORNIONJOKI	SE_6233911	Sweden	67.23	23.35	11038.1
97	BREGENZER ACH	AT_6235100	Austria	47.36	9.88	228.6
98	ALM	AT_6242110	Austria	48.05	13.92	445
99	ILZBACH	AT_6242200	Austria	47.08	15.94	190.1
100	SALZA	AT_6242240	Austria	47.74	15.31	280
101	ENNS	AT_6242250	Austria	48.04	14.43	5915.4
102	STEYR	AT_6242260	Austria	47.77	14.17	184.9
103	PALTENBACH	AT_6242290	Austria	47.55	14.31	368.7
104	DANUBE RIVER	AT_6242401	Austria	48.38	15.46	95970
105	KREMS	AT_6242420	Austria	48.45	15.57	305.9
106	DANUBE RIVER	AT_6242501	Austria	48.36	16.34	101536.6
107	TRAUN (TRIB. DANUBE)	AT_6242600	Austria	47.81	13.77	1257.6
108	KITZBUEHLER ACHE	AT_6243200	Austria	47.46	12.39	153
109	KITZBUEHLER ACHE	AT_6243201	Austria	47.52	12.42	332.4
110	BERCHDESGADENER ACHE	AT_6243240	Austria	47.73	13.04	428.2
111	BRIXENTALER ACHE	AT_6243300	Austria	47.49	12.10	322.3
112	OETZTALER ACHE	AT_6243400	Austria	47.16	10.91	785.5
113	ANTIESEN	AT_6243800	Austria	48.27	13.45	164.9
114	INN	AT_6243850	Austria	48.44	13.44	25663.8
115	LIESINGBACH	AT_6246100	Austria	47.39	14.91	265.6
116	LAVANT	AT_6246110	Austria	46.66	14.95	954.5
117	GAIL	AT_6246130	Austria	46.62	13.25	594.9
118	ISEL	AT_6246160	Austria	46.97	12.55	518.4
119	SCHWARZACH	AT_6246170	Austria	46.92	12.51	268.1
120	MUR	AT_6246611	Austria	47.16	15.32	6791.5
121	MUR	AT_6246612	Austria	47.41	15.28	6214
122	TAURACH	AT_6246630	Austria	47.14	13.80	377.9
123	MUERZ	AT_6246700	Austria	47.53	15.47	727.7
124	MUERZ	AT_6246701	Austria	47.66	15.59	229.4
125	RHINE RIVER	DE_6335020	Germany	51.76	6.40	159300
126	RUHR	DE_6335030	Germany	51.40	7.16	4078
127	RUHR	DE_6335031	Germany	51.44	7.58	1988
128	RUHR	DE_6335032	Germany	51.35	8.28	425
129	SWIST	DE_6335040	Germany	50.77	6.85	285
130	SIEG	DE_6335045	Germany	50.80	7.16	2832
131	SIEG	DE_6335046	Germany	50.78	7.44	1472

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
132	RHINE RIVER	DE_6335050	Germany	51.23	6.77	147680
133	RHINE RIVER	DE_6335060	Germany	50.94	6.96	144232
134	RHINE RIVER	DE_6335070	Germany	50.44	7.39	139549
135	WIED	DE_6335076	Germany	50.50	7.45	697
136	LIPPE	DE_6335081	Germany	51.66	8.09	2005
137	LIPPE	DE_6335082	Germany	51.75	8.63	1018
138	RHINE RIVER	DE_6335100	Germany	50.09	7.76	103488
139	NAHE	DE_6335115	Germany	49.91	7.91	4013
140	NAHE	DE_6335116	Germany	49.78	7.72	2832
141	ALSENZ	DE_6335117	Germany	49.78	7.83	318
142	KINZIG (TRIB. RHEIN)	DE_6335125	Germany	48.39	8.03	954
143	RHINE RIVER	DE_6335150	Germany	50.00	8.28	98206
144	RHINE RIVER	DE_6335170	Germany	49.32	8.45	53131
145	RHINE RIVER	DE_6335180	Germany	49.64	8.38	68827
146	MAIN	DE_6335240	Germany	49.72	9.22	21505
147	KOCHER	DE_6335290	Germany	49.26	9.29	1929
148	KOCHER	DE_6335291	Germany	49.00	9.77	726
149	MAIN	DE_6335301	Germany	50.03	10.22	12715
150	MAIN	DE_6335302	Germany	50.01	9.60	17914
151	MAIN	DE_6335303	Germany	49.93	10.76	12010
152	MAIN	DE_6335304	Germany	50.11	8.71	24764
153	LAHN	DE_6335350	Germany	50.55	8.36	3571
154	LAHN	DE_6335351	Germany	50.80	8.76	1667
155	RHINE RIVER	DE_6335400	Germany	47.56	7.80	34550
156	WUTACH	DE_6335410	Germany	47.62	8.33	627.13
157	ARGEN	DE_6335450	Germany	47.63	9.60	639.34
158	SCHUSSEN	DE_6335460	Germany	47.67	9.53	782.01
159	MAIN	DE_6335500	Germany	49.80	9.93	14031
160	TAUBER	DE_6335520	Germany	49.63	9.67	1584
161	REGNITZ	DE_6335530	Germany	49.83	10.94	7005
162	RODACH	DE_6335540	Germany	50.18	11.23	713
163	NECKAR	DE_6335600	Germany	49.44	9.01	12710
164	NECKAR	DE_6335601	Germany	49.07	9.15	7916
165	PFINZ	DE_6335640	Germany	49.01	8.52	235
166	ENZ	DE_6335660	Germany	48.90	8.73	1476
167	MURR	DE_6335675	Germany	48.96	9.26	506
168	JAGST	DE_6335680	Germany	48.93	10.14	178.59
169	JAGST	DE_6335681	Germany	49.27	9.22	1825
170	REMS	DE_6335690	Germany	48.85	9.32	569
171	MURG	DE_6335710	Germany	48.82	8.30	468.8
172	WIESE	DE_6335730	Germany	47.70	7.85	209
173	MAIN	DE_6335800	Germany	49.95	10.87	4251
174	KRAICHBACH	DE_6335830	Germany	49.16	8.63	161.1
175	MOSELLE RIVER	DE_6336050	Germany	50.14	7.17	27088
176	MOSELLE RIVER	DE_6336500	Germany	49.73	6.63	23857

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
177	MOSELLE RIVER	DE_6336800	Germany	49.47	6.37	11522
178	SAAR RIVER	DE_6336900	Germany	49.41	6.65	6983
179	NIED	DE_6336910	Germany	49.34	6.59	1337.4
180	HUNTE	DE_6337050	Germany	52.83	8.47	1313
181	WUEMME	DE_6337060	Germany	53.08	9.21	955
182	WESER	DE_6337100	Germany	52.18	8.86	17618
183	ALLER	DE_6337250	Germany	52.79	9.38	14730
184	HASEL	DE_6337340	Germany	50.55	10.48	321
185	SCHLEUSE	DE_6337350	Germany	50.50	10.72	256
186	WESER	DE_6337400	Germany	51.43	9.64	12442
187	HAUNE	DE_6337410	Germany	50.81	9.73	421.8
188	ALLER	DE_6337501	Germany	52.68	9.70	7209
189	EDER	DE_6337504	Germany	51.16	8.90	1202
190	EDER	DE_6337505	Germany	51.17	9.09	1452
191	FULDA	DE_6337506	Germany	51.19	9.50	2975
192	FULDA	DE_6337507	Germany	51.23	9.47	6366
193	FULDA	DE_6337508	Germany	51.00	9.72	2523
194	LEINE	DE_6337509	Germany	52.39	9.68	5304
195	LEINE	DE_6337510	Germany	52.68	9.60	6443
196	WERRA	DE_6337512	Germany	51.13	10.20	4302
197	WERRA	DE_6337513	Germany	51.41	9.71	5487
198	WESER	DE_6337514	Germany	51.97	9.52	15924
199	WESER	DE_6337515	Germany	52.85	9.21	22110
200	WESER	DE_6337516	Germany	51.65	9.44	14794
201	WESER	DE_6337517	Germany	52.59	9.11	19910
202	WESER	DE_6337518	Germany	52.25	8.92	19162
203	WESER	DE_6337519	Germany	51.63	9.52	12996
204	EDER	DE_6337520	Germany	51.04	8.62	489.7
205	LEINE	DE_6337542	Germany	51.38	9.97	275
206	GROSSE AUE	DE_6337550	Germany	52.60	8.90	1014
207	WERRE	DE_6337570	Germany	52.13	8.67	874
208	SCHWALM (DE-HE)	DE_6337600	Germany	50.85	9.28	250
209	EMS	DE_6338100	Germany	52.74	7.24	8369
210	EMS	DE_6338110	Germany	52.60	7.25	4981
211	EMS	DE_6338120	Germany	52.09	7.60	2842
212	EMS	DE_6338130	Germany	52.29	7.43	3740
213	EMS	DE_6338150	Germany	51.97	7.90	1499
214	OSTE	DE_6338250	Germany	53.34	9.17	604
215	SOHOLMER AU	DE_6338270	Germany	54.70	9.02	352
216	TREENE	DE_6338800	Germany	54.51	9.32	481
217	ILMENAU	DE_6340050	Germany	53.15	10.46	1434
218	ELBE RIVER	DE_6340110	Germany	53.23	10.89	131950
219	ELBE RIVER	DE_6340120	Germany	51.06	13.74	53096
220	ELBE RIVER	DE_6340130	Germany	51.86	12.65	61879
221	ELBE RIVER	DE_6340140	Germany	51.99	11.88	94060

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
222	ELBE RIVER	DE_6340150	Germany	52.99	11.76	123532
223	ELBE RIVER	DE_6340160	Germany	52.55	11.98	97780
224	ELBE RIVER	DE_6340170	Germany	51.86	12.06	70093
225	ELBE RIVER	DE_6340180	Germany	52.14	11.64	94942
226	ELBE RIVER	DE_6340190	Germany	51.56	13.01	55211
227	UNSTRUT	DE_6340200	Germany	51.23	11.68	6218
228	ALSTER (BINNEN-ALSTER)	DE_6340210	Germany	53.66	10.09	320.5
229	SAALE	DE_6340300	Germany	51.92	11.81	23719
230	SAALE	DE_6340301	Germany	51.52	11.95	17979
231	SAALE	DE_6340303	Germany	50.32	11.91	521
232	ZORGE	DE_6340315	Germany	51.51	10.78	303.6
233	ILM	DE_6340320	Germany	51.07	11.58	894.3
234	MUEGLITZ	DE_6340410	Germany	50.95	13.85	198
235	WESENITZ	DE_6340420	Germany	51.02	13.99	227
236	LACHSBACH	DE_6340430	Germany	50.94	14.13	267
237	HAVEL	DE_6340501	Germany	52.61	12.35	19288
238	HAVEL	DE_6340510	Germany	52.48	12.85	16173
239	VEREINIGTE MULDE	DE_6340600	Germany	51.59	12.58	6171
240	SPREE	DE_6340610	Germany	52.37	14.00	6171
241	SPREE	DE_6340611	Germany	52.51	13.41	9707
242	SPREE	DE_6340612	Germany	51.16	14.41	275.9
243	ZWICKAUER MULDE	DE_6340621	Germany	50.74	12.49	1029.7
244	FLOEHA / FLAJSKY PATOK	DE_6340628	Germany	50.70	13.24	385
245	PRESSNITZ	DE_6340635	Germany	50.63	13.08	206
246	WEISSE ELSTER	DE_6340670	Germany	50.33	12.26	171
247	SCHWARZE ELSTER	DE_6340700	Germany	51.52	13.40	3184
248	SPREE	DE_6340800	Germany	51.58	14.37	2092
249	TOLLENSE	DE_6341500	Germany	53.79	13.31	1403
250	WOERNITZ	DE_6342130	Germany	48.78	10.69	1578
251	ILLER	DE_6342200	Germany	47.73	10.32	954.6
252	AITRACH	DE_6342230	Germany	47.88	10.04	308.1
253	DANUBE RIVER	DE_6342502	Germany	48.27	9.73	4036
254	LECH	DE_6342510	Germany	48.41	10.89	3800
255	LECH	DE_6342512	Germany	47.70	10.80	1713.9
256	LECH	DE_6342513	Germany	48.04	10.88	2295
257	DANUBE RIVER	DE_6342600	Germany	49.02	12.14	35399
258	NAAB	DE_6342610	Germany	49.12	11.94	5426
259	NAAB	DE_6342611	Germany	49.54	12.15	2004
260	AMPER	DE_6342650	Germany	48.46	11.87	3043
261	GLONN	DE_6342655	Germany	48.42	11.52	392
262	VILS	DE_6342670	Germany	48.61	12.69	719.5
263	DANUBE RIVER	DE_6342800	Germany	48.68	13.12	47496
264	SCHMIECHA	DE_6342810	Germany	48.09	9.15	155
265	ILZ	DE_6342830	Germany	48.69	13.45	762
266	DANUBE RIVER	DE_6342900	Germany	48.58	13.50	76653

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
267	DANUBE RIVER	DE_6342920	Germany	48.88	12.75	37687
268	ISAR	DE_6342925	Germany	48.67	12.69	8467
269	ISAR	DE_6342927	Germany	47.78	11.54	1558.8
270	ISAR	DE_6342928	Germany	47.44	11.27	404
271	LOISACH	DE_6342931	Germany	47.50	11.06	393.5
272	DANUBE RIVER	DE_6342970	Germany	48.07	9.40	2647.01
273	INN	DE_6343100	Germany	48.06	12.23	11983
274	SALZACH	DE_6343500	Germany	48.16	12.83	6649
275	TRAUN (TRIB. INN)	DE_6343530	Germany	47.99	12.54	367.4
276	INN	DE_6343900	Germany	48.56	13.44	26084
277	PO	IT_6348400	Italy	45.02	9.67	42030
278	PO	IT_6348500	Italy	44.90	10.55	55183
279	PO	IT_6348800	Italy	44.88	11.60	70091
280	ODER RIVER	DE_6357010	Germany	52.87	14.14	109564
281	ODER RIVER	DE_6357500	Germany	52.15	14.69	52033
282	NEISSE RIVER / NYSA LUZYCKA	DE_6357501	Germany	51.97	14.71	4125
283	NEISSE RIVER / NYSA LUZYCKA	DE_6357502	Germany	51.16	15.00	1621
284	NEISSE RIVER / NYSA LUZYCKA	DE_6357503	Germany	50.87	14.82	376
285	PLIESSNITZ	DE_6357510	Germany	51.06	14.94	162
286	MEUSE	NL_6421102	Netherlands	51.38	6.17	26040
287	MEUSE	NL_6421500	Netherlands	50.87	5.72	21301
288	MEUSE	NL_6421501	Netherlands	51.82	5.54	28980
289	RHINE RIVER	NL_6435060	Netherlands	51.84	6.11	160800
290	SAVA	SI_6545050	Slovenia	45.89	15.61	10186
291	SAVA	SI_6545101	Slovenia	46.12	15.09	5177
292	SAVA	SI_6545102	Slovenia	46.08	14.58	2285
293	SAVA	SI_6545190	Slovenia	46.34	14.17	908
294	SAVINJA	SI_6545300	Slovenia	46.09	15.18	1842
295	KOLPA	SI_6545500	Slovenia	45.63	15.32	2002
296	KOLPA	SI_6545501	Slovenia	45.46	14.85	460
297	SAVINJA	SI_6545600	Slovenia	46.15	15.23	1664
298	SAVINJA	SI_6545601	Slovenia	46.23	15.26	1189
299	MUR	SI_6546610	Slovenia	46.68	16.00	10197
300	PESNICA	SI_6546810	Slovenia	46.41	16.03	478
301	SOCA / INSONZO	SI_6549100	Slovenia	45.98	13.66	1573
302	SOCA / INSONZO	SI_6549180	Slovenia	46.33	13.58	325
303	CAMOWEN	GB_6603100	UK	54.61	-7.29	276.6
304	MOURNE	GB_6603120	UK	54.82	-7.46	1843.8
305	SILLEES	GB_6603130	UK	54.31	-7.68	166.3
306	LOWER BANN	GB_6603300	UK	54.98	-6.54	5209.8
307	MOYOLA	GB_6603310	UK	54.76	-6.51	304.3
308	BLACKWATER (N. IRELAND)	GB_6603320	UK	54.41	-6.73	970.2
309	SIX-MILE WATER	GB_6603350	UK	54.72	-6.23	277.6
310	LAGAN	GB_6603500	UK	54.56	-5.95	491.6
311	EWE	GB_6604100	UK	57.76	-5.60	441.1

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
312	SHIEL	GB_6604160	UK	56.77	-5.82	256
313	ORCHY	GB_6604180	UK	56.44	-4.86	251.2
314	CLYDE (SCOTLAND)	GB_6604201	UK	55.80	-4.07	1704.2
315	LEVEN (SCOTLAND)	GB_6604220	UK	55.99	-4.58	784.3
316	IRVINE	GB_6604230	UK	55.60	-4.63	380.7
317	GIRVAN	GB_6604240	UK	55.26	-4.81	245.5
318	LUCE	GB_6604250	UK	54.90	-4.84	171
319	CREE (SCOTLAND)	GB_6604260	UK	54.96	-4.48	368
320	URR	GB_6604270	UK	54.93	-3.84	199
321	NITH (SCOTLAND)	GB_6604280	UK	55.28	-3.80	471
322	ANNAN	GB_6604500	UK	55.02	-3.27	925
323	LYON	GB_6604605	UK	56.61	-3.98	391.1
324	TAY	GB_6604610	UK	56.51	-3.39	4587.1
325	NORTH ESK (SCOTLAND)	GB_6604620	UK	56.77	-2.49	732
326	EARN	GB_6604630	UK	56.35	-3.55	782.2
327	TEITH	GB_6604640	UK	56.19	-4.06	517.7
328	ALLAN WATER	GB_6604645	UK	56.22	-3.95	161
329	SPEY (SCOTLAND)	GB_6604650	UK	57.55	-3.14	2861.2
330	DULNAIN	GB_6604655	UK	57.30	-3.70	272.2
331	TWEED (SCOTLAND)	GB_6604690	UK	55.59	-2.80	1500
332	LYNE WATER	GB_6604710	UK	55.65	-3.28	175
333	TWEED (SCOTLAND)	GB_6604750	UK	55.72	-2.16	4390
334	WHITEADDER WATER	GB_6604760	UK	55.79	-2.19	503
335	DEE (SCOTLAND)	GB_6604800	UK	57.05	-2.60	1370
336	GAIRN	GB_6604815	UK	57.06	-3.08	150
337	DEE (SCOTLAND)	GB_6604820	UK	57.08	-2.33	1844
338	DON (SCOTLAND)	GB_6604830	UK	57.22	-2.19	1273
339	UGIE	GB_6604840	UK	57.53	-1.83	325
340	DEVERON	GB_6604850	UK	57.54	-2.49	954.9
341	FINDHORN	GB_6604860	UK	57.60	-3.64	781.9
342	CONON	GB_6604880	UK	57.56	-4.54	961.8
343	TEVIOT	GB_6604920	UK	55.43	-2.76	323
344	OYKEL	GB_6604950	UK	57.96	-4.70	330.7
345	HALLADALE	GB_6604960	UK	58.48	-3.90	204.6
346	RIBBLE	GB_6605100	UK	53.78	-2.63	1145
347	MERSEY	GB_6605210	UK	53.44	-2.34	660
348	TYNE	GB_6605300	UK	54.95	-1.94	2175.6
349	SOUTH TYNE	GB_6605310	UK	54.94	-2.51	321.9
350	BLYTH	GB_6605320	UK	55.11	-1.62	269.4
351	COQUET	GB_6605330	UK	55.33	-1.63	569.8
352	LIDDEL WATER	GB_6605350	UK	55.07	-2.92	319
353	EDEN (N. ENGLAND)	GB_6605360	UK	54.91	-2.95	2286.5
354	LEVEN (TRIB. IRISH SEA)	GB_6605390	UK	54.27	-2.97	247
355	DERWENT (TRIB. NORTH SEA)	GB_6605420	UK	54.02	-0.88	1586
356	AIRE	GB_6605440	UK	53.91	-1.98	282.3

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
357	LUNE	GB_6605450	UK	54.08	-2.72	983
358	WHARFE	GB_6605550	UK	53.92	-1.36	758.9
359	URE	GB_6605560	UK	54.10	-1.46	914.6
360	OUSE	GB_6605570	UK	53.99	-1.13	3315
361	NIDD	GB_6605575	UK	53.97	-1.35	484.3
362	TEES	GB_6605590	UK	54.52	-1.60	818.4
363	TRENT (N. ENGLAND)	GB_6605600	UK	52.95	-1.08	7486
364	IDLE	GB_6605610	UK	53.40	-0.96	529
365	DERWENT (TRIB. TRENT)	GB_6605620	UK	52.93	-1.47	1054
366	DOVE (TRIB. TRENT)	GB_6605630	UK	52.86	-1.65	883.2
367	WEAKE	GB_6605640	UK	52.71	-1.09	413.8
368	SOAR	GB_6605650	UK	52.57	-1.20	183.9
369	ANKER	GB_6605660	UK	52.63	-1.61	368
370	SOW	GB_6605680	UK	52.84	-2.17	163
371	BROWNEY	GB_6605700	UK	54.74	-1.60	179
372	WEAR	GB_6605730	UK	54.73	-1.59	657.8
373	LITTLE OUSE	GB_6606250	UK	52.43	0.72	688.5
374	ISE BROOK	GB_6606300	UK	52.33	-0.68	194
375	BEDFORD OUSE	GB_6606400	UK	52.13	-0.46	1460
376	BEDFORD OUSE	GB_6606401	UK	52.16	-0.32	1660
377	GLEN	GB_6606560	UK	52.72	-0.36	341.9
378	WITHAM	GB_6606590	UK	53.00	-0.76	297.9
379	CAM	GB_6606700	UK	52.13	0.14	198
380	IVEL	GB_6606720	UK	52.14	-0.32	541.3
381	CHELMER	GB_6606790	UK	51.74	0.48	190.3
382	STOUR (E. ENGLAND)	GB_6606850	UK	51.97	0.94	578
383	BURE	GB_6606880	UK	52.82	1.25	164.7
384	WAVENEY	GB_6606900	UK	52.38	1.28	370
385	COLNE	GB_6606950	UK	51.90	0.85	238.2
386	DART	GB_6607100	UK	50.48	-3.76	247.6
387	CAMEL	GB_6607115	UK	50.48	-4.80	208.8
388	FOWEY	GB_6607120	UK	50.43	-4.68	169.1
389	TORRIDGE	GB_6607140	UK	50.95	-4.14	663
390	TAW	GB_6607150	UK	51.00	-3.98	826.2
391	EXE	GB_6607200	UK	50.80	-3.51	600.9
392	CULM	GB_6607210	UK	50.84	-3.39	226.1
393	AXE	GB_6607220	UK	50.75	-3.05	288.5
394	OTTER	GB_6607225	UK	50.69	-3.29	202.5
395	TONE (S. ENGLAND)	GB_6607230	UK	51.02	-3.13	202
396	PIDDLE	GB_6607250	UK	50.69	-2.12	183.1
397	BLACKWATER (S. ENGLAND)	GB_6607330	UK	51.38	-0.95	354.8
398	CHERWELL	GB_6607340	UK	51.86	-1.30	551.7
399	KENNET	GB_6607345	UK	51.43	-1.07	1033.4
400	LAMBOURN	GB_6607347	UK	51.41	-1.32	234.1
401	STOUR (S. ENGLAND)	GB_6607500	UK	50.76	-1.84	1073

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
402	TEST	GB_6607520	UK	50.97	-1.50	1040
403	ITCHEN	GB_6607550	UK	50.99	-1.34	360
404	THAMES	GB_6607651	UK	51.41	-0.31	9948
405	THAMES	GB_6607701	UK	51.64	-1.18	3445
406	NADDER	GB_6607705	UK	51.08	-1.87	220.6
407	AVON (S. ENGLAND)	GB_6607710	UK	51.17	-1.78	323.7
408	DARENT	GB_6607800	UK	51.42	0.23	191.4
409	LEE RIVER	GB_6607830	UK	51.76	0.01	1036
410	RODING	GB_6607840	UK	51.58	0.04	303.3
411	BEULT	GB_6607850	UK	51.20	0.52	277.1
412	GREAT STOUR	GB_6607950	UK	51.26	1.03	345
413	DEE (WALES)	GB_6608100	UK	52.97	-2.97	1019.3
414	ALYN	GB_6608110	UK	53.09	-2.99	227.1
415	DYFI	GB_6608170	UK	52.60	-3.85	471.3
416	YSTWYTH	GB_6608190	UK	52.38	-4.07	169.6
417	TEIFI	GB_6608200	UK	52.04	-4.56	893.6
418	TYWI	GB_6608210	UK	51.86	-4.20	1090.4
419	COTHI	GB_6608220	UK	51.88	-4.17	297.8
420	RHYMNEY	GB_6608240	UK	51.53	-3.12	178.7
421	WYE (WALES)	GB_6608500	UK	52.30	-3.50	174
422	WYE (WALES)	GB_6608501	UK	51.80	-2.68	4010
423	LUGG	GB_6608520	UK	52.28	-2.93	203.3
424	ELWY	GB_6608530	UK	53.24	-3.57	194
425	LEADON	GB_6609110	UK	51.91	-2.32	293
426	AVON (CENTRAL ENGLAND)	GB_6609400	UK	52.09	-1.94	2210
427	SEVERN (CENTRAL ENGLAND)	GB_6609500	UK	52.38	-2.32	4325
428	WORFE	GB_6609520	UK	52.56	-2.38	258
429	PERRY	GB_6609530	UK	52.77	-2.84	180.8
430	TANAT	GB_6609538	UK	52.80	-3.11	229
431	RODEN	GB_6609540	UK	52.72	-2.61	259
432	AVON (CENTRAL ENGLAND)	GB_6609560	UK	52.34	-1.51	347
433	REA BROOK	GB_6609570	UK	52.68	-2.79	178
434	LYGNA	NO_6731250	Norway	58.40	7.23	266
435	AUSTENA	NO_6731280	Norway	58.84	8.10	286
436	JONDALSELV	NO_6731320	Norway	59.70	9.55	150
437	NEIDEN	NO_6731330	Norway	59.37	8.53	2911
438	HOBOLELV	NO_6731350	Norway	59.55	10.87	297
439	GLAMA	NO_6731401	Norway	60.88	11.56	15426
440	GLAMA	NO_6731402	Norway	60.25	11.68	20300
441	ATNA	NO_6731410	Norway	61.85	10.22	465
442	LAKSELV (E)	NO_6731900	Norway	69.23	17.78	178
443	KOKEMAENJOKI	FI_6854101	Finland	61.34	22.11	26117
444	KOKEMAENJOKI	FI_6854102	Finland	61.31	23.75	8641
445	KOKEMAENJOKI	FI_6854103	Finland	61.50	23.76	7672
446	KOKEMAENJOKI	FI_6854104	Finland	61.85	23.91	6102

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
447	KOKEMAENJOKI	FI_6854105	Finland	61.14	22.69	2652
448	KOKEMAENJOKI	FI_6854107	Finland	62.28	24.04	546
449	EURAJOKI	FI_6854150	Finland	61.20	21.75	1229
450	LESTIJOKI	FI_6854301	Finland	63.99	23.76	1283
451	OULUJOKI	FI_6854593	Finland	64.40	29.85	403
452	SIMOJOKI	FI_6854620	Finland	65.69	25.09	3109
453	KYRONJOKI	FI_6854900	Finland	63.14	21.84	4833
454	LAPVAEAERTINJOKI	FI_6854950	Finland	62.24	21.58	976
455	VANTAA / VANDA A	FI_6855100	Finland	60.23	24.98	1680
456	KYMIJOKI	FI_6855200	Finland	60.70	26.82	36275
457	KYMIJOKI	FI_6855201	Finland	61.57	26.04	1421
458	KYMIJOKI	FI_6855250	Finland	62.24	25.89	17684
459	KYMIJOKI	FI_6855272	Finland	62.75	24.84	409
460	PORVOONJOKI	FI_6855300	Finland	60.47	25.61	1128
461	VIROJOKI	FI_6855320	Finland	60.62	27.63	328
462	VUOKSI	FI_6855401	Finland	62.77	30.13	20816
463	VUOKSI	FI_6855402	Finland	62.55	27.77	16270
464	VUOKSI	FI_6855409	Finland	63.10	31.03	596
465	VUOKSI	FI_6855412	Finland	61.84	28.30	788
466	AARE	CH_6935020	Switzerland	46.93	7.45	2945
467	RHINE RIVER	CH_6935053	Switzerland	47.56	7.80	34526
468	RHINE RIVER	CH_6935054	Switzerland	47.57	8.33	14718
469	RHINE RIVER	CH_6935055	Switzerland	47.68	8.63	11887
470	BIRSE	CH_6935060	Switzerland	47.28	7.38	183
471	SUZE	CH_6935065	Switzerland	47.20	7.17	150
472	ERGOLZ	CH_6935070	Switzerland	47.49	7.73	261
473	RHINE RIVER	CH_6935145	Switzerland	46.84	9.46	3229
474	AARE	CH_6935300	Switzerland	47.52	8.23	17601
475	AARE	CH_6935301	Switzerland	47.48	8.19	11726
476	AARE	CH_6935302	Switzerland	47.27	7.83	10119
477	EMME	CH_6935320	Switzerland	46.97	7.74	443
478	SIMME	CH_6935330	Switzerland	46.66	7.44	344
479	SENSE	CH_6935350	Switzerland	46.89	7.35	352
480	THUR (CH)	CH_6935400	Switzerland	47.60	8.68	1696
481	THUR (CH)	CH_6935401	Switzerland	47.53	9.17	1085
482	SITTER	CH_6935410	Switzerland	47.41	9.32	261
483	RHINE RIVER	CH_6935500	Switzerland	47.38	9.64	6119
484	LANDQUART	CH_6935540	Switzerland	46.97	9.61	616
485	TOESS	CH_6935560	Switzerland	47.52	8.65	342
486	PLESSUR	CH_6935600	Switzerland	46.86	9.51	263
487	RHONE	CH_6939200	Switzerland	46.35	6.89	5244
488	RHONE	CH_6939500	Switzerland	46.13	7.09	3752
489	MASSA	CH_6939510	Switzerland	46.39	8.01	195
490	LUETSCHINE	CH_6939540	Switzerland	46.66	7.87	379
491	WEISSE LUETSCHINE	CH_6939541	Switzerland	46.63	7.90	164

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
492	INN	CH_6943100	Switzerland	46.89	10.47	1945
493	TICINO	CH_6948100	Switzerland	46.19	9.01	1515
494	POSCHIAVINO	CH_6948110	Switzerland	46.30	10.08	169
495	MAGGIA (TRIB. PO)	CH_6948120	Switzerland	46.16	8.79	926
496	VENTA	LV_6973010	Latvia	56.97	21.98	8320
497	DAUGAVA	LV_6973300	Latvia	55.88	26.53	64500
498	ORIA	ES_1080	Spain	43.23	-2.03	765
499	URUMEA	ES_1105	Spain	43.24	-1.94	215
500	BIDASOA	ES_1106	Spain	43.30	-1.73	681
501	IBAIZABAL	ES_1163	Spain	43.21	-2.75	251
502	ASON	ES_1196	Spain	43.33	-3.43	485
503	PAS	ES_1215	Spain	43.30	-3.97	357
504	CARES	ES_1276	Spain	43.32	-4.68	455
505	PILOÑ'A	ES_1302	Spain	43.37	-5.18	486
506	NORA	ES_1343	Spain	43.42	-5.89	314
507	NARCEA	ES_1353	Spain	43.19	-6.54	531
508	PIGŕEÁ'A	ES_1358	Spain	43.35	-6.20	403
509	NARCEA	ES_1359	Spain	43.37	-6.15	1705
510	ALLER	ES_1365	Spain	43.17	-5.73	265
511	CAUDAL	ES_1369	Spain	43.28	-5.87	893
512	CUBIA	ES_1378	Spain	43.39	-6.07	210
513	ESVA O NARAVAL	ES_1395	Spain	43.50	-6.44	411
514	IBIAS	ES_1404	Spain	43.06	-6.87	294
515	EO	ES_1427	Spain	43.41	-7.14	712
516	ORO	ES_1433	Spain	43.56	-7.38	163
517	LANDRO	ES_1438	Spain	43.62	-7.59	198
518	MANDEO	ES_1464	Spain	43.25	-8.05	248
519	ALLONES	ES_1485	Spain	43.23	-8.89	438
520	FURELOS	ES_1542	Spain	42.86	-8.02	150
521	ULLA	ES_1544	Spain	42.85	-8.02	516
522	DEZA	ES_1552	Spain	42.78	-8.34	545
523	UMIA	ES_1564	Spain	42.60	-8.64	190
524	MIÁ'O	ES_1607	Spain	43.14	-7.60	999
525	PARGA	ES_1617	Spain	43.17	-7.79	301
526	LADRA	ES_1619	Spain	43.15	-7.69	843
527	TEA	ES_1645	Spain	42.18	-8.51	286
528	LOURO	ES_1647	Spain	42.07	-8.63	150
529	CUA	ES_1724	Spain	42.65	-6.73	482
530	CABRERA	ES_1734	Spain	42.42	-6.81	558
531	SIL	ES_1739	Spain	42.42	-6.96	4268
532	CABE	ES_1765	Spain	42.56	-7.48	353
533	LIMIA	ES_1805	Spain	42.02	-7.88	684
534	DUERO	ES_2002	Spain	41.80	-2.45	1500
535	DUERO	ES_2004	Spain	41.49	-3.01	5055
536	UCERO	ES_2005	Spain	41.58	-3.08	900

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
537	DURATON	ES_2012	Spain	41.30	-3.74	480
538	DUERO	ES_2013	Spain	41.66	-3.68	7356
539	DUERO	ES_2015	Spain	41.57	-4.67	12740
540	CEGA	ES_2016	Spain	41.17	-3.85	280
541	ODRA	ES_2018	Spain	42.25	-4.20	796
542	PISUERGA	ES_2029	Spain	42.12	-4.25	4227
543	ARLANZA	ES_2030	Spain	42.06	-3.51	1200
544	ARLANZA	ES_2031	Spain	42.08	-4.07	2413
545	ARLANZA	ES_2036	Spain	42.06	-4.24	5256
546	UCIEZA	ES_2041	Spain	42.32	-4.52	312
547	CARRION	ES_2042	Spain	42.05	-4.55	2222
548	PISUERGA	ES_2043	Spain	41.74	-4.64	14283
549	ESGUEVA	ES_2044	Spain	41.67	-4.72	997
550	ADAJA	ES_2046	Spain	40.65	-4.71	770
551	ERESMA	ES_2048	Spain	41.33	-4.62	2746
552	ERESMA	ES_2050	Spain	40.95	-4.15	252
553	MOROS	ES_2052	Spain	40.82	-4.28	252
554	DUERO	ES_2054	Spain	41.51	-4.93	36570
555	ADAJA	ES_2056	Spain	41.49	-4.77	5202
556	PIRON	ES_2057	Spain	41.11	-4.12	172
557	DUERO	ES_2062	Spain	41.52	-5.41	41808
558	CURUEÑA'O	ES_2068	Spain	42.90	-5.40	154
559	BERNESGA	ES_2070	Spain	42.81	-5.63	340
560	ESLA	ES_2074	Spain	41.98	-5.64	6783
561	OMAÑA'AS	ES_2076	Spain	42.70	-5.90	481
562	TORMES	ES_2085	Spain	40.36	-5.53	900
563	TORMES	ES_2087	Spain	40.96	-5.65	4010
564	ERIA	ES_2089	Spain	42.22	-6.25	280
565	ESLA	ES_2095	Spain	41.87	-5.76	14263
566	PISUERGA	ES_2097	Spain	41.66	-4.73	15638
567	TERA	ES_2099	Spain	41.94	-5.78	2350
568	CEA	ES_2104	Spain	42.66	-5.03	355
569	VALDERADUEY	ES_2105	Spain	42.22	-5.10	283
570	NEGRO	ES_2113	Spain	42.04	-6.27	391
571	ARLANZON	ES_2116	Spain	42.27	-3.89	1702
572	UBIerna	ES_2125	Spain	42.42	-3.68	281
573	DUERO	ES_2132	Spain	41.63	-4.36	12093
574	ORBIGO	ES_2145	Spain	42.04	-5.74	4959
575	VALDERADUEY	ES_2148	Spain	41.54	-5.70	3546
576	ALHANDIGA	ES_2149	Spain	40.71	-5.60	255
577	TORIO	ES_2150	Spain	42.84	-5.52	222
578	DURATON	ES_2161	Spain	41.44	-3.97	1125
579	DUERO	ES_2162	Spain	41.67	-2.39	1717
580	DUERO	ES_2163	Spain	41.50	-2.51	2959
581	ESLA	ES_2710	Spain	42.40	-5.55	3980

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
582	TAMEGA	ES_2818	Spain	41.85	-7.43	719
583	TAJO	ES_3001	Spain	40.60	-1.93	410
584	TAJO	ES_3005	Spain	40.70	-2.58	3253
585	TAJO	ES_3015	Spain	39.96	-4.81	33849
586	GALLO	ES_3030	Spain	40.83	-1.96	944
587	GUADIELA	ES_3041	Spain	40.51	-2.32	666
588	ESCABAS	ES_3045	Spain	40.44	-2.31	345
589	HENARES	ES_3060	Spain	40.95	-2.92	1036
590	HENARES	ES_3062	Spain	40.46	-3.42	4031
591	TAJUÑA	ES_3082	Spain	40.31	-3.19	2029
592	GUADARRAMA	ES_3100	Spain	40.63	-4.00	234
593	GUADARRAMA	ES_3102	Spain	39.99	-4.03	1353
594	AMBROZ	ES_3144	Spain	40.17	-6.10	386
595	JERTE	ES_3146	Spain	40.11	-5.94	313
596	JERTE	ES_3147	Spain	39.98	-6.27	631
597	SORBE	ES_3159	Spain	40.99	-3.20	439
598	TIETAR	ES_3161	Spain	40.15	-5.04	730
599	ERJAS	ES_3163	Spain	39.82	-6.98	887
600	ALMONTE	ES_3168	Spain	39.66	-5.96	787
601	MAYOR	ES_3172	Spain	40.17	-2.69	430
602	MANZANARES	ES_3177	Spain	40.33	-3.55	1240
603	COFIO	ES_3180	Spain	40.41	-4.32	629
604	ALAGON	ES_3182	Spain	40.50	-5.95	426
605	TIETAR	ES_3184	Spain	39.93	-5.90	4090
606	TRABAQUE	ES_3186	Spain	40.43	-2.33	361
607	GUADARRAMA	ES_3194	Spain	40.49	-3.94	366
608	PERALES	ES_3198	Spain	40.37	-4.14	261
609	PUSA	ES_3212	Spain	39.92	-4.58	423
610	GEBALO	ES_3213	Spain	39.77	-4.85	196
611	SANGUSIN	ES_3217	Spain	40.43	-5.89	177
612	TAMUJA	ES_3220	Spain	39.50	-6.11	457
613	IBOR	ES_3221	Spain	39.77	-5.52	266
614	GUALIJA	ES_3222	Spain	39.76	-5.42	184
615	ALBERCHE	ES_3231	Spain	40.41	-4.70	698
616	GUAZALETE	ES_3233	Spain	39.89	-3.89	256
617	ARRAGO	ES_3238	Spain	40.06	-6.63	469
618	AYUELA	ES_3246	Spain	39.34	-6.49	252
619	MAGASCA	ES_3250	Spain	39.46	-5.92	231
620	SANGRERAS	ES_3251	Spain	39.92	-4.70	223
621	DULCE	ES_3254	Spain	40.96	-2.73	263
622	ALBURREL	ES_3278	Spain	39.47	-7.21	169
623	TOZO	ES_3279	Spain	39.61	-5.91	260
624	ALAGON	ES_3940	Spain	39.98	-6.30	2672
625	GUADIANA	ES_4004	Spain	38.98	-2.89	856
626	GUADIANA	ES_4030	Spain	38.86	-7.02	48530

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
627	TIRTEAFUERA	ES_4108	Spain	38.89	-4.40	734
628	ALBUERA	ES_4165	Spain	38.88	-6.78	430
629	ARDILA	ES_4174	Spain	38.23	-6.88	1800
630	BULLAQUE	ES_4214	Spain	38.99	-4.29	2029
631	ZANCARA	ES_4224	Spain	39.36	-2.58	2020
632	ALJUCEN	ES_4257	Spain	39.07	-6.29	228
633	GUADIANA	ES_4904	Spain	39.20	-3.59	10319
634	GUADALQUIVIR	ES_5004	Spain	37.99	-3.80	16166
635	GUADAJOZ O ALMEDINILLA	ES_5016	Spain	37.82	-4.80	2420
636	FRAILES	ES_5042	Spain	37.26	-3.77	357
637	GENIL	ES_5048	Spain	37.56	-5.08	8028
638	GUADIAMAR	ES_5056	Spain	37.53	-6.19	240
639	GUADAIRA	ES_5057	Spain	37.35	-5.96	1318
640	GUADIAMAR	ES_5076	Spain	37.30	-6.26	960
641	CORBONES	ES_5125	Spain	37.57	-5.61	1377
642	GUADAHORTUNA	ES_5133	Spain	37.60	-3.12	403
643	OJAIEN	ES_5138	Spain	38.66	-4.02	228
644	FARDES	ES_5140	Spain	37.54	-3.10	1685
645	SEGURA	ES_7001	Spain	38.39	-2.21	1218
646	SEGURA	ES_7057	Spain	38.41	-2.01	2185
647	TURIA O GUADALAVIAR	ES_8018	Spain	39.84	-1.15	4052
648	JUCAR	ES_8032	Spain	40.07	-2.14	984
649	MARIMOTA	ES_8087	Spain	39.82	-2.34	187
650	CABRIEL	ES_8090	Spain	39.94	-1.71	829
651	JUCAR	ES_8091	Spain	39.85	-2.29	1793
652	OJOS DE MOYA	ES_8092	Spain	39.67	-1.56	720
653	EBRON	ES_8104	Spain	40.11	-1.29	241
654	JUCAR	ES_8126	Spain	40.22	-1.85	250
655	MIJARES	ES_8134	Spain	40.12	-0.60	1396
656	LEZUZA	ES_8137	Spain	38.90	-2.25	217
657	CABRIEL	ES_8139	Spain	39.74	-1.63	1255
658	EBRO	ES_9001	Spain	42.69	-2.95	5481
659	EBRO	ES_9002	Spain	42.18	-1.69	25194
660	EGA I	ES_9003	Spain	42.38	-1.95	1445
661	ARAGON	ES_9005	Spain	42.34	-1.65	5469
662	PIEDRA	ES_9008	Spain	41.20	-1.79	732
663	JILOCA	ES_9010	Spain	41.11	-1.42	2202
664	ESERA	ES_9013	Spain	42.21	0.35	893
665	GUADALOPE	ES_9015	Spain	41.06	-0.13	3476
666	GARONA	ES_9019	Spain	42.74	0.74	440
667	VALIRA	ES_9022	Spain	42.36	1.45	559
668	SEGRE	ES_9025	Spain	41.45	0.42	12782
669	GUATIZALEMA	ES_9032	Spain	41.91	-0.13	362
670	ALCANADRE	ES_9033	Spain	41.91	-0.12	765
671	NAJERILLA	ES_9038	Spain	42.50	-2.68	1090

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
672	CIDACOS	ES_9044	Spain	42.11	-2.33	223
673	ISABENA	ES_9047	Spain	42.20	0.40	426
674	NAJERILLA	ES_9048	Spain	42.26	-2.78	541
675	TIRON	ES_9050	Spain	42.54	-2.97	698
676	JILOCA	ES_9055	Spain	41.25	-1.58	2502
677	MESA	ES_9056	Spain	41.19	-1.88	537
678	JALON	ES_9058	Spain	41.20	-2.34	196
679	GALLEGO	ES_9059	Spain	42.27	-0.75	1901
680	VERAL	ES_9062	Spain	42.66	-0.78	161
681	ESCA	ES_9063	Spain	42.65	-1.01	506
682	SALAZAR	ES_9064	Spain	42.72	-1.16	396
683	IRATI	ES_9065	Spain	42.62	-1.29	1546
684	IRATI	ES_9066	Spain	42.95	-1.25	236
685	ULZAMA	ES_9067	Spain	42.89	-1.61	240
686	ARAQUIL	ES_9068	Spain	42.83	-1.79	782
687	ARGA	ES_9069	Spain	42.79	-1.79	1756
688	EGA I	ES_9071	Spain	42.67	-2.03	943
689	ONSELLA	ES_9073	Spain	42.55	-1.25	275
690	ZADORRA	ES_9074	Spain	42.68	-2.90	1357
691	AYUDA	ES_9075	Spain	42.68	-2.88	307
692	ERRO	ES_9079	Spain	42.77	-1.41	180
693	JALON	ES_9087	Spain	41.73	-1.17	9694
694	GALLEGO	ES_9089	Spain	41.67	-0.84	4008
695	ALCANADRE	ES_9091	Spain	42.09	-0.11	501
696	NELA	ES_9092	Spain	42.80	-3.40	1093
697	OCA	ES_9093	Spain	42.74	-3.41	1051
698	GUADALOPE	ES_9099	Spain	41.21	0.01	3845
699	HUERVA	ES_9105	Spain	41.43	-1.08	620
700	SEGRE	ES_9111	Spain	42.23	1.34	2384
701	EBRO	ES_9120	Spain	42.42	-2.20	12010
702	GALLEGO	ES_9123	Spain	42.41	-0.65	1391
703	HUERVA	ES_9124	Spain	41.30	-1.08	456
704	NOGUERA RIBAGORZANA	ES_9137	Spain	42.40	0.74	558
705	LINARES	ES_9139	Spain	42.08	-2.01	326
706	TIRON	ES_9158	Spain	42.39	-3.21	192
707	ARGA	ES_9159	Spain	42.84	-1.59	178
708	EBRO	ES_9163	Spain	41.19	0.57	82245
709	BAYAS	ES_9165	Spain	42.69	-2.93	318
710	JEREA	ES_9166	Spain	42.81	-3.36	290
711	QUEILES	ES_9174	Spain	41.88	-1.78	194
712	ALHAMA	ES_9185	Spain	42.08	-1.81	1120
713	ARBA DE BIEL	ES_9187	Spain	42.12	-0.95	336
714	OMECILLO	ES_9188	Spain	42.78	-3.04	350
715	ORONCILLO O GRILLERA	ES_9189	Spain	42.67	-2.98	217
716	FLUMEN	ES_9190	Spain	42.14	-0.35	160

ID	River (GRDC name)	Station				Area (km ²)
		Code	Country	Lat (°)	Lon (°)	
717	FLUMEN	ES_9191	Spain	41.98	-0.42	534
718	HUERVA	ES_9215	Spain	41.22	-1.21	315
719	NOGUERA PALLARESA	ES_9252	Spain	42.55	1.16	444
720	CIDACOS	ES_9253	Spain	42.22	-2.23	405
721	TRUEBA	ES_9254	Spain	42.93	-3.48	465
722	ARBA DE LUESIA	ES_9260	Spain	41.91	-1.28	2193
723	ESCA	ES_9268	Spain	42.86	-0.93	188
724	ASSINO	IT_0001	Italy	43.32	12.39	165
725	CAINA	IT_0003	Italy	43.02	12.26	230
726	CERFONE	IT_0004	Italy	43.49	12.17	284
727	CHIANI	IT_0005	Italy	42.76	12.10	457
728	MARROGGIA	IT_0009	Italy	42.81	12.76	258
729	NESTORE	IT_0011	Italy	42.92	12.34	725

REFERENCES

- Afshar, M.H., Yilmaz, M.T., 2017. The added utility of nonlinear methods compared to linear methods in rescaling soil moisture products. *Remote Sensing of Environment*, 196, 224-237.
- Ahlström, A., Smith, B., Lindström, J., Rummukainen, M., Uvo, C.B., 2013. GCM characteristics explain the majority of uncertainty in projected 21st century terrestrial ecosystem carbon balance. *Biogeosciences* 10, 1517–1528.
- Albergel, C., C. Rüdiger, T. Pellarin, J.-C. Calvet, N. Fritz, F. Froissard, D. Suquia, A. Petitpa, B. Piguet, and E. Martin. 2008. From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations. *Hydrology and Earth System Sciences* 12 (6): 1323–1337.
- Albergel, C., C. Rüdiger, D. Carrer, J.-C. Calvet, N. Fritz, V. Naeimi, Z. Batalis, and S. Hasenauer. 2009. An evaluation of ASCAT surface soil moisture products with in-situ observations in Southwestern France. *Hydrology and Earth System Sciences* 13 (2): 115–124.
- Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y., Wagner, W., 2012. Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. *Remote Sens. Environ.* 118, 215–226.
- Albergel, C., L. Brocca, W. Wagner, P. de Rosnay, and J.-C Calvet. 2013. Selection of Performance Metrics for Global Soil Moisture Products: The Case of ASCAT Product. In *Remote Sensing of Energy Fluxes and Soil Moisture Content*, edited by George P. Petropoulos, 427-443. Boca Raton, FL: CRC Press.
- Alvarez-Garreton, C., Ryu, D., Western, A.W., Crow, W.T., Robertson, D.E., 2013. Impact of observation error structure on satellite soil moisture assimilation into a rainfall–runoff model. In: *MODSIM, 20th International Congress on Modelling and Simulation*. Modelling and Simulation Society of Australia and New Zealand, pp. 3071–3077.
- 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: 2763–2774.
- Alvarez-Garreton, C., Ryu, D., Western, A. W., Su, C.-H., Crow, W. T., Robertson, D. E., and 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, 1659-1676.
- Alvarez-Garreton, C., Ryu, D., Western, A., Crow, W., Su, C., & Robertson, D., 2016. Dual assimilation of satellite soil moisture to improve stream flow prediction in data-scarce catchments. *Water Resour. Res.*, 52, 5357–5375.
- Álvarez-Mozos, J., M. Gonzalez-Audicana, and J. Casali. 2007. Evaluation of empirical and semiempirical backscattering models for surface soil moisture estimation. *Canadian Journal of Remote Sensing* 33:176–188.

- Al-Yaari, A. , Wigneron, J. , Kerr, Y. , De Jeu, R. , Rodriguez-Fernandez, N. , Van Der Schalie, R. , Al Bitar, A , Mialon, A. , Richaume, P. , Dolman, A. , 2016b. Testing regression equations to derive long-term global soil moisture datasets from passive microwave observations. *Remote Sens. Environ.* 180, 453–464 .
- Amri, R., Zribi, M., Chabaane, Z. L., Wagner, W., and Hasenauer, S, 2012. Analysis of C-band scatterometer moisture estimations derived over a semiarid region, *IEEE T. Geosci. Remote*, 50, 2630–2638.
- Arnold, J.G., Srinivasan, R., Muttiah, R.S., and Williams, J.R. 1998. Large area hydrologic modeling and assessment: Part I. Model development. *Journal of the American Water Resources Association*, Vol. 34, No. 1, pp. 73–89.
- Arya, L. M., J. C. Richter, and J. F. Paris. 1983. Estimating profile water storage from surface zone soil moisture measurements under bare field conditions. *Water Resources Research* 19 (2): 403-412.
- Attema, E.P.W., 1991. The active microwave instrument on-board the ERS-1 satellite. *Proc. IEEE*, vol. 79, no. 6, pp. 791–799.
- Attema, E.P.W., Ulaby, F.T., 1978. Vegetation modeled as a water cloud. *Radio Sci.* 13 (2), 357–364.
- Baghdadi, N., and M. Zribi. 2006. Evaluation of radar backscatter models IEM, Oh and Dubois using experimental observations. *International Journal of Remote Sensing* 27:3831–3852.
- Baghdadi, N., King, C., Bourguignon, A., Remond, A., 2002. Potential of ERS and Radarsat data for surface roughness monitoring over bare agricultural fields: application to catchments in Northern France. *Int. J. Remote Sens.* 23, 3427–3442.
- Baghdadi, N., Zribi, M., Loumagne, C., Ansart, P., Anguela, T.P., 2008. Analysis of TerraSAR-X data and their sensitivity to soil surface parameters over bare agricultural fields. *Remote Sens. Environ.* 112, 4370–4379.
- Baldwin, D., Manfreda, S., Keller, K., and Smithwick, E.A.H., 2017. Predicting root zone soil moisture with soil properties and satellite near-surface moisture data across the conterminous United States. *J. Hydrol.* 546 393–404.
- Barrett, B. W., E. Dwyer, and P. Whelan. 2009. Soil moisture retrieval from active spaceborne microwave observations: An evaluation of current techniques. *Remote Sensing* 1(3):210–242.
- Barrett, B. W., and Petropoulos, G. P., 2013. Satellite Remote Sensing of Surface Soil Moisture. In *Remote Sensing of Energy Fluxes and Soil Moisture Content*, edited by George P. Petropoulos, 85-120. Boca Raton, FL: CRC Press.
- Bauer-Marschallinger, B.; Freeman, V.; Cao, S.; Paulik, C.; Schauffer, S.; Stachl, T.; Modanesi, S.; Massari, C.; Ciabatta, L.; Brocca, L.; Wagner, W.; 2019. Toward Global Soil Moisture Monitoring With Sentinel-1: Harnessing Assets and Overcoming Obstacles. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 520-539.
- Berthet, L., Andréassian, V., Perrin, C., and Javelle, P., 2009. How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments. *Hydrol. Earth Syst. Sci.*, 13, 819-831.

- Bierkens, M.F.; Bell, V.A.; Burek, P.; Chaney, N.; Condon, L.E.; David, C.H.; de Roo, A.; Döll, P.; Drost, N.; Famiglietti, J.S.; et al. 2015. Hyper-resolution global hydrological modelling: What is next? Everywhere and locally relevant. *Hydrol. Process.*, 29, 310–320.
- Bindlish, R., Barros, A.P., 2001. Parameterization of vegetation backscatter in radar-based, soil moisture estimation. *Remote Sens. Environ.* 76 (1), 130–137.
- Bittelli, M., 2011. Measuring soil water content: A review. *HortTechnology* , 21, 293–300.
- Blaney, H.F., Criddle, W.D., 1950. Determining Water Requirements in Irrigated Areas from Climatological and Irrigation Data. In: USDA, SCS-TP-96.
- Blöschl, G., Sivapalan, M., 1995. Scale issues in hydrological modelling - a review. *Hydrol. Processes* 9, 251–290.
- Blümich, B., Casanova, F., Dabrowski, M., Danieli, E., Evertz, L., Haber, A., van Landeghem, M., Haber-Pohlmeier, S., Olaru, A., Perlo, J., Sucre, O., 2011. Small-scale instrumentation for nuclear magnetic resonance of porous media. *New J. Phys.* 13, 1–15.
- Bonan, G. B., 1998. The land surface climatology of the NCAR Land Surface Model coupled to the NCAR Community Climate Model. *J. Climate*, 11, 1307–1326.
- Bouttier, F., and Courtier, P., 2002. Data assimilation concepts and methods. March 1999. ECMWF, Meteorological Training Course Lecture Series, available at <https://www.ecmwf.int/en/elibrary/16928-data-assimilation-concepts-and-methods>.
- Bracaglia, M., Ferrazzoli, P., Guerriero, L., 1995. A fully polarimetric multiple scattering model for crops. *Remote Sens. Environ.* 54 (3), 170–179.
- Brocca, L.; Morbidelli, R.; Melone, F.; Moramarco, T., 2007. Soil moisture spatial variability in experimental areas of central Italy. *J. Hydrol.*, 333, 356–373.
- Brocca, L., Melone, F., and Moramarco, T., 2008. On the estimation of antecedent wetness conditions in rainfall–runoff modelling. *Hydrol. Process.*, 642, 629–642.
- Brocca, L., Melone, F., and Moramarco, T., 2009. Antecedent wetness conditions based on ERS scatterometer data in support to rainfall-runoff modeling. *J. Hydrol.*, 364(1–2), 73–86.
- Brocca, L., F. Melone, T. Moramarco, W. Wagner, V. Naeimi, Z. Bartalis, and S. Hasenauer. 2010a. Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrology and Earth System Sciences* 14 (10): 1881–1893.
- Brocca, L., F. Melone, T. Moramarco, W. Wagner, and S. Hasenauer. 2010b. ASCAT Soil Wetness Index validation through in-situ and modeled soil moisture data in central Italy. *Remote Sensing of Environment* 114 (11): 2745–2755.
- Brocca, L., S. Hasenauer, T. Lacava, F. Melone, T. Moramarco, W. Wagner, W. Dorigo, et al. 2011a. Soil moisture estimation through ASCAT and AMSR-E sensors: an inter-comparison and validation study across Europe. *Remote Sensing of Environment* 115 (12): 3390–3408.
- Brocca, L.; Melone, F.; Moramarco, T., 2011b. Distributed rainfall-runoff modelling for flood frequency estimation and flood forecasting. *Hydrol. Process.*, 25, 2801–2813.

- Brocca, L., T. Moramarco, F. Melone, W. Wagner, S. Hasenauer, and S. Hahn, 2012a. 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.
- Brocca, L., Ponziani, F., Moramarco, T., Melone, F., Berni, N., & Wagner, W. 2012b. Improving landslide forecasting using ASCAT-derived soil moisture data: a case study of the Torgiovannetto landslide in central Italy. *Remote Sens.*, 4, 1232–1244.
- Brocca, L., F. Melone, T. Moramarco, W. Wagner, and C. Albergel. 2013. Scaling and Filtering Approaches for the Use of Satellite Soil Moisture Observations. In *Remote Sensing of Energy Fluxes and Soil Moisture Content*, edited by George P. Petropoulos, 411-425. Boca Raton, FL: CRC Press.
- Brocca, L., W. T. Crow, L. Ciabatta, C. Massari, P. de Rosnay, M. Enenkel, S. Hahn, et al. 2017a. A review of the applications of ASCAT soil moisture products. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10 (5): 2285–2306.
- Brocca, L., L. Ciabatta, C. Massari, S. Camici, A. Tarpanelli, 2017b. Soil moisture for hydrological applications: open questions and new opportunities. *Water*, 9 (2), p. 140.
- Burgers, G., Jan van Leeuwen, P., and Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, 126 (6), 1719-1724.
- Calvet, J.C.; Noilhan, J. 2000. From near-surface to root-zone soil moisture using year-round data. *J. Hydrometeorol.*, 1, 393–411.
- Castillo V.M., Gomez-Plaza A., Martinez-Mena M., 2003. The role of antecedent soil water content in the runoff response of semiarid catchments: a simulation approach. *Journal of Hydrology*, 284, 114–130.
- Ceballos, A., K. Scipal, W. Wagner, and J. Martínez-Fernández. 2005. Validation of ERS scatterometer-derived soil moisture data in the central part of the Duero Basin, Spain. *Hydrological Processes* 19 (8): 1549–1566.
- Cenci, L., Laiolo, P., Gabellani, S., Campo, L., Silvestro, F., Delogu, F., Boni, G., and Rudari, R. 2016. Assimilation of H-SAF Soil Moisture Products for Flash Flood Early Warning Systems. Case Study: Mediterranean Catchments, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 9, 5634–5646.
- Cenci, L., Pulvirenti, L., Boni, G., Chini, M., Matgen, P., Gabellani, S., et al. 2017. An evaluation of the potential of Sentinel 1 for improving flash flood predictions via soil moisture–data assimilation. *Adv. Geosci.*, 44, 89–100.
- Chanzy, A., Tarussov, A., Judge, A., Bonn, F., 1996. Soil water content determination using a digital ground-penetrating radar. *Soil Sci. Soc. Am. J.* 60, 1318–1326.
- Chan, S. K., Bindlish, R., O’Neill, P. E., Njoku, E., Jackson, T., Colliander, A., Chen, F., Burgin, M., Dunbar, S., Piepmeier, J., et al., 2016. Assessment of the SMAP Passive Soil Moisture Product, *IEEE T. Geosci. Remote*, 54, 4994–5007.
- Chan, S., Bindlish, R., O’Neill, P., Jackson, T., Njoku, E., Dunbar, S., Chaubell, J., Piepmeier, J., Yueh, S., Entekhabi, D., et al., 2018. Development and assessment of the SMAP enhanced passive soil moisture product, *Remote Sens. Environ.*, 204, 931–941.

- Chen, K. S., S. K. Yen, and W. P. Huang. 1995. A simple model for retrieving bare soil moisture from radar-scattering coefficients. *Remote Sensing of Environment* 54:121–126.
- Chen, F., et al., 1996. Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. *J. Geophys. Res.*, 101 (D3), 7251–7268.
- Chen, F., W. T. Crow, P. J. Starks, and D. N. Moriasi, 2011. Improving hydrologic predictions of a catchment model via assimilation of surface soil moisture, *Adv. Water Resour.*, 34, 526-536.
- Chen, F., W. T. Crow, and D. Ryu, 2014. Dual forcing and state correction via soil moisture assimilation for improved rainfall–runoff modelling. *J. Hydrometeorol.*, 15(5), 1832–1848.
- Chen, F., Crow, W.T., Colliander, A., Cosh, M.H., Jackson, T., Bindlish, R., Reichle, R., Chan, S.K., Bosch, D., Starks, P., Goodrich, D., Seyfried, M., 2017. Application of triple collocation in ground-based validation of soil moisture active/passive (SMAP) level 2 data products. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 10 (2), 489–502.
- Cho, E., Choi, M., Wagner, W., 2015. An assessment of remotely sensed surface and root zone soil moisture through active and passive sensors in northeast Asia. *Remote Sens. Environ.*, 160, 166–179.
- Choudhury, B. , Schugge, T. , Mo, T. , 1982. A parameterization of effective soil temperature for microwave emission. *J. Geophys. Res. Oceans* 87, 1301–1304 .
- Ciais, P., et al., 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* 437, 529–533.
- Clemente, R. S., R. Dejong, H. N. Hayhoe, W. D. Reynolds, and M. Hares, 1994. Testing and comparison of three unsaturated soil-water flow models, *Agric. Water Manage.*, 25, 135 – 152.
- Colliander, A., Jackson, T., Bindlish, R., Chan, S.K., Kim, S., Cosh, M.H., Dunbar, R., et al., 2017. Validation of SMAP surface soil moisture products with core validation sites. *Remote Sens. Environ.* 191, 215–231.
- Crow, W.T., van Loon, E., 2006. Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture. *J. Hydrometeorol.* 7, 421–432.
- Crow, W. T., and M. J. van den Berg. 2010. An improved approach for estimating observation and model error parameters in soil moisture data assimilation. *Water Resources Research* 46 (12), W12519.
- Crow, W. T., A. A. Berg, M. H. Cosh, A. Loew, B. P. Mohanty, R. Panciera, P. de Rosnay et al., 2012. Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products, *Rev. Geophys.*, 50(2), RG2002.
- Das, N. N., and B. P. Mohanty. 2006. Root zone soil moisture assessment using passive microwave remote sensing and vadose zone modeling. *Vadose Zone Journal* 5: 296–307.

- Das, N. N., D. Entekhabi, E. G. Njoku, J. C. Shi, J. T. Johnson, and A. Colliander, 2014. Tests of the SMAP combined radar and radiometer algorithm using airborne field campaign observations and simulated data. *IEEE Trans. Geosci. Rem. Sens.*, vol. 52, pp. 2018–2028.
- Dee, D.P., and da Silva, A.M., 1998. Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, 124, 269-295.
- Dee, D.P., 2005. Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131, 3323-3343.
- de Jeu, R.A., Holmes, T.R., Parinussa, R.M., Owe, M., 2014. A spatially coherent global soil moisture product with improved temporal resolution. *J. Hydrol.* 516, 284–296.
- 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 spaceborne microwave radiometers and scatterometers. *Surveys in Geophysics*, 29, 399–420.
- De Lannoy, G. J., Houser, P. R., Pauwels, V., and Verhoest, N. E., 2006. Assessment of model uncertainty for soil moisture through ensemble verification, *J. Geophys. Res.-Atmos.*, 111, D10101.
- De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Pauwels, V. R. N., and Verhoest, N. E. C. (2007) Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. *Water Resources Research*, Vol. 43, W09410.
- de Rosnay, P., M. Drusch, D. Vasiljevic, G. Balsamo, C. Albergel, and L. Isaksen, 2013. A simplified extended Kalman filter for the global operational soil moisture analysis at ECMWF. *Quart. J. Roy. Meteorol. Soc.*, vol. 139, no. 674, pp. 1199–1213.
- de Rosnay, P.; Balsamo, G.; Albergel, C.; Muñoz-Sabater, J.; Isaksen, L.; 2014. Initialisation of land surface variables for Numerical Weather Prediction. *Surv. Geophys.*, 35, 607–621.
- de Wit, A. J. W., and van Diepen, C. A., 2007. Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, 146(1–2), 38–56
- Dharssi, I., K. Bovis, B. Macpherson, and C. Jones. 2011. Operational assimilation of ASCAT surface soil wetness at the Met Office. *Hydrology and Earth System Sciences* 15 (8): 2729–2746.
- Dickinson, R.E., K.W. Oleson, G. Bonan, F. Hoffman, P. Thornton, M. Vertenstein, Z.-L. Yang, and X. Zeng, 2006. The Community Land Model and its climate statistics as a component of the Community Climate System Model, *J. Climate*, 19, 2302-2324.
- Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki and N. Hanasaki, 2006. The Second Global Soil Wetness Project (GSWP-2): Multi-model analysis and implications for our perception of the land surface. *Bull. Amer. Meteor. Soc.*, 87, 1381-1397.
- Dirmeyer, P.A., Halder, S., 2016. Application of the land–atmosphere coupling paradigm to the operational coupled forecast system, version 2 (CFSv2). *J. Hydrometeorol.* 18, 85–108.

- Dobriyal, P., A. Qureshi, R. Badola, and S. A. Hussain. 2012. A review of the methods available for estimating soil moisture and its implications for water resource management. *Journal of Hydrology* 458–459: 110–117.
- Dong, J., and Crow, W. T., 2017. An improved triple collocation analysis algorithm for decomposing autocorrelated and white soil moisture retrieval errors. *Journal of Geophysical Research: Atmospheres*, 122, 13,081–13,094.
- Doolittle, J.A., Collins, M.E., 1995. Use of soil information to determine application of ground-penetrating radar. *J. Appl. Geophys.* 33, 101–108.
- 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, 2605–2616.
- Dorigo, W., W. Wagner, R. Hohensinn, S. Hahn, C. Paulik, A. Xaver, A. Gruber, et al. 2011. The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences* 15 (5): 1675–1698.
- Dorigo, W., A. Xaver, M. Vreugdenhil, A. Gruber, A. Hegyiová, A. Sanchis-Dufau, D. Zamojski, C. Cordes, W. Wagner, and M. Drusch. 2013. Global automated quality control of in-situ soil moisture data from the International Soil Moisture Network. *Vadose Zone Journal* 12 (3).
- Dorigo, W.A., Gruber, A., De Jeu, R.A.M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R.M., Kidd, R., 2015. Evaluation of the ESA CCI Soil Moisture product using ground-based observations. *Remote Sens. Environ.* 162, 380–395.
- Dorigo, W.A., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., et al., 2017. ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sensing of Environment*, 203, 185–215.
- Doubkova, M., A. I. J. M. van Dijk, D. Sabel, W. Wagner, and G. Bloschl. 2012. Evaluation of the predicted error of the soil moisture retrieval from C-band SAR by comparison against modelled soil moisture estimates over Australia. *Remote Sensing of Environment* 120 (2): 188–196.
- Draper, C., J.-F. Mahfouf, J.-C. Calvet, E. Martin, and W. Wagner. 2011. Assimilation of ASCAT near-surface soil moisture into the SIM hydrological model over France. *Hydrology and Earth System Sciences* 15 (12): 3829–3841.
- Draper, C., R. Reichle, G. De Lannoy, and Q. Liu. 2012. Assimilation of passive and active microwave soil moisture retrievals. *Geophysical Research Letters* 39 (4), L04401.
- Draper, C., R. Reichle, R. de Jeu, V. Naeimi, R. Parinussa, and W. Wagner. 2013. Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. *Remote Sensing of Environment* 137: 288–298.
- Drusch, M., Wood, E. F., and Gao, H., 2005. Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture. *Geophysical Research Letters*, 32, L15403.

- Drusch, M., and Viterbo, P., 2007. Assimilation of screen-level variables in ECMWF's integrated forecast system: a study on the impact on the forecast quality and analyzed soil moisture. *Mon. Weather Rev.* 135, 300–314.
- Dubois, P. C., J. Van Zyl, and T. Engman. 1995. Measuring soil moisture with imaging radars. *IEEE Transactions on Geoscience and Remote Sensing* 33(4):915–926.
- Engman, E. T. and N. S. Chauhan, 1995. Status of microwave soil moisture measurements with remote sensing. *Remote Sens. Environ.*, vol. 51, no. 1, pp. 189–198.
- Entekhabi, D., R. Reichle, R. Koster, and W. T. Crow. 2010a. Performance metrics for soil moisture retrievals and application requirements. *Journal of Hydrometeorology* 11: 832–840.
- Entekhabi, D., E. G. Njoku, P. E. O'Neill, K. H. Kellogg, W. T. Crow, W. N. Edelstein, J. K. Entin, et al. 2010b. The Soil Moisture Active Passive (SMAP) mission. *Proceedings of the IEEE* 98 (5): 704–716.
- Escorihuela, M.J., Chanzy, A., Wigneron, J.P., Kerr, Y.H., 2010. Effective soil moisture sampling depth of L-band radiometry: a case study. *Rem. Sens. Environ.* 114, 995–1001.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99 (C5), 10143-10162.
- Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, 53 (4), 343-367.
- Evensen, G., 2009. *Data Assimilation. The Ensemble Kalman Filter.* Springer.
- Fairbairn, D., Barbu, A. L., Napoly, A., Albergel, C., Mahfouf, J.-F., and Calvet, J.-C., 2017. The effect of satellite-derived surface soil moisture and leaf area index land data assimilation on streamflow simulations over France, *Hydrol. Earth Syst. Sci.*, 21, 2015-2033.
- Famiglietti, J.S., and Wood, E.F. 1994a. Application of Multiscale water and energy balance models on a tallgrass prairie. *Water Resources Research*, Vol. 30, No. 11, pp. 3061–3078.
- Famiglietti, J.S., and Wood, E.F. 1994b. Multiscale modeling of spatially variable water and energy balance processes. *Water Resources Research*, Vol. 30, No. 11, pp. 3079–3093.
- Famiglietti, J., J. Rudnicki, and M. Rodell, 1998. Variability in surface moisture content along a hillslope transect: Rattlesnake Hill, Texas, *J. Hydrol.*, 210, 259–281.
- Faridani, F., Farid, A., Ansari, H., and Manfreda, S., 2016. Estimation of the root-zone soil moisture using passive microwave remote sensing and SMAR model. *Journal of Irrigation and Drainage Engineering*, ASCE, Volume 143(1), pp. 1-9.
- Faridani, F., Farid, A., Ansari, H., and Manfreda, S., 2017. A modified version of the SMAR model for estimating root-zone soil moisture from time-series of surface soil moisture. *Water SA*, 43 (3), pp. 492-498.
- Fernandez-Moran R., A. Al-Yaari, A. Mialon, A. Mahmoodi, A. Al Bitar, G. De Lannoy, N. Rodriguez-Fernandez, E. Lopez-Baeza, Y. Kerr and J.-P. Wigneron, 2017.

SMOS-IC: An Alternative SMOS Soil Moisture and Vegetation Optical Depth Product. *Remote Sensing*, 9, 457.

Figa-Saldana, J., J. J. W. Wilson, E. Attema, R. Gelsthorpe, M. R. Drinkwater, and A. Stoffelen. 2002. The advanced scatterometer (ASCAT) on the meteorological operational (MetOp) platform: A follow on for European wind scatterometers. *Canadian Journal of Remote Sensing* 28 (3): 404–412.

Findell, K.L., Gentine, P., Lintner, B.R., Kerr, C., 2011. Probability of afternoon precipitation in eastern United States and Mexico enhanced by high evaporation. *Nat. Geosci.* 4, 434–439.

Flores, A. N., R. L. Bras, and D. Entekhabi. 2012. Hydrologic data assimilation with a hillslope-scale-resolving model and L band radar observations: Synthetic experiments with the ensemble Kalman filter. *Water Resources Research*, 48: W08509.

Ford, T. W., E. Harris, and S. M. Quiring. 2014. Estimating Root Zone Soil Moisture Using Near-Surface Observations from SMOS. *Hydrology and Earth System Sciences* 18 (1): 139-154.

Francois, C.; Quesney, A.; Otlé, C. 2003. Sequential assimilation of ERS-1 SAR data into a coupled land surface-hydrological model using an extended Kalman filter. *J. Hydrometeorol.*, 4, 473–487.

Fujii,H., Koike,T., Imaoka,K., 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*, Vol.29, No.1, pp.282-292.

Fung, A. K., Z. Li, and K. S. Chen. 1992. Backscattering from a randomly rough dielectric surface. *IEEE Transactions on Geoscience and Remote Sensing* 30(2):356–369.

Galantowicz, J. F., D. Entekhabi, and E. G. Njoku. 1999. Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness. *IEEE Transactions on Geoscience and Remote Sensing*, 37(4): 1860–1870.

González-Zamora, A., Sánchez, N., Martínez-Fernández, J., & Wagner, W. 2016. Root-zone plant available water estimation using the SMOS-derived soil water index. *Adv. Water Resour.*, 96, 339–353.

Goodrich, D.C.; Schmugge, T.J.; Jackson, T.J.; Unkrich, C.L.; Keefer, T.O.; Parry, R.; Bach, L.B.; Amer, S.A.; 1994. Runoff simulation sensitivity to remotely sensed initial soil water content. *Water Resour. Res.*, 30, 1393–1405.

Graham, L. P. and Bergström, S., 2000. Land surface modelling in hydrology and meteorology – lessons learned from the Baltic Basin, *Hydrol. Earth Syst. Sci.*, 4, 13-22.

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

Gruber, A., C.-H. Su, S. Zwieback, W. T. Crow, W. Dorigo, and W. Wagner. 2016a. Recent advances in (soil moisture) triple collocation analysis. *International Journal of Applied Earth Observation and Geoinformation* 45(B): 200–211.

- Gruber, A., C. H. Su, W. T. Crow, S. Zwieback, W. A. Dorigo, and W. Wagner. 2016b. Estimating error cross-correlations in soil moisture data sets using extended collocation analysis. *J. Geophys. Res. Atmospheres*, vol. 121, no. 3, pp. 1208–1219.
- Gruber, A.; Dorigo, W.A.; Crow, W.; Wagner, W.; 2017. Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals. *IEEE Trans. Geosci. Remote Sens.*, 55, 6780–6792.
- Gruber, A.; Crow, W.; Dorigo, W.; 2018. Assimilation of spatially sparse in situ soil moisture networks into a continuous model domain. *Water Resour. Res.*, 54, 1353–1367.
- Guerriero, L., Ferrazzoli, P., Rahmoune, R., 2012, March. A synergistic view of L-band active and passive remote sensing of vegetated soil. In: *Microwave Radiometry and Remote Sensing of the Environment (MicroRad)*, 2012 12th Specialist Meeting on, pp. 1–3. IEEE.
- Guilod, B.P., Orlowsky, B., Miralles, D.G., Teuling, A.J., Seneviratne, S.I., 2015. Reconciling spatial and temporal soil moisture effects on afternoon rainfall. *Nat. Commun.* 6.
- Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, 377, 80–91.
- Hallikainen, M.T. , Ulaby, F.T. , Dobson, M.C. , El-Rayes, M.A. , Wu, L.-K. , 1985. Microwave dielectric behavior of wet soil-part 1: Empirical models and experimental observations. *IEEE Trans. Geosci. Remote Sens.* 25–34 .
- Han, X., X. Li, H. J. H. Franssen, H. Vereecken, and C. Montzka. 2012. Spatial horizontal correlation characteristics in the land data assimilation of soil moisture. *Hydrology and Earth System Sciences*, 16(5), 1349–1363.
- Han, E., W. T. Crow, T. Holmes, and J. Bolten, 2014. Benchmarking a soil moisture data assimilation system for agricultural drought monitoring. *J. Hydrometeorol.*, vol. 15, no. 3, pp. 1117–1134.
- Haylock, M., N. Hofstra, A. K. Tank, E. Klok, P. Jones, and M. New, 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, *J. Geophys. Res.*, 113, D20119.
- Hirschi, M., Seneviratne, S.I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O.B., Formayer, H., Orlowsky, B., Stepanek, P., 2011. Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nat. Geosci.* 4, 17–21.
- Hirschi, M., Mueller, B., Dorigo, W., Seneviratne, S.I., 2014. Using remotely sensed soil moisture for land–atmosphere coupling diagnostics: the role of surface vs. rootzone soil moisture variability. *Remote Sens. Environ.* 154, 246–252.
- Hofstra, N., M. Haylock, M. New, P. Jones, and C. Frei, 2008. Comparison of six methods for the interpolation of daily, European climate data, *J. Geophys. Res.*, 113, D21110.
- Hofstra, N., M. Haylock, M. New, and P. Jones, 2009. Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature, *J. Geophys. Res.*, 114, D21101.

- Holah, N., N. Baghdadi, M. Zribi, A. Bruand, and C. King. 2005. Potential of ASAR/ENVISAT for the characterization of soil surface parameters over bare agricultural fields. *Remote Sensing of Environment* 96(1):78–86.
- Holmes, T. , De Rosnay, P. , De Jeu, R. , Wigneron, R.P. , Kerr, Y. , Calvet, J.C. , Escorihuela, M. , Saleh, K. , Lemaître, F. , 2006. A new parameterization of the effective temperature for L band radiometry. *Geophys. Res. Lett.* 33 .
- 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(D04), 113.
- Hopmans, J. W., D. R. Nielsen, and K. L. Bristow, 2002. How useful are small-scale soil hydraulic property measurements for large-scale vadose zone modeling? in *Environmental Mechanics: Water, Mass and Energy Transfer in the Biosphere: The Philip Volume*, P. A.C. Raats, D. Smiles, and A. W. Warrick Eds. Washington, DC, USA: American Geophysical Union, pp. 247–258.
- Hornacek, M.; Wagner, W.; Sabel, D.; Truong, H.-L.; Snoeij, P.; Hahmann, T.; Diedrich, E.; Doubková, M.; 2012. Potential for high resolution systematic global surface soil moisture retrieval via change detection using Sentinel-1. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 5, 1303–1311.
- Houser, P.R., et al., 1998. Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*, 34 (12), 3405–3420.
- H-SAF (Satellite Application Facility on Support to Operational Hydrology and Water Management). 2014. Product User Manual (PUM) for H25/SM-OBS-4 METOP ASCAT Soil Moisture Time Series (version 0.9). http://hsaf.meteoam.it/documents/PUM/SAF_HSAF_CDOP_PUM_25_0_9.pdf
- Ide, K., Courtier, P., Ghil, M., and Lorenc, A.C., 1997. Unified notation for data assimilation: Operational, sequential and variational. *Journal of the Meteorological Society of Japan*, 75 (1B), 181-189.
- Imaoka, K.; Kachi, M.; Fujii, H.; Murakami, H.; Hori, M.; Ono, A.; Igarashi, T.; Nakagawa, K.; Oki, T.; Honda, Y.; et al. 2010. Global Change Observation Mission (GCOM) for monitoring carbon, water cycles, and climate change. *IEEE Proc.*, 98, 717–734.
- Jackson, T.J. 1980. Profile soil moisture from space measurements. *Journal of Irrigation and Drainage Engineering - ASCE* 106: 81–92.
- Jackson T.J., Schmugge J., Engman E.T.; 1996. Remote sensing applications to hydrology: soil moisture, *Hydrological Sciences Journal*, 41:4, 517-530.
- Jackson, T.J., Cosh, M.H., Bindlish, R., Starks, P.J., Bosch, D.D., Seyfried, M., Goodrich, D.C., Moran, M.S., Du, J., 2010. Validation of advanced microwave scanning radiometer soil moisture products. *IEEE Trans. Geosci. Remote Sens.* 48, 4256–4272.
- Jackson, T.J., R. Bindlish, M.H. Cosh, T.J. Zhao, P.J. Starks, D.D. Bosch, et al., 2012. Validation of Soil Moisture and Ocean Salinity (SMOS) soil moisture over watershed networks in the U.S. *IEEE Trans. Geosci. Remote Sens.* 50:1530–1543.

- Jackson, T.J., 1993. Measuring surface soil moisture using passive microwave remote sensing. *Hydrol. Process.* 7, 139–152.
- Jackson, T.J., T. Schmugge, and J. Wang. 1982. Passive microwave remote sensing of soil moisture under vegetation canopies. *Water Resour. Res.* 18, 1137–1142.
- Jackson, T.J., Le Vine, D.M. , Hsu, A.Y. , Oldak, A. , Starks, P.J. , Swift, C.T. , Isham, J.D. , Haken, M. , 1999. Soil moisture mapping at regional scales using microwave radiometry: The southern great plains hydrology experiment. *IEEE Trans. Geosci. Remote Sens.* , 37, 2136–2151 .
- Javelle, P., Fouchier, C., Arnaud, P., and Lavabre, J., 2010. Flash flood warning at ungauged locations using radar rainfall and antecedent soil moisture estimations. *J. Hydrol.*, 397, 267–274.
- Jazwinski, A. H., 1970. *Stochastic Processes and Filtering.* Mathematics in Science and Engineering. New York: Academic Press. p. 376.
- Jonard, F. et al., 2018. Ground-Based Soil Moisture Determination. In: Li X., Vereecken H. (eds) *Observation and Measurement. Ecohydrology.* Springer, Berlin, Heidelberg.
- Jones, A.S.; Vukicevic, T.; Vonder Haar, T.H. 2004. A microwave satellite observational operator for variational data assimilation of soil moisture. *J. Hydrometeorol.*, 5, 213–229.
- Kalman, R.E., 1960. A new approach to linear filtering and prediction. *ASME Transactions*, Vol. 82, Part D (Journal of Basic Engineering), 35-45.
- Karthikeyan, L., Pan, M., Wanders, N., Kumar, D.N., Wood, E.F., 2017a. Four decades of microwave satellite soil moisture observations: Part 1. A review of retrieval algorithms. *Adv. Water Resour.*, 109, 106–120.
- Karthikeyan, L., Pan, M., Wanders, N., Kumar, D.N., Wood, E.F., 2017b. Four decades of microwave satellite soil moisture observations: Part 2. Product validation and inter-satellite comparisons. *Adv. Water Resour.*, 109, 236–252.
- Kerr, Y. H., Waldteufel, P., Wigneron, J. P., Martinuzzi, J. M., 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, 1729–1735.
- Kerr, Y. H., 2007. Soil moisture from space: Where are we? *Hydrogeol. J.*, vol. 15, no. 1, pp. 117–120.
- Kerr, Y. H., P. Waldteufel, P. Richaume, J.-P. Wigneron, P. Ferrazzoli, A. Mahmoodi, A. Al Bitar, et al. 2012. The SMOS soil moisture retrieval algorithm. *IEEE Transactions on Geoscience and Remote Sensing* 50 (5): 1384–1403.
- Kerr, Y.H., Al-Yaari, A., Rodriguez-Fernandez, N., Parrens, M., Molero, B., Leroux, D., et al., 2016. Overview of SMOS performance in terms of global soil moisture monitoring after six years in operation. *Remote Sens. Environ.* 180, 40–63.
- Kim, S. B., M. Moghaddam, L. Tsang, M. Burgin, X. Xu, and E. G. Njoku. 2014. Models of L-band radar backscattering coefficients over the global terrain for soil moisture retrieval. *IEEE Trans. Geosci. Remote Sens.*, 52, 1381-1396.

- Kim, S. B., L. Tsang, J. T. Johnson, S. Huang, J. J. van Zyl, and E. G. Njoku. 2012. Soil moisture retrieval using time-series radar observations over bare surfaces. *IEEE Trans. Geosci. Remote Sens.*, 50, 1853-1863.
- Kim, S.; Liu, Y.Y.; Johnson, F.M.; Parinussa, R.M.; Sharma, A.; 2015. A global comparison of alternate AMSR2 soil moisture products: Why do they differ? *Remote Sens. Environ.*, 161, 43–62.
- Köhli, M., M. Schrön, M. Zreda, U. Schmidt, P. Dietrich, S. Zacharias, 2015. Footprint characteristics revised for field-scale soil moisture monitoring with cosmic-ray neutrons. *Water Resour. Res.* 51, 5772–5790.
- Koike, T., Y. Nakamura, I. Kaihotsu, G. Davva, N. Matsuura, K. Tamagawa and H. Fujii; 2004. Development of an Advanced Microwave Scanning Radiometer (AMSR-E) Algorithm of Soil Moisture and Vegetation Water Content. *Annual Journal of Hydraulic Engineering, JSCE*, Vol. 48(2), 217-222.
- Konings, A.G. , Piles, M. , Rötzer, K. , McColl, K.A. , Chan, S.K. , Entekhabi, D. , 2016. Vegetation optical depth and scattering albedo retrieval using time series of dual-polarized L-band radiometer observations. *Remote Sens. Environ.* 172, 178–189 .
- Koren, V., J. Schaake, K. Mitchell, Q. Y. Duan, F. Chen, and J. M. Baker, 1999. A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *J. Geophys. Res.*, 104, 19 569–19 585.
- Kornelsen, K. C., and P. Coulibaly. 2013. Advances in Soil Moisture Retrieval from Synthetic Aperture Radar and Hydrological Applications. *Journal of Hydrology* 476: 460–489.
- Kornelsen, K. C., and P. Coulibaly. 2014. Root-zone soil moisture estimation using data-driven methods. *Water Resources Research* 50: 2946-2962.
- Kornelsen, K. C., and P. Coulibaly. 2015. Reducing multiplicative bias of satellite soil moisture retrievals. *Remote Sens. Environ.*, 165, 109–122.
- Koster, R. D., and M. J. Suarez, 1996. Energy and water balance calculations in the Mosaic LSM. *NASA Tech. Memo.* 104606, Vol. 9, 76 pp.
- Koster, R. D., and P. C. D. Milly, 1997. The interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models. *J. Climate*, 10, 1578-1591.
- Koster, R. D., M. J. Suarez, A. Ducharne, M. Stiglitz, and P. Kumar, 2000. A catchment-based approach to modeling land surface processes in a GCM. Part 1: Model structure. *J. Geophys. Res.*, 105 (D20), 24 809–24 822.
- Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K., & Puma, M. J., 2009. On the nature of soil moisture in land surface models. *Journal of Climate*, 22, 4322–4335.
- Kumar, S. V., R. H. Reichle, R. D. Koster, W. T. Crow, and C. D. Peters-Lidard, 2009. Role of subsurface physics in the assimilation of surface soil moisture observations. *J. Hydrometeorol.*, 10(6), 1534–1547.
- Kumar, S. V., R. H. Reichle, K. W. Harrison, C. D. Peters-Lidard, S. Yatheendradas, and J. A. Santanello, 2012. A comparison of methods for a priori bias correction in soil moisture data assimilation. *Water Resour. Res.*, 48, W03515,

- Kustas, W. P. and M. Anderson, 2009. Advances in thermal infrared remote sensing for land surface modelling. *Agricultural. Forest Meteorol.*, vol. 149, pp. 2071–2081.
- Lahoz, W.A., Schneider, P., 2014. Data assimilation: making sense of earth observation. *Front. Environ. Sci.* 2.
- Laiolo, P., S. Gabellani, L. Campo, F. Silvestro, F. Delogu, R. Rudari, L. Pulvirenti, et al. (2016). Impact of different satellite soil moisture products on the predictions of a continuous distributed hydrological model. *International Journal of Applied Earth Observation and Geoinformation*, 48, 131–145.
- Leng, P.; Song, X.; Li, Z.-L., 2014. Bare surface soil moisture retrieval from the synergistic use of optical and thermal infrared data. *Int. J. Remote Sens.*, 3, 988–1003.
- Li, J., and S. Islam. 1999. On the estimation of soil moisture profile and surface fluxes partitioning from sequential assimilation of surface layer soil moisture. *Journal of Hydrology*, 220(1–2): 86–103.
- Li, Y.; Grimaldi, S.; Walker, J.P.; Pauwels, V.R.N.; 2016. Application of Remote Sensing Data to Constrain Operational Rainfall-Driven Flood Forecasting: A Review. *Remote Sens.*, 8, 456.
- Li, L., Gaiser, P.W., Gao, B.C., Bevilacqua, R.M., Jackson, T.J., Njoku, E.G., Rüdiger, C., Calvet, J.C., Bindlish, R., 2010. WindSat global soil moisture retrieval and validation. *IEEE Trans. Geosci. Remote Sens.* 48, 2224–2241.
- Li, B., Rodell, M., 2013. Spatial variability and its scale dependency of observed and modeled soil moisture over different climate regions. *Hydrol. Earth Syst. Sci.* 17(3), 1177–1188.
- Liang, X., Lettenmaier, D.P., Wood, E.F., and Burges, S.J. 1994. A simple hydrologically based model of land surface water and energy fluxes for GSMs. *Journal of Geophysical Research*, Vol. 99, No. D7, pp. 14 415 – 14 428.
- Liang, X., Wood, E.F., and Lettenmaier, D.P. 1996. Surface soil moisture parameterization of the VIC-2L model: Evaluation and modifications. *Global and Planetary Change*, Vol. 13, pp. 195–206.
- Liu, W., Baret, F., Gu, X., Tong, Q., Zheng, L., and Zhang, B. 2002. Relating soil surface moisture to reflectance. *Remote Sensing of Environment*, 81(2–3), 238–246.
- Liu, Y. Y., Parinussa, R. M., Dorigo, W. A., De Jeu, R. A.M., Wagner, W., Van Dijk, A. I. J. M., et al., 2011. Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrology and Earth System Sciences*, 15, 425–436.
- Liu, Y.Y., Dorigo, W.A., Parinussa, R.M., de Jeu, R.A.M. , Wagner, W., McCabe, M.F., Evans, J.P., van Dijk, A.I.J.M., 2012a. Trend-preserving blending of passive and active microwave soil moisture retrievals, *Remote Sensing of Environment*, 123, 280–297.
- Liu, Y., Weerts, A.H., Clark, M., Franssen, H.-J.H., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A.I.J.M., et al., 2012b. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*, 16(10), 3863–3887.

- Lievens, H.; Tomer, S.K.; Al Bitar, A.; De Lannoy, G.J.; Drusch, M.; Dumedah, G.; Franssen, H.J.; Kerr, Y.H.; Martens, B.; Pan, M.; et al. 2015. SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sens. Environ.*, 168, 146–162.
- Lievens, H., G. J. M. De Lannoy, A. Al Bitar, M. Drusch, G. Dumedah, H.-J. Hendricks Franssen, Y. H. Kerr, S. K. Tomer, B. Martens, O. Merlin, M. Pan, J. K. Roundy, H. Vereecken, and J. P. Walker, 2016. Assimilation of SMOS soil moisture and brightness temperature products into a land surface model, *Remote Sens. Environ.*, 180, 292-304.
- Lobell, D. B., and G. P. Asner. 2002. Moisture effects on soil reflectance. *Soil Science Society of America Journal*, 66(3), 722–727.
- Loew, A., Schlenz, F., 2011. A dynamic approach for evaluating coarse scale satellite soil moisture products. *Hydrol. Earth Syst. Sci.*, 15, 75–90.
- Loew, A., et al., 2017. Validation practices for satellite-based Earth observation data across communities. *Rev. Geophys.*, 55.
- Lorenc, A.C. 2003. Modelling of error covariances by 4d-var data assimilation. *Q. J. R. Meteorol. Soc.*, 129, 3167–3182.
- Loizu, J.; Massari, C.; Álvarez-Mozos, J.; Tarpanelli, A.; Brocca, L.; Casali, J., 2018. On the assimilation set-up of ASCAT soil moisture data for improving stream flow catchment simulation. *Adv. Water Resour.*, 111, 86–104.
- Loumagne, C.; Normand, M.; Riffard, M.; Weisse, A.; Quesney, A.; Hégarat-Masclé, S.L.; Alem, F. 2001. Integration of remote sensing data into hydrological models for reservoir management. *Hydrol. Sci. J.*, 46, 89–102.
- Lu, H., Koike, T., Fujii, H., Ohta, T., & Tamagawa, K.; 2009. Development of a physically-based soil moisture retrieval algorithm for spaceborne passive microwave radiometers and its application to AMSR-E. *Journal of the Remote Sensing Society of Japan*, 29.
- Lv, S., Zeng, Y., Wen, J., Zhao, H. and Su, B., 2018. Estimation of Penetration Depth from Soil Effective Temperature in Microwave Radiometry. *Remote Sensing*, 10, 519.
- Mahfouf, J.-F. 2010. Assimilation of satellite-derived soil moisture from ASCAT in a limited-area NWP model. *Quarterly Journal of the Royal Meteorological Society* 136 (648): 784-798.
- Manfreda, S., L. Brocca, T. Moramarco, F. Melone, and J. Sheffield. 2014. A physically based approach for the estimation of root-zone soil moisture from surface measurements. *Hydrology and Earth System Sciences* 18 (3): 1199–1212.
- Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne. 2002. Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. *Water Resources Research*, 38(12).
- Martens, B., D. Miralles, H. Lievens, D. Fernandez-Prieto, and Verhoest, N., 2015. Improving terrestrial evaporation estimates over continental Australia through assimilation of SMOS soil moisture. *Int. J. Appl. Earth Observation Geoinf.*, vol. 48, pp. 146–162.

- Massari, C.; Brocca, L.; Moramarco, T.; Tramblay, Y.; Didon Lescot, J.-F. 2014a. Potential of soil moisture observations in flood modelling: Estimating initial conditions and correcting rainfall. *Adv. Water Resour.*, 74, 44–53.
- Massari, C.; Brocca, L.; Barbeta, S.; Papathanasiou, C.; Mimikou, M.; Moramarco, T. 2014b. Using globally available soil moisture indicators for flood modelling in Mediterranean catchments. *Hydrol. Earth Syst. Sci.*, 18, 839–853.
- Massari, C.; Brocca, L.; Tarpanelli, A.; Moramarco, T.; 2015. Data assimilation of satellite soil moisture into rainfall-runoff modelling: a complex recipe? *Remote Sens.*, 7, 11403–11433.
- Massari, C.; Camici, S.; Ciabatta, L.; Brocca, L.; 2018. Exploiting satellite-based surface soil moisture for flood forecasting in the Mediterranean area: state update versus rainfall correction. *Remote Sens.*, 10, 292.
- Matgen, P.; Fenicia, F.; Heitz, S.; Plaza, D.; de Keyser, R.; Pauwels, V.R.; Wagner, W.; Savenije, H.; 2012. Can ASCAT-derived soil wetness indices reduce predictive uncertainty in well-gauged areas? A comparison with in situ observed soil moisture in an assimilation application. *Adv. Water Resour.*, 44, 49–65.
- Mathieu, R., M. Sbih, A. A. Viau, F. Anctil, L. E. Parent, and J. Boisvert. 2003. Relationships between Radarsat SAR data and surface moisture content of agricultural organic soils. *International Journal of Remote Sensing* 24(24):5265–5281.
- McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., Lucier, A., Houborg, R., Verhoest, N. E. C., Franz, T. E., Shi, J., Gao, H., and Wood, E. F., 2017. The future of Earth observation in hydrology. *Hydrol. Earth Syst. Sci.*, 21, 3879–3914.
- McColl, K. A., J. Vogelzang, A. G. Konings, D. Entekhabi, M. Piles, and A. Stoffelen, 2014. Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target, *Geophys. Res. Lett.*, 41, 6229–6236.
- McLaughlin, D., 1995. Recent developments in hydrologic data assimilation. *Reviews of Geophysics*, 33 (S2), 977–984.
- McLaughlin, D., 2002. An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering. *Advances in Water Resources*, 25 (8-12), 1275-1286.
- Mecklenburg, S., M. Drusch, L. Kaleschke, N. Rodriguez-Fernandez, N. Reul, Y. Kerr, et al. 2016. ESA's Soil Moisture and Ocean Salinity mission: From science to operational applications. *Remote Sens. Environ.* 180:3–18.
- Melone, F.; Corradini, C.; Singh, V.P., 2002. Lag prediction in ungauged basins: An investigation through actual data of the upper Tevere River valley. *Hydrol. Processes*, 16, 1085–1094.
- Milly, P.C.D., 2001. A minimalist probabilistic description of root zone soil water. *Water Resour. Res.* 37, 457–463.
- Miralles, D.G., Holmes, T.R.H., De Jeu, R.A.M., Gash, J.H., Meesters, A.G.C.A., Dolman, A.J., 2011. Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.* 15, 453–469.

- Miralles, D. G., M. V. den Berg, A. J. Teuling, and R. D. Jeu, 2012. Soil moisture-temperature coupling: A multiscale observational analysis. *Geophys. Res. Lett.*, vol. 39, no. 21.
- Miralles, D.G., Teuling, A.J., van Heerwaarden, C.C., Vila-Guerau de Arellano, J., 2014. Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nat. Geosci.* 7, 345–349.
- Mladenova, I.E., Jackson, T.J., Njoku, E., Bindlish, R., Chan, S., Cosh, M.H., Holmes, T.R.H., de Jeu, R.A.M., Jones, L., Kimball, J., Santi, E., 2014. Remote monitoring of soil moisture using passive microwave-based techniques—theoretical basis and overview of selected algorithms for AMSR-E. *Remote Sens. Environ.* 144, 197–213.
- Mo, T. , Choudhury, B. , Schmugge, T. , Wang, J. , Jackson, T. , 1982. A model for microwave emission from vegetation-covered fields. *J. Geophys. Res. Oceans* 87, 11229–11237 .
- Mohanty, B. P., and T. H. Skaggs, 2001. Spatio-temporal evolution and time stable characteristics of soil moisture within remote sensing footprints with varying soils, slopes, and vegetation, *Adv. Water Resour.*, 24, 1051-1067.
- Montaldo, N., J. D. Albertson, M. Mancini, and G. Kiely. 2001. Robust simulation of root zone soil moisture from assimilation of surface soil moisture. *Water Resources Research* 37 (12): 2889-2900.
- Montaldo, N., and J. D. Albertson, 2003. Temporal dynamics of soil moisture variability: 2. Implications for land surface models, *Water Resour. Res.*, 39(10), 1275.
- Montzka, C., Pauwels, R.N.V., Franssen, H.-J.H., Han, X., and Vereecken, H., 2012. Multivariate and multiscale data assimilation in terrestrial systems: a review. *Sensors*, 12 (12), 16291-16333.
- Montzka, C. 2013. Soil Moisture Remote Sensing and Data Assimilation. In *Remote Sensing of Land Surface Turbulent Fluxes & Soil Moisture*; Taylor & Francis Group, CRC Press: Boca Raton, FL, USA.
- Moradkhani, H., Sorooshian, S., Gupta, H.V., and Houser, P.R., 2005. Dual state-parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, 28 (2), 135-147.
- Moradkhani, H., 2008. Hydrologic remote sensing and land surface data assimilation. *Sensors*, 8 (5), 2986-3004.
- Moran, M.S., Hymer, D.C., Qi, J.G., Sano, E.E., 2000. Soil moisture evaluation using multi-temporal synthetic aperture radar (SAR) in semiarid rangeland. *Agric. Forest Meteorol.* 105 (1–3), 69–80.
- Moran, M. S., Peters-Lidard, C. D., Watts, J. M., & McElroy, S., 2004. Estimating soil moisture at the watershed scale with satellite-based radar and land surface models. *Canadian Journal of Remote Sensing*, 30(5), 805–826.
- Morbideilli, R., C. Corradini, and R. S. Govindaraju, 2006. A field-scale infiltration model accounting for spatial heterogeneity of rainfall and soil saturated hydraulic conductivity. *Hydrol. Process.*, vol. 20, no. 7, pp. 1465–1481.

- Morbidelli, R., C. Saltalippi, A. Flammini, E. Rossi, and C. Corradini. 2014. Soil water content vertical profiles under natural conditions: Matching of experiments and simulations by a conceptual model. *Hydrological Processes* 28 (17): 4732-4742.
- Murphy, A., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. *Monthly Weather Review* 116, 2417–2424.
- Naeimi, V., K. Scipal, Z. Bartalis, S. Hasenauer, and W. Wagner. 2009. An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations. *IEEE Transactions on Geoscience and Remote Sensing* 47 (7): 1999–2013.
- Naeimi, V., C. Paulik, A. Bartsch, W. Wagner, S. Member, R. Kidd, S.-e. Park, K. Elger, and J. Boike. 2012. ASCAT Surface State Flag (SSF): extracting information on surface freeze/thaw conditions from backscatter data using an empirical threshold-analysis algorithm. *IEEE Transactions on Geoscience and Remote Sensing* 50 (7): 2566-2582.
- Nearing, G. S., S. Yatheendradas, W. T. Crow, D. D. Bosch, M. H. Cosh, D. C. Goodrich, M. S. Seyfried, and P. J. Starks, 2017. Nonparametric triple collocation, *Water Resour. Res.*, 53, 5516–5530.
- Neusch, T., and M. Sties. 1999. Application of the Dubois-model using experimental synthetic aperture radar data for the determination of soil moisture and surface roughness. *ISPRS Journal of Photogrammetry and Remote Sensing* 54:273–278.
- Ni-Meister, W., 2008. Recent Advances On Soil Moisture Data Assimilation. *Physical Geography*, 29 (1), 19-37.
- Njoku, E.G.; Jackson, T.J.; Lakshmi, V.; Chan, T.K.; Nghiem, S.V.; 2003. Soil moisture retrieval from AMSR-E. *IEEE Trans. Geosci. Remote Sens.*, 41, 215–229.
- Njoku, E.G., and Chan, S.K.; 2006. Vegetation and surface roughness effects on AMSR-E land observations. *Remote Sensing of Environment*, 100(2), 190–199.
- Njoku, E.G.; Entekhabi, D.; 1996. Passive Microwave Remote Sensing of Soil Moisture. *Journal of Hydrology*, 184 (1–2), 101–129.
- Njoku, E.G. , Li, L. , 1999. Retrieval of land surface parameters using passive microwave measurements at 6-18 GHz. *IEEE Trans. Geosci. Remote Sens.* 37, 79–93 .
- Norbiato, D., Borga, M., Degli Esposti, S., Gaume, E., and Anquetin, S., 2008. Flash flood warning based on rainfall depth-duration thresholds and soil moisture conditions: An assessment for gauged and ungauged basins. *J. Hydrol.*, 362(3–4), 274–290.
- Ochsner, T. E., M. H. Cosh, R. H. Cuenca, W. Dorigo, C. Draper, Y. Hagimoto, Y. H. Kerr, E. G. Njoku, E. E. Small, and M. Zreda. 2013. State of the art in large-scale soil moisture monitoring. *Soil Science Society of America Journal* 77: 1888-1919.
- Oh, Y. 2004. Quantitative retrieval of soil moisture content and surface roughness from multi-polarized radar observations of bare soil surfaces. *IEEE Transactions on Geoscience and Remote Sensing* 42:596–601.
- Oh, Y., K. Sarabandi, and F. T. Ulaby. 1992. An empirical model and an inversion technique for radar scattering from bare soil surfaces. *IEEE Transactions on Geoscience and Remote Sensing* 30(2):370–381.

- Oh, Y., K. Sarabandi, and F. T. Ulaby. 2002. Semiempirical model of the ensemble-averaged differential Mueller matrix for microwave backscattering from bare soil surfaces. *IEEE Transactions on Geoscience and Remote Sensing* 40(6):1348–1355.
- Ottlé, C.; Vidal-Madjar, D.; 1994. Assimilation of soil moisture inferred from infrared remote sensing in a hydrological model over the Hapex-Mobilhy Region. *J. Hydrol.*, 158, 241–264.
- Owe, M., de Jeu, R., Walker, J., 2001. A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Trans. Geosci. Remote Sens.* 39, 1643–1654.
- Owe, M. , de Jeu, R. , Holmes, T. , 2008. Multisensor historical climatology of satellite derived global land surface moisture. *J. Geophys. Res.: Earth Surf.* 113, F01002.
- Pachepsky, Y., Guber, A., and Jacques, D., 2005. Temporal persistence in vertical distributions of soil moisture contents, *Soil Sci. Soc. Am. J.*, 69, 347–352.
- Pal, M., Maity, R., & Dey, S., 2016. Statistical Modelling of Vertical Soil Moisture Profile: Coupling of Memory and Forcing. *Water Resources Management*, 30(6), 1973-1986.
- Paloscia, S.; Pampaloni, P.; Pettinato, S.; Santi, E.; 2008. A comparison of algorithms for retrieving soil moisture from ENVISAT/ASAR images. *IEEE Transactions on Geoscience and Remote Sensing* 46(10):3274–3284.
- Paloscia, S.; Pettinato, S.; Santi, E.; Notarnicola, C.; Pasolli, L.; Reppucci, A.; 2013. Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation. *Remote Sens. Environ.*, 134, 234–248.
- Paloscia, S.; Santi, E.; 2013. Passive Microwave Remote Sensing Techniques for the Retrieval of Surface Soil Moisture from Space. In *Remote Sensing of Energy Fluxes and Soil Moisture Content*, edited by George P. Petropoulos, 303-322. Boca Raton, FL: CRC Press.
- Pan, M. , Sahoo, A.K. , Wood, E.F. , 2014. Improving soil moisture retrievals from a physically-based radiative transfer model. *Remote Sens. Environ.* 140, 130–140 .
- Panciera, R., Tanase, M.A., Lowell, K., Walker, J.P., 2014. Evaluation of IEM, Dubois, and Oh radar backscatter models using airborne L-band SAR. *IEEE Trans. Geosci. Remote Sens.* 52, 4966–4979 .
- Panciera, R., 2009. Effect of land surface heterogeneity on satellite near-surface soil moisture observations, PhD thesis, Univ. of Melbourne, Melbourne, Vic., Australia.
- Parinussa, R., Meesters, A., Liu, Y., Dorigo, W., Wagner, W., and de Jeu, R., 2011. Error estimates for near-real-time satellite soil moisture as derived from the Land Parameter Retrieval Model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8, 779–783.
- Parinussa, R.M., Holmes, T.R.H., De Jeu, R.A.M., 2012. Soil moisture retrievals from the windSat spaceborne polarimetric microwave radiometer. *IEEE Trans. Geosci. Remote Sens.* 50, 2683–2694.
- Parinussa, R.M., Holmes, T.R.H., Wanders, N., Dorigo, W.A., de Jeu, R.A.M., 2015. A preliminary study toward consistent soil moisture from AMSR2. *Journal of Hydrometeorol.*, 16, 932–947.

- Parrens, M. , Wigneron, J.-P. , Richaume, P. , Mialon, A. , Al Bitar, A. , Fernandez–Moran, R. , Al-Yaari, A. , Kerr, Y.H. , 2016. Global-scale surface roughness effects at L-band as estimated from SMOS observations. *Remote Sens. Environ.* 181, 122–136 .
- Parrish, D.F.; Derber, J.C. 1992. The national-meteorological-centers spectral statistical-interpolation analysis system. *Mon. Weather Rev.*, 120, 1747–1763.
- Pathe, C., W. Wagner, D. Sabel, M. Doubkova, and J. B. Basara. 2009. Using ENVISAT ASAR global mode data for surface soil moisture retrieval over Oklahoma, USA. *IEEE Transactions on Geoscience and Remote Sensing* 47 (2): 468–480.
- Paulik, C., W. Dorigo, W. Wagner, and R. Kidd. 2014. Validation of the ASCAT Soil Water Index using in-situ data from the International Soil Moisture Network. *International Journal of Applied Earth Observation and Geoinformation* 30: 1–8.
- Pauwels, V. R. N., R. Hoeben, N. E. C. Verhoest, F. P. De Troch, and P. A. Troch. 2002. Improvement of TOPLATS-based discharge predictions through assimilation of ERS-based remotely sensed soil moisture values. *Hydrological Processes*, 16(5): 995–1013.
- Pellarin, T., Calvet, J.-C., and Wagner, W., 2006. Evaluation of ERS scatterometer soil moisture products over a half-degree region in southwestern France, *Geophys. Res. Lett.*, 33, L17401.
- Penna, D., H. van Meerveld, A. Gobbi, M. Borga, and G. Fontana, 2011. The influence of soil moisture on threshold runoff generation processes in an alpine headwater catchment, *Hydrol. Earth Syst. Sci.*, 15(3), 689–702.
- Peters-Lidard, C.D., Zion, M.S., and Wood, E.F. 1997. A soil-vegetation-atmosphere transfer scheme for modeling spatially variable water and energy balance processes. *Journal of Geophysical Research*, Vol. 102, No. D4, pp. 4303–4324.
- Petropoulos, G.; Griffiths, H.; Dorigo, W.; Xaver, A.; Gruber, A. 2013. Surface Soil Moisture Estimation: Significance, Controls, and Conventional Measurement Techniques. In *Remote Sensing of Land Surface Turbulent Fluxes & Soil Moisture*; Taylor & Francis Group, CRC Press: Boca Raton, FL, USA.
- Petropoulos, G. P., G. Ireland, and B. Barrett. 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: 36-56.
- Pierdicca, N., Pulvirenti, L., Bignami, C., Ticconi, F., 2013. Monitoring soil moisture in an agricultural test site using SAR data: design and test of a pre-operational procedure. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 6, 1199–1210.
- Pierdicca, N.; Pulvirenti, L.; Pace, G.; 2014. A prototype software package to retrieve soil moisture from Sentinel-1 data by using a bayesian multitemporal algorithm. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 7, 153–166.
- Pierdicca, N., L. Pulvirenti, F. Fascetti, L. Crapolicchio, R. Crapolicchio, and J. Munoz-Sabater, 2015. Quadruple collocation analysis for soil moisture product assessment. *IEEE Geosci. Remote Sens. Lett.*, vol. 12, pp. 1595–1599.
- Pierdicca, N., F. Fascetti, L. Pulvirenti, L. Crapolicchio, and R. Crapolicchio. 2017. Error Characterization of Soil Moisture Satellite Products: Retrieving Error Cross-

- Correlation Through Extended Quadruple Collocation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, pp. 4522–4530.
- Pires, L., O. S. Bacchi, and K. Reichardt, 2005. Soil water retention curve determined by gamma ray beam attenuation, *Soil Tillage Res.*, 82(1), 89–97.
- Porporato, A.; Daly, E.; Rodriguez-Iturbe, I. 2004. Soil water balance and ecosystem response to climate change. *Amer. Nat.*, 164, 625–632.
- Qiu, J., Crow, W. T., Nearing, G. S., Mo, X., & Liu, S. 2014. The impact of vertical measurement depth on the information content of soil moisture times series data. *Geophys. Res. Lett.*, 41, 4997–5004.
- Rahimzadeh-Bajgiran, P.; Berg, A.; 2016. Soil Moisture Retrievals Using Optical/TIR Methods. In *Satellite Soil Moisture Retrievals: Techniques & Applications*; Petropoulos, G.P., Srivastava, P., Kerr, Y., Eds.; Elsevier: Amsterdam, The Netherlands, Volume 3, pp. 47–72.
- Raju, S. , Chanzy, A. , Wigneron, J.-P. , Calvet, J.-C. , Kerr, Y. , Laguerre, L. , 1995. Soil moisture and temperature profile effects on microwave emission at low frequencies. *Remote Sens. Environ.* 54, 85–97 .
- Refsgaard, J.C., and Storm, B. 1995. MIKE SHE. In *Computer Models of Watershed Hydrology*. Edited by V.P. Singh. Water Resources Publications, Highlands Ranch, Colo. pp. 809–846.
- Reggiani, P., M. Sivapalan, and S. M. Hassanizadeh, 2000. Conservation equations governing hillslope responses: Exploring the physical basis of water balance. *Water Resour. Res.*, 36, 1845 – 1863.
- Reichle, R.H.; McLaughlin, D.B.; Entekhabi, D. 2001. Variational data assimilation of microwave radiobrightness observations for land surface hydrology applications. *IEEE Trans. Geosci. Remote Sens.*, 39, 1708–1718.
- Reichle, R.H., Walker, J.P., Koster, R.D., and Houser, P. R., 2002a. Extended versus ensemble Kalman filtering for land data assimilation. *Journal of Hydrometeorology*, 3 (6), 728-740.
- Reichle, R.H., McLaughlin, D.B., and Entekhabi, D., 2002b. Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130 (1), 103-114.
- Reichle, R., and R. Koster. 2004. Bias reduction in short records of satellite soil moisture. *Geophysical Research Letters* 31 (19), L19501.
- Reichle, R., R. D. Koster, J. Dong, and A. A. Berg, 2004. Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation. *J. Hydrometeorol.*, 5, 430– 442.
- Reichle, R. H. and R. D. Koster, 2005. Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model. *Geophysical Research Letters*, 32, L02404.
- Reichle, R.H., 2008. Data assimilation methods in the Earth sciences. *Advances in Water Resources*, 31 (11), 1411-1418.

- Reichstein, M., Beer, C., 2008. Soil respiration across scales: The importance of a model–data integration framework for data interpretation. *J. Plant Nutr. Soil Sci.* 171, 344–354.
- Ridler, M.-E.; Madsen, H.; Stisen, S.; Bircher, S.; Fensholt, R.; 2014. Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark. *Water Resour. Res.*, 50, 8962–8981.
- Riley, W.J.; Shen, C. 2014. Characterizing coarse-resolution watershed soil moisture heterogeneity using fine-scale simulations. *Hydrol. Earth Syst. Sci.*, 18, 2463–2483.
- Robinson, D. A., C. S. Campbell, J. W. Hopmans, B. K. Hornbuckle, S. B. Jones, R. Knight, F. Ogden, J. Selker, and O. Wendroth. 2008. Soil moisture measurements for ecological and hydrological watershed scale observatories: a review. *Vadose Zone Journal* 7: 358–389.
- Robock, A., K. Y. Vinnikov, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai, 2000. The global soil moisture data bank, *Bull. Am. Meteorol. Soc.*, 81(6), 1281–1299.
- Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., et al., 2004. The global land data assimilation system. *Bulletin of the American Meteorological Society*, 85, 381–394.
- Rodríguez-Blanco, M. L., Taboada-Castro, M. M., and Taboada-Castro, M. T., 2012. Rainfall–runoff response and event-based runoff coefficients in a humid area (northwest Spain). *Hydrolog. Sci. J.*, 57, 1–15.
- Rodríguez-Fernández, N.J., Aires, F., Richaume, P., Kerr, Y.H., Prigent, C., Kolassa, J., Cabot, F., Jiménez, C., Mahmoodi, A., Drusch, M., 2015. Soil moisture retrieval using neural networks: application to SMOS. *IEEE Trans. Geosci. Remote Sens.* 53, 5991–6007.
- Rodríguez-Fernández, N., Kerr, Y., van der Schalie, R., Al-Yaari, A., Wigneron, J.-P., de Jeu, R., Richaume, P., Dutra, E., Mialon, A., and Drusch, M., 2016. Long Term Global Surface Soil Moisture Fields Using an SMOS-Trained Neural Network Applied to AMSR-E Data. *Remote Sens.*, 8, 959.
- Rodríguez-Iturbe, I., Porporato, A., 2004. *Ecohydrology of Water-controlled Ecosystems – Soil Moisture and Plant Dynamics*. Cambridge University Press, New York, USA.
- Romano, N., 2014. Soil moisture at local scale: Measurements and simulations. *Journal of Hydrology*, 516, 6-20.
- Rosenbaum, U.; Bogena, H.R.; Herbst, M.; Huisman, J.A.; Peterson, T.J.; Weuthen, A.; Western, A.; Vereecken, H. 2002. Seasonal and event dynamics of spatial soil moisture patterns at the small catchment scale. *Water Resour. Res.*, 48, W10544.
- Ryu, D., Crow, W. T., Zhan, X., and Jackson, T. J, 2009. Correcting Unintended Perturbation Biases in Hydrologic Data Assimilation, *J. Hydrometeorol.*, 10, 734–750.
- Sabater, J.M.; Jarlan, L.; Calvet, J.C.; Bouyssel, F.; de Rosnay, P. 2007. From near-surface to root-zone soil moisture using different assimilation techniques. *J. Hydrometeorol.*, 8, 194–206.

- Satalino, G., F. Mattia, M.W. J. Davidson, T. Le Toan, G. Pasquariello, and M. Borgeaud, 2002. On current limits of soil moisture retrieval from ERS-SAR data. *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2438–2447.
- Sayde, C., C. Gregory, M. Gil-Rodriguez, N. Tuffiaro, S. Tyler, N. van de Giesen, et al. 2010. Feasibility of soil moisture monitoring with heated fiber optics. *Water Resour. Res.* 46:W06201.
- Schoups, G., J.W. Hopmans, C. A. Young, J. A. Vrugt, and W.W. Wallender, 2005. Multi-criteria optimization of a regional spatially-distributed subsurface water flow model, *J. Hydrol.*, 311, 20 – 48.
- Schrott, L., Sass, O., 2008. Application of field geophysics in geomorphology: advances and limitations exemplified by case studies. *Geomorphology* 93 (1–2), 55–73.
- Schwen, A., Zimmermann, M., Bodner, G., 2014. Vertical variations of soil hydraulic properties within two soil profiles and its relevance for soil water simulations. *J. Hydrol.*, 516, 169–181.
- Scipal, K., M. Drusch, and W. Wagner. 2008a. Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system. *Advances in Water Resources* 31 (8): 1101–1112.
- Scipal, K., T. Holmes, R. de Jeu, V. Naeimi, and W. Wagner. 2008b. A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophysical Research Letters* 35 (24), L24403.
- Seneviratne, S.I., Lüthi, D., Litschi, M., Schär, C., 2006. Land-atmosphere coupling and climate change in Europe. *Nature* 443, 205–209.
- Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, 2010. Investigating soil moisture-climate interactions in a changing climate: A review, *Earth Sci. Rev.*, 99, 125–161.
- Seo, D-J; Saifuddin, M.M.; Lee, H. Conditional bias-penalized Kalman filter for improved estimation and prediction of extremes. *Stoch. Environ. Res. Risk Assess.* 2018, 32, 183–201.
- Shutko, A. M. 1982. Microwave radiometry of lands under natural and artificial moistening. *IEEE Trans. Geosci. Remote Sci.* 20, 18–26.
- Simunek, J., and van Genuchten, M. Th. 1999. Using the HYDRUS-1D and HYDRUS-2D codes for estimating unsaturated soil hydraulic and solute transport parameters. In *Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media*. Edited by M.T. van Genuchten, F.J. Leij, and L. Wu. University of California, Riverside, Calif. pp. 1523–1536.
- Simunek, J., Huang, K., and van Genuchten, M. Th. 1998. The HYDRUS code for simulating the one-dimensional movement of water, heat, and multiple solutes in variably saturated media. Version 6.0. US Salinity Laboratory, Agricultural Research Service, US Department of Agriculture, Riverside, Calif. 164 pp.
- Soil Conservation Service, 1972. *National Engineering Handbook, Supplement A, Section 4 “Hydrology”, Chapter 10 “Estimation of Direct Runoff from Storm Rainfall”*. US Department of Agriculture, Washington, D.C.

- Srinivasan, R., and Arnold, J.G. 1994. Integration of a basin-scale water quality model with GIS. *Water Resources Bulletin*, Vol. 30, No. 3, pp. 453–462.
- Srivastava, S. K., N. Yograjan, V. Jayaraman, P. P. Nageswara Rao, and M. G. Chandrasekhar. 1997. On the relationship between ERS-1 SAR/backscatter and surface/sub-surface soil moisture variations in vertisols. *Acta Astronautica* 40 (10): 693–699.
- Steele-Dunne, S.C., M.M. Rutten, D.M. Krzeminska, M. Hausner, S.W. Tyler, J. Selker, et al. 2010. Feasibility of soil moisture estimation using passive distributed temperature sensing. *Water Resour. Res.* 46:W03534.
- Stiles, J.M., Sarabandi, K., 2000. Electromagnetic scattering from grassland. Part I: A fully phase-coherent scattering model. *IEEE Trans. Geosci. Remote Sens.* 38 (1), 339–348.
- Stoffelen, A. 1998. Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research: Oceans* 103 (C4): 7755–7766.
- Stoner, E. R., and M. F. Baumgardner. 1981. Characteristic variations in reflectance of surface soils. *Soil Science Society of America Journal*, 45(6), 1161–1165.
- Su, C.H., Ryu, D., Crow, W.T., Western, A.W., 2014a. Beyond triple collocation: applications to soil moisture monitoring. *J. Geophys. Res. Atmos.* 119, 6419–6439.
- Su, C.H., Ryu, D., Crow, W.T., Western, A.W., 2014b. Stand-alone error characterisation of microwave satellite soil moisture using a Fourier method. *Remote Sens. Environ.* 154, 115–126.
- Sun, L., Seidou, O., Nistor, I., and Liu, K., 2016. Review of the Kalman-type hydrological data assimilation. *Hydrological Sciences Journal*, 61 (13), 2348-2366.
- Talagrand, O., 1997. Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan*, 75(1B), 191-209.
- Tayfur, G., Zucco, G., Brocca, L., Moramarco, T., 2014. Coupling soil moisture and precipitation observations for predicting hourly runoff at small catchment scale. *Journal of Hydrology*, 510, 363-371.
- Taylor, C.M., De Jeu, R.A.M., Guichard, F., Harris, P.P., Dorigo, W.A., 2012. Afternoon rain more likely over drier soils. *Nature* 489, 282–286.
- Teuling, A. J., Uijlenhoet, R., Hupet, F., van Loon, E. E., and Troch, P. A., 2006. Estimating spatial mean root-zone soil moisture from point-scale observations, *Hydrol. Earth Syst. Sci.*, 10, 755-767.
- Tian, Y.; Huffman, G.J.; Adler, R.F.; Tang, L.; Sapiano, M.; Maggioni, V.; Wu, H., 2013. Modeling errors in daily precipitation measurements: Additive or multiplicative? *Geophys. Res. Lett.*, 40, 2060–2065.
- Tobin, K. J., Torres, R., Crow, W. T., and Bennett, M. E., 2017. Multi-decadal analysis of root-zone soil moisture applying the exponential filter across CONUS, *Hydrol. Earth Syst. Sci.*, 21, 4403-4417.

- Todini, E., and Biondi, D., 2016. Calibration, Parameters Estimation, Uncertainty, Data Assimilation, Sensitivity Analysis and Validation. In: Handbook of Applied Hydrology, Second Edition, edited by V. P. Singh, McGraw-Hill Education.
- Todisco, F., Brocca, L., Termiti, L. F., & Wagner, W. 2015. Use of satellite and modeled soil moisture data for predicting event soil loss at plot scale. *Hydrol. Earth Syst. Sci.*, 19, 3845–3856.
- Topp, G.C., J.L. Davis, and A.P. Annan. 1980. Electromagnetic determination of soil water content: Measurements in coaxial transmission lines. *Water Resour. Res.* 16:574–582.
- Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., et al., 2012. GMES Sentinel-1 mission. *Remote Sens. Environ.* 120, 9–24.
- Ulaby, F.T., Dubois, P.C., and Van Zyl, J. 1996. Radar mapping of surface soil moisture. *Journal of Hydrology*, 184, 57–84.
- Ulaby, F.T., Sarabandi, K., McDonald, K., Whitt, M., Dobson, M.C., 1990. Michigan microwave canopy scattering model. *Int. J. Remote Sens.* 11 (7), 1223–1253.
- Vachaud, G., Passerat De Silans, A., Balabanis, P., and Vauclin, M., 1985. Temporal stability of spatially measured soil water probability density function, *Soil Sci. Soc. Am. J.*, 49, 822–828.
- van de Griend, A., Camillo, P.J., Gurney, R.J., 1985. Discrimination of soil physical parameters, thermal inertia and soil moisture from diurnal surface temperature fluctuations. *Water Resour. Res.* 21 (7), 997–1009.
- van de Griend, A. A. , Wigneron, J.-P. , 2004. On the measurement of microwave vegetation properties: Some guidelines for a protocol. *IEEE Trans. Geosci. Remote Sens.* 42, 2277–2289 .
- van Oevelen, P.J., Hoekman, D.H., 1999. Radar backscatter inversion techniques for estimation of surface soil moisture: EFEDA-Spain and HAPEX-Sahel case studies. *IEEE Trans. Geosci. Remote Sens.* 1999 (37), 113–123.
- Van Steenbergen, N. and Willems, P., 2013. Increasing River Flood Preparedness by Real-time Warning Based on Wetness State Conditions. *J. Hydrol.*, 489, 227–237.
- Verstraeten, W. W., F. Veroustraete, and J. Feyen, 2008. Assessment of evapotranspiration and soil moisture content across different scales of observation, *Sensors*, 8, 70–117.
- Vereecken, H., Huisman, J.A., Bogaen, H., Vanderborght, J., Vrugt, J.A., Hopmans, J.W., 2008. On the value of soil moisture measurements in vadose zone hydrology: a review. *Water Resour. Res.* 44, W00D06.
- Wagner, W., G. Lemoine, and H. Rott. 1999. A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sensing of Environment* 70 (2): 191–207.
- Wagner, W., G. Blöschl, P. Pampaloni, J. C. Calvet, B. Bizzarri, J. P. Wigneron, and Y. Kerr; 2007. Operational readiness of microwave remote sensing of soil moisture for hydrologic applications. *Nord. Hydrol.*, 8, 1 – 20.

- Wagner, W., Pathe, C., Doubkova, M., Sabel, D., Bartsch, A., Hasenauer, S., Blöschl, G., Scipal, K., Martínez-Fernández, J., & Löw, A., 2008. Temporal stability of soil moisture and radar backscatter observed by the Advanced Synthetic Aperture Radar (ASAR). *Sensors*, 8, 1174–1197.
- Wagner, W., Dorigo, W., de Jeu, R., Fernandez-Prieto, D., Benveniste, J., Haas, E., et al., 2012. Fusion of active and passive microwave observations to create an Essential Climate Variable data record on soil moisture. XXII ISPRS Congress. Melbourne, Australia.
- Wagner, W., S. Hahn, R. Kidd, T. Melzer, Z. Bartalis, S. Hasenauer, J. Figa-Saldana, et al. 2013a. The ASCAT soil moisture product: a review of its specifications, validation results, and emerging applications. *Meteorologische Zeitschrift* 22 (1): 5–33.
- Wagner, W. et al., 2013b. Operations, challenges, and prospects of satellite-based surface soil moisture data services. In *Remote Sensing of Energy Fluxes and Soil Moisture Content*, edited by George P. Petropoulos, 463-488. Boca Raton, FL: CRC Press.
- Walker, J. P., G. R. Willgoose, and J. D. Kalma. 2002. Three-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: Simplified Kalman filter covariance forecasting and field application. *Water Resources Research*, 38(12): 1301.
- Walker, J. P., and Houser, P. R., 2005. Hydrologic data assimilation. In: *Advances in Water Science Methodologies*, edited by A. Aswathanarayana, 25-48, Taylor & Francis.
- Waller, J. A., García-Pintado, J., Mason, D. C., Dance, S. L., and Nichols, N. K., 2018. Technical note: Assessment of observation quality for data assimilation in flood models, *Hydrol. Earth Syst. Sci.*, 22, 3983-3992.
- Wanders, N., D. Karssenbergh, M. Bierkens, R. Parinussa, R. de Jeu, J. van Dam, and S. de Jong. 2012. Observation uncertainty of satellite soil moisture products determined with physically-based modelling. *Remote Sensing of Environment* 127: 341–356.
- Wanders, N.; Karssenbergh, 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, 2343–2357.
- Wang, J.R., Choudhury, B. , 1981. Remote sensing of soil moisture content, over bare field at 1.4 GHz frequency. *J. Geophys. Res. Oceans* 86, 5277–5282 .
- Wang, J.R. , Schmugge, T.J. , 1980. An empirical model for the complex dielectric permittivity of soils as a function of water content. *IEEE Trans. Geosci. Remote Sens.* 288–295 .
- Western, A.W., Grayson, R.B., 1998. The Tarrawarra data set: soil moisture patterns, soil characteristics and hydrological flux measurements. *Water Resources Research* 34 (10), 2765–2768.
- Western, A.W., Blöschl, G., 1999. On the spatial scaling of soil moisture. *Journal of Hydrology* 217, 203–224.
- Western, A. W., R. B. Grayson, G. Blöschl, G. R. Willgoose, and T. A. McMahon, 1999. Observed spatial organization of soil moisture and its relation to terrain indices, *Water Resour. Res.*, 35(3), 797–810.

- Western, A. W., R. B. Grayson, and G. Blöschl, 2002. Scaling of soil moisture: a hydrologic perspective. *Annual Review of Earth and Planetary Sciences*, 30, 149–180.
- Western, A. W., R. B. Grayson, G. Blöschl, and D. J. Wilson, 2003. Spatial variability of soil moisture and its implications for scaling, in: *Scaling Methods in Soil Physics*, edited by Y. Pachepsky, D. E. Radcliffe, and H. M. Selim, pp. 119–142, CRC Press, Boca Raton, Fla.
- Wigneron, J.-P. , Calvet, J.-C. , Pellarin, T. , Van de Griend, A. , Berger, M. , Ferrazzoli, P. , 2003. Retrieving near-surface soil moisture from microwave radiometric observations: current status and future plans. *Remote Sens. Environ.* 85, 489–506 .
- Wigneron, J.-P. , Chanzy, A. , De Rosnay, P. , Rudiger, C. , Calvet, J.-C. , 2008. Estimating the effective soil temperature at L-band as a function of soil properties. *IEEE Trans. Geosci. Remote Sens.* 46, 797–807 .
- Wigneron, J.-P. , Kerr, Y. , Waldteufel, P. , Saleh, K. , Escorihuela, M.-J. , Richaume, P. , Ferrazzoli, P. , De Rosnay, P. , Gurney, R. , Calvet, J.-C. , 2007. L-band microwave emission of the biosphere (L-MEB) model: Description and calibration against experimental data sets over crop fields. *Remote Sens. Environ.* 107, 639–655 .
- Wigneron, J.-P. , Laguerre, L. , Kerr, Y.H. , 2001. A simple parameterization of the L-band microwave emission from rough agricultural soils. *IEEE Trans. Geosci. Remote Sens.* 39, 1697–1707 .
- Wigneron, J.-P. , Waldteufel, P. , Chanzy, A. , Calvet, J.-C. , Kerr, Y. , 2000. Two-dimensional microwave interferometer retrieval capabilities over land surfaces (SMOS mission). *Remote Sens. Environ.* 73, 270–282 .
- Wilson D.J., Western A.W., Grayson R.B., 2004. Identifying and quantifying sources of variability in temporal and spatial soil moisture observations. *Water Resour Res*, 40(2):W02507.
- Yang, K.; Watanabe, T.; Koike, T.; Li, X.; Fuji, H.; Tamagawa, K.; Ma, Y.M.; Ishikawa, H. 2007. Auto-calibration system developed to assimilate AMSR-E data into a land surface model for estimating soil moisture and the surface energy budget. *J. Meteorol. Soc. Jpn.*, 85A, 229–242.
- Yilmaz, M.T., W.T. Crow, 2013. The optimality of potential rescaling approaches in land data assimilation, *J. Hydrometeorol.* 14, 650–660.
- Zehe, E., Graeff, T., Morgner, M., Bauer, A., and Bronstert, A., 2010. Plot and field scale soil moisture dynamics and subsurface wetness control on runoff generation in a headwater in the Ore Mountains. *Hydrol. Earth Syst. Sci.*, 14, 873–889.
- Zhang, D.; Zhou, G.; 2016. Estimation of Soil Moisture from Optical and Thermal Remote Sensing: A Review. *Sensors*, 16, 1308.
- Zhang, N., S. Quiring, T. Ochsner, and T. Ford. 2017. Comparison of three methods for vertical extrapolation of soil moisture in Oklahoma. *Vadose Zone J.* 16(10).
- Zhao, W., Li, Z.L., 2013. Sensitivity study of soil moisture on the temporal evolution of surface temperature over bare surfaces. *Int. J. Remote Sens.* 34 (9–10), 3314–3331.
- Zhao, D., C. Künzer, C. Fu, and W. Wagner, 2008. Evaluation of the ERS scatterometer-derived soil water index to monitor water availability and precipitation

- distribution at three different scales in China. *J. Hydrometeorol.*, vol. 9, no. 3, pp. 549–561.
- Zou, G. Y. 2007. Toward using confidence intervals to compare correlations. *Psychological methods*, 12(4), 399-413.
- Zwieback, S., Scipal, K., Dorigo, W., Wagner, W., 2012. Structural and statistical properties of the collocation technique for error characterization. *Nonlin. Process. Geophys.* 19(1), 69–80.
- Zreda, M., D. Desilets, T.P.A. Ferre, R.L. Scott, 2008. Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons. *Geophys. Res. Lett.* 35(21), L21402.
- Zreda, M., Shuttleworth, W.J., Zeng, X., Zweck, C., Desilets, D., Franz, T., Rosolem, R., 2012. COSMOS: the COsmic-ray Soil Moisture Observing System. *Hydrol. Earth Syst. Sci.* 16, 4079–4099.
- Zribi, M., Baghdadi, N., Holah, N., Fafin, O., 2005. New methodology for soil surface moisture estimation and its application to ENVISAT-ASAR multi-incidence data inversion. *Rem. Sens. Environ.* 96, 485–496.
- Zribi, M., Dechambre, M., 2003. A new empirical model to retrieve soil moisture and roughness from C-band radar data. *Remote Sens. Environ.* 84, 42–52.
- Zribi, M., T. Paris Anguela, B. Duchemin, Z. Lili, W. Wagner, S. Hasenauer, and A. Chehbouni, 2010. Relationship between soil moisture and vegetation in the Kairouan plain region of Tunisia using low spatial resolution satellite data. *Water Resour. Res.*, 46, W06508.