

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *06/12/2016* **par :**

Vivien STEFAN

Modélisation mixte et télédétection multi-résolution de l'évaporation du sol

---

JURY

Jean-Philippe Gastellu-Etchegorry	Professeur d'Université UT3	Président
Jean-Pierre WIGNERON	Directeur de Recherche INRA	Rapporteur
Albert Olioso	Directeur de Recherche INRA	Rapporteur
Olivier MERLIN	Chargé de Recherche CNRS	Directeur de thèse
Yann KERR	Directeur de Recherche CNES	Co-Directeur de thèse
Sylvain MANGIAROTTI	Chargé de Recherche IRD	Examinateur
Maria-José ESCORIHUELA	Ingénieur R & D isardSAT	Examinateur
Salah ER-RAKI	Professeur d'Université Cadi Ayyad Marrakech	Examinateur

---

École doctorale et spécialité :
SDU2E : Surfaces et Interfaces Continentales, Hydrologie

Unité de Recherche :
Centre d'Études Spatiales de la Biosphère - CESBIO (UMR 5126)

Directeur(s) de Thèse :
Olivier MERLIN et *Yann KERR*

Rapporteurs :
Jean-Pierre WIGNERON et *Albert OLIOSO*

Contents

List of Figures v

List of Tables ix

ACKNOWLEDGEMENTS xi

ABSTRACT xiii

RESUME xv

1 INTRODUCTION 1
- 1.1 Context . . . 1
- 1.2 Modeling approaches of soil evaporation . . . 3
  - 1.2.1 Mechanistic models . . . 3
  - 1.2.2 Simplified models . . . 4
  - 1.2.3 Phenomenological models . . . 7
- 1.3 Remote sensing approaches for soil evaporation . . . 8
  - 1.3.1 Vegetation index–based approaches . . . 8
  - 1.3.2 Temperature–based approaches . . . 8
  - 1.3.3 Soil moisture–based approaches . . . 10
- 1.4 Objectives . . . 11

INTRODUCTION (Français) 15
- 1.5 Contexte . . . 15
- 1.6 Approches de modélisation de l'évaporation du sol . . . 17
  - 1.6.1 Modèles mécanistes . . . 17
  - 1.6.2 Modèles simplifiés . . . 18
  - 1.6.3 Modèles phénoménologiques . . . 20
- 1.7 Approches pour l'évaporation du sol basées sur la télédétection . . . 21
  - 1.7.1 Approches basées sur l'indice de la végétation . . . 21
  - 1.7.2 Approches basées sur la température . . . 22
  - 1.7.3 Approches basées sur l'humidité du sol . . . 24
- 1.8 Objectifs . . . 26

2 DATA 29
- 2.1 Introduction . . . 29
- 2.2 Sites and *in situ* data description . . . 30
  - 2.2.1 Yaqui site . . . 30
    - 2.2.1.1 *In situ* flux data . . . 30

2.2.1.2 Meteorological data . . . 30
2.2.2 Haouz plain (R3 perimeter) . . . 31
2.2.2.1 *In situ* temperature data . . . 31
2.2.2.2 Meteorological data . . . 34
2.2.3 Urgell site . . . 35
2.2.3.1 *In situ* soil moisture data . . . 35
2.2.3.2 Meteorological data . . . 35
2.3 Spatial data . . . 36
2.3.1 Optical data . . . 37
2.3.2 Microwave data . . . 38
2.3.3 SAFRAN data . . . 39
2.3.4 DEM data . . . 40
2.3.5 Data pre–processing . . . 40
2.4 Conclusion . . . 42

3 Improving the wet and dry boundaries of a contextual evapotranspiration model 43
3.1 Introduction . . . 43
3.2 A contextual method . . . 45
3.2.1 SEB–1S overview . . . 45
3.2.2 Image–based extreme temperatures . . . 46
3.2.3 Adjustments to the estimation of the image–based extreme temperatures 48
3.3 Modeling extreme temperatures . . . 49
3.3.1 Different aerodynamic resistance formulations . . . 49
3.3.2 Vegetation extreme temperatures . . . 50
3.4 Results . . . 51
3.4.1 Evaluation of extreme soil temperatures . . . 52
3.4.1.1 Validation with *in situ* measurements . . . 52
3.4.1.2 Consistency with image–based extreme temperatures . . . 53
3.4.2 Application to evapotranspiration estimation at multiple resolutions . . . 55
3.5 Summary and Conclusions . . . 57
3.6 Article . . . 58

4 Global modeling of the temporal behavior of soil evaporative efficiency 95
4.1 Introduction . . . 96
4.2 Global Modeling . . . 97
4.2.1 Concepts . . . 97
4.2.2 Method formulation . . . 98
4.2.3 Algorithm . . . 99
4.3 Data description . . . 100
4.3.1 Transfert Eau Chaleur (TEC) model . . . 100
4.3.2 TEC Outputs . . . 100
4.3.3 Data pre-processing . . . 102
4.4 Application and Results . . . 103
4.4.1 SEE Global Model . . . 103
4.4.2 Model validation . . . 109
4.5 Summary and Conclusions . . . 110
4.6 Article . . . 111

5 Evaporation–based disaggregation of soil moisture data **119**
5.1 Introduction . . . 120
5.2 Methodology . . . 121
5.2.1 DISPATCH . . . 121
5.2.2 Temperature–based SEE model . . . 121
5.2.3 Soil moisture–based SEE model . . . 124
5.2.3.1 Linear model . . . 124
5.2.3.2 Non–linear model . . . 124
5.2.4 DISPATCH–E . . . 125
5.3 Results . . . 127
5.3.1 DISPATCH/DISPATCH–E sensitivity analysis . . . 127
5.3.1.1 Influence of soil/vegetation partitioning . . . 128
5.3.1.2 Influence of number of MODIS days and CNT . . . 133
5.3.2 Temperature–based *versus* soil moisture–based SEE estimates . . . 137
5.3.3 Validation of disaggregation methodology . . . 139
5.4 Summary and Conclusions . . . 141
5.5 Article . . . 142

6 CONCLUSIONS AND PERSPECTIVES **181**

CONCLUSIONS ET PERSPECTIVES **187**

Annex **193**

Bibliography **233**

List of Figures

2.1 The Yaqui site along with the six flux stations situated in the study area. 31

2.2 The dry soil plot located in the R3 perimeter. A pit of 15 cm depth was dug **(a)** and was kept fully dry by using a plastic sheeting to cover it in between experiments **(b)**. iButton sensors **(c)** and the handheld infrared thermometer **(d)** used to perform temperature measurements are also illustrated. 32

2.3 Mean and standard deviation of the *in situ* measurements acquired over the wet and dry soil plots in the R3 perimeter (April 30th, 2014). Measurements were acquired using either iButton (here abbreviated as "iB") sensors or the infrared thermometer (here abbreviated as "trad"). Vertical lines mark the begining and ending of the experiment. The average of the standard deviation of the instantaneous measurements is also computed. 33

2.4 The wet and dry soil plots along with the meteorological station located in the R3 perimeter. 34

2.5 The Urgell 60 km by 60 km (encompassing the 20 km by 20 km area where *in situ* measurements where performed) study area, alongside the *in situ* sampling areas and the XEMA meteorological stations (with a 2 m acquisition height). The Mollerussa (XI) station whose data was used in this study is shown in light blue. 36

2.6 40 km SMOS SM data corresponding to one MODIS tile that encompasses the 60 km by 60 km Urgell study area. SAFRAN 5 km resolution data is also shown. . 41

2.7 An example of one of the new, smaller SMOS grids (amongst a total of four). . . 41

3.1 For a given point J in the $LST - \alpha$ space, the EF is calculated as the ratio of IJ to IK. Underlying grey points correspond to 90 m resolution LST data acquired over Yaqui, on April 27th, 2008. 45

3.2 Meaning behind the edges of the polygons identified in the $LST - \alpha$ **(a)** and the $LST - f_{vg}$ **(b)** spaces. Underlying grey points correspond to 90 m resolution LST data acquired over Yaqui, on April 11th, 2008. 46

3.3 Soil T_{ext} simulated by EBsoil are plotted against the *in situ* measured T_{ext}, at the two soil plots in the R3 perimeter. Both RI (**a**) and MO (**b**) formulations of the aerodynamic resistance are taken into account when deriving modeled T_{ext}. . . . 52

3.4 Soil T_{ext} simulated by EBsoil using either the RI(**a**) or MO (**b**) aerodynamic resistance are plotted against the image–based T_{ext} for the Yaqui and Urgell regions separately. 53

3.5 Polygons obtained using either image–based Text_RS_1km (blue) and Text_RS_90m (light blue), or model–derived Text_EBsoil for the RI (magenta) and MO (red) aerodynamic resistance are shown in the $LST - \alpha$ and $LST - f_{vg}$ spaces. The spaces correspond to either 1 km resolution data (black dots) or 90 m resolution data (grey dots) acquired over the Yaqui area on December 30th. 54

3.6 1 km resolution ET estimates *versus* 90 m resolution simulated ET, aggregated at 1 km resolution, for the Yaqui site. The T_{ext} used as input to the SEB–1S consist of the T_{ext} simulated by EBsoil with the RI (**a**) and MO (**b**) aerodynamic resistance, and the T_{ext} retrieved from 1 km resolution data (**c**). 56

4.1 Soil evaporation (LE_s) and potential soil evaporation (LE_p) as derived from TEC, using four formulations of the pedotransfer functions: *Wösten* (1997) (a), *Cosby et al.* (1984) (b), *Rawls and Brackensiek* (1985) (c) and *Vereecken et al.* (1989) (d) 101

4.2 Daily cycle of soil evaporation (LE_s) and potential soil evaporation (LE_p). Data was generated using TEC, for the *Wösten* (1997) pedotransfer function. 101

4.3 Evaporation (LE_s) generated using TEC with the "wos" formulation of the pedotransfer function, using four different soil textures. 102

4.4 Original (circle) and processed (lines) SEE time series. The processed SEE time series is used as input to the global modeling. 103

4.5 Original (black) and model (red) phase portraits for various projections of the phase space. 105

4.6 Original (black) and reconstructed (red) SEE time series. 106

4.7 Original (black) and reconstructed (red) SEE time series. The reconstructed SEE corresponds to simulations starting from different initial conditions (red circles). The green curves represent the error between the model and the original data. . 107

4.8 Phase portraits of the original (black) and reconstructed (red) SEE series for various projections of the phase space. The reconstructed series corresponds to simulations starting from different initial conditions (red circles). The green curves represent the phase portraits of the error between the model and the original data. 108

4.9 Error growth $e_t(h)$ (in color) as a function of t (expressed as the Day of Year - DoY - in the abscissa) and the horizon of prediction h (expressed in hours, as the ordinate). 109

4.10 The standard deviation of the error growth is plotted (in corresponding SEE units and in percentage) as a function of the prediction horizon. 110

5.1 The four different zones that can be distinguished in the $LST - f_{vg}$ space (*Merlin et al.*, 2012a). 123

5.2 Visual representation of three non–linear SM–based SEE models. 125

5.3 Schematic overview of DISPATCH (left) and DISPATCH–E (right). 126

5.4 The Urgell study area on which both DISPATCH and DISPATCH–E were applied. 128

5.5 The temperature–based SEE is plotted against the SMOS SM, for the entire time series, for the Aqua overpass, for the ABCD zones defined in the $LST - f_{vg}$ space. Temperature–based SEE is derived either from MODIS, or from EBsoil, with its two aerodynamic resistance formulations: MO and RI. The number of MODIS days taken into account is equal to 1. 129

5.6 Same as Figure 5.5, but for the ABC zones. 130

5.7 Same as Figure 5.5, but for zone A. 130

5.8 Same as Figure 5.5, but for the Terra overpass. 131

5.9 Same as Figure 5.6, but for the Terra overpass. 131

5.10 Same as Figure 5.7, but for the Terra overpass. 132

5.11 The temperature–based SEE is plotted against the SMOS SM, for the entire time series, for the Aqua overpass, for the ABC zones defined in the $LST - f_{vg}$ space. Temperature–based SEE is derived either from MODIS, or from EBsoil, with its two aerodynamic resistance formulations: MO and RI. The CNT value considered is 12. 134
5.12 Same as Figure 5.11, but for a CNT value equal to 8. 135
5.13 Same as Figure 5.11, but for a CNT value equal to 4. 135
5.14 Same as Figure 5.11, but for the Terra overpass. 136
5.15 Same as Figure 5.12, but for the Terra overpass. 136
5.16 Same as Figure 5.13, but for the Terra overpass. 137
5.17 The observed (temperature–based) SEE and the simulated (SM–based) SEE is plotted against the SMOS SM, for the entire time series, for the Aqua overpass. The observed SEE is derived either from MODIS, or from EBsoil. The SM–based SEE uses the non–linear K03 model and is calibrated from the observed SEE. . . 138
5.18 Same as Figure 5.17 but for the Terra overpass. 139
5.19 The SMOS, DISPATCH and DISPATCH–E SM are plotted against 3 km aggregated *in situ* measurements, for the Aqua overpass. 140
5.20 Same as Figure 5.19, but for the Terra overpass. 140

List of Tables

1.1 A list of simplified E modules used by some of the most well-known land surface models (Merlin 2016, *Mémoire HDR*). 6
1.2 Une liste des modules simplifiés de l'E utilisés par certains modèles de surface les plus connus (Merlin 2016, *Mémoire HDR*). 20

2.1 List of variables and acquisition heights, as measured by the XEMA meteorological stations 36
2.2 List of variables and acquisition heights, as estimated by SAFRAN 39

3.1 Root mean square error (RMSE), bias, correlation coefficient (R) and slope of the linear regression between soil T_{ext} (as given by EBsoil) and the soil T_{ext} observed *in situ* at the two soil plots, located in the R3 perimeter. Modeled T_{ext} take into account either the RI or MO aerodynamic resistance. 52
3.2 Difference between the soil T_{ext} simulated by EBsoil with the RI and MO aerodynamic resistance and the soil T_{ext} retrieved using 90 m resolution data, for the Yaqui and Urgell regions separately. 54
3.3 Mean and standard deviation (σ) of the root mean square difference (RMSD), bias, correlation coefficient (R) and slope of the linear regression between the ET simulated at 1 km resolution and the 90 m resolution simulated ET aggregated at 1 km resolution, for data over the Yaqui region. The T_{ext} used as input to SEB-1S consist of T_{ext} simulated by EBsoil with the RI and MO aerodynamic resistance and the T_{ext} retrieved from 1 km resolution data. 57

4.1 Parameters obtained for the 4 dimensional model presented in Equation 4.3. . . . 104

5.1 Correlation coefficient, slope of the linear regression, bias and RMSD between SMOS SM/DISPATCH SM/DISPATCH-E SM and 3 km *in situ* data are reported for each MODIS overpass individually, as well as both overpasses combined. Both DISPATCH and DISPATCH-E are tested in linear and nonlinear modes. 140

ACKNOWLEDGEMENTS

First and foremost I would like to thank my two supervisors, Olivier Merlin and Yann Kerr, whose contributions in stimulating suggestions and encouragement helped me to coordinate my PhD thesis in the best possible manner. I would like to express a special gratitude to Olivier, who has also been my internship supervisor, and who, for three and a half years has been nothing less than a continuous source of support and motivation. I appreciate all his contributions of time and ideas to make my PhD experience productive and stimulating.

A special thanks goes to all the people I have worked throughout my PhD: Sylvain, Maria–José, Salah, André, Jeff, Beatriz and Yoann, who have been a source of good advice and collaboration. I am especially grateful to Sylvain for his enthusiasm and willingness to guide me in my work with the global modeling. I thank Maria–José for the opportunity she offered me to be a part of the REC project, and the warm welcome she has provided as part of my stay in Barcelona. I greatly appreciate André's consent to use the TEC simulations in this PhD, as well as the great work Beatriz has done with the DISPATCH Input Interface. A sincere thanks goes to Jeff as well, who has given me the opportunity to take part in the SMAPEx–4 campaign that took place in Australia.

I would also like to thank the CESBIO staff, with whom I always interacted in a great way, especially Delphine, Monique, Dominique and Emilie, who from an administrative point of view have been very prompt and helpful. A special gratitude goes to the isardSAT team in Barcelona, who have been imensely welcoming, helpful and friendly throughout my stay there. I have not only discovered a new type of working environment, but also felt part of a small family.

A very sincere thanks goes to the members of the jury and to the reviewers, who have kindly accepted this role. I also gratefully acknowledge the funding sources (ANR JC MIXMOD–E and H2020 RISE REC projects) that made my PhD work possible.

A warm thanks to all my friends in Romania, Toulouse, Barcelona, Australia and Morocco, who have offered their support and understanding throughout these years. And last, but not least, my deep appreciation to my family, for all their love and encouragement, especially my parents who have always sincerely supported me in all my pursuits, regardless of where these have taken me. Thank you.

ABSTRACT

Evapotranspiration (ET) and its two main components, soil evaporation and plant transpiration, play a crucial role in the hydrological cycle, especially over semi–arid regions. Monitoring soil evaporation at multiple scales is especially relevant to a multitude of agronomic, hydrologic, meteorological and climatic applications, where remote sensing–based approaches have great potential via two evaporation–related observable variables: the land surface temperature (LST) and the surface soil moisture (SM).

The aim of this PhD was to try and improve the spatio–temporal representation of soil evaporation by developing a multi–scale modeling approach that makes use of the synergy between readily available remote sensing data. A new methodology was developed to estimate the soil evaporative efficiency or SEE (ratio of actual to potential soil evaporation) at 100 m to 40 km resolution by combining 40 km resolution microwave–derived SM, 100 m to 1 km resolution thermal–derived LST, 100 m to 1 km resolution reflectance–derived vegetation cover/surface albedo and available meteorological data. The approach was tested over three different semi–arid irrigated areas in Morocco, Mexico and Spain.

In a first step, a soil energy balance model (EBsoil) was proposed to improve the determination of the dry/wet boundaries of a LST–based ET model. EBsoil outputs were validated against *in situ* measurements of dry/wet soil temperatures over the Moroccan site. ET estimates were subsequently derived either from remote sensing data solely or by using EBsoil and intercompared at 100 m and 1 km resolutions with a reference ET over the Mexican site. Results showed that integrating EBsoil improves ET estimates at a medium (1 km) resolution.

The next step consisted in looking into the temporal dynamics of SEE, using the global modeling technique, approach used for a very small number of environmental contexts, with even fewer models obtained that are able to reproduce the original dynamics. For the first time ever, a four dimensional chaotic model was obtained for the daily cycle of SEE, able to reproduce the dynamics of the variable with a good approximation in the phase domain.

In a final step, the two remote sensing approaches used in deriving SEE - as a function of LST or SM - were combined within a disaggregation algorithm (DISPATCH) of SMOS (Soil Moisture and Ocean Salinity) SM data. EBsoil was integrated in DISPATCH and a new algorithm was developed: DISPATCH–E. An indirect validation of the SEE modeling approach was performed by assessing DISPATCH–E's results in terms of 1 km resolution SM estimates over the Spanish site. Implementing EBsoil significantly improves the downscaled SM. Furthermore it provides reference points for calibrating multi–resolution SEE models, as well as being an independent way of assessing uncertainty in remotely sensed data.

RESUME

L'évapotranspiration (ET) et ses deux composantes, l'évaporation du sol et la transpiration des plantes, jouent un rôle crucial dans le cycle hydrologique, en particulier dans les régions semi–arides. Le suivi de l'évaporation du sol à des échelles multiples est particulièrement intéressant pour une multitude d'applications agronomiques, hydrologiques, météorologiques et climatiques, pour lesquelles la télédétection a un grand potentiel via deux observables clés liées à l'évaporation: la température de surface (Land Surface Temperature - LST) et l'humidité du sol en surface (Soil Moisture - SM).

Cette thèse avait pour but d'améliorer la représentation spatio–temporelle de l'évaporation du sol, par une modélisation multi-échelle capable d'utiliser la synergie des données de télédétection disponibles. Une nouvelle méthodologie a été développée afin d'estimer l'efficacité évaporative du sol ou SEE (Soil Evaporative Efficiency, définie comme le rapport entre l'évaporation du sol réelle et celle potentielle), avec une résolution comprise entre 100 m et 40 km. Elle combine l'humidité du sol dérivée à 40 km de résolution des données micro–ondes, la température de surface, le couvert végétal et l'albédo de surface issues des capteurs optiques avec une résolution spatiale variant de 100 m and 1 km, et des données météorologiques. Cette approche a été testée sur des zones irriguées semi–arides au Maroc, au Mexique et en Espagne.

Dans une première étape, un modèle du bilan d'énergie du sol (EBsoil) a été proposé pour estimer les limites humide/sèche d'un modèle de l'ET basé sur la LST. Les sorties du modèle EBsoil ont été validées avec des mesures *in situ* de la température du sol réalisées au Maroc. Les estimations de l'ET obtenues soit uniquement à partir des données de télédétection, soit en utilisant EBsoil, ont ensuite été comparées avec une ET de référence sur le site mexicain pour des résolutions de 100 m et 1 km. Les résultats ont montré que l'intégration de EBsoil améliore les estimations de l'ET à la résolution de 1 km.

L'étape suivante s'est concentrée sur la dynamique temporelle de la SEE, en utilisant la technique de la modélisation globale. Cette approche a été utilisée jusqu'à présent pour un petit nombre de contextes environnementaux, avec encore moins de modèles obtenus qui sont en mesure de reproduire la dynamique d'origine. Pour la première fois, un modèle chaotique de dimension quatre a été obtenu pour le cycle journalier de la SEE, capable de reproduire la dynamique de la variable avec une bonne approximation dans l'espace des phases.

Dans une dernière étape, les deux approches de télédétection utilisées pour calculer la SEE - en fonction de la LST ou de la SM - ont été combinées dans un algorithme de désagrégation (DISPATCH) des données SM SMOS (Soil Moisture and Ocean Salinity). EBsoil a été intégré dans DISPATCH et un nouvel algorithme a été développé: DISPATCH–E. Une évaluation indirecte de l'approche de modélisation de la SEE a été implémentée sur le site espagnol en comparant la SM DISPATCH-E aux mesures *in situ*. L'intégration de EBsoil améliore significativement la SM désagrégée à 1 km de résolution. De plus, elle fournit des points de référence pour l'étalonnage des modèles multi–résolution de la SEE, ainsi qu'une évaluation indépendante de l'incertitude des données de télédétection.

Chapter 1

INTRODUCTION

Contents

1.1 Context 1
1.2 Modeling approaches of soil evaporation 3
1.2.1 Mechanistic models 3
1.2.2 Simplified models 4
1.2.3 Phenomenological models 7
1.3 Remote sensing approaches for soil evaporation 8
1.3.1 Vegetation index–based approaches 8
1.3.2 Temperature–based approaches 8
1.3.3 Soil moisture–based approaches 10
1.4 Objectives 11

1.1 Context

In the context of nowadays increasing population and growing economy, more and more water resources are used for expanding irrigation over agricultural areas, as well as by different industries, which demand more and more energy consumption. Concerns arise when dealing with limited water resources given the circumstances of growing population, climate change and increasing environmental issues. At a global scale, approximately 10% of the total fresh water is needed per year to account for the needs of the entire Earth's population (*Oki and Kanae*, 2006). It might not seem as a high percentage given the total available fresh water, but one must keep in mind that there are areas where water scarcity is a real problem, affecting more than two billion people worldwide. This entails a real problem, when considering that in order to meet rising food needs (linked with a growing population), an increase of 70–100% in crop production is needed over the next three decades, as reported by several organizations. Agriculture thus is a notable constraint on water resources, as the relationship between water availability and food production is crucial (*FAO*, 2009). 70% of total water is used for agriculture, with a 40% of the total food supply being produced by irrigated agriculture.

The importance of water resources in agriculture is of particular interest over Mediterranean countries, where up to 80% of the total water is used for irrigation (*FAO*, 2015; *Garrido et al.*, 2010). These areas are also the most prone to climate change (*IPCC*, 2013; *Stocker et al.*, 2014), while having only 1% of global fresh water resources. With the annual rainfall predictions over the next decades being largely uncertain in terms of magnitude and spatial distribution

(*Solomon*, 2007), IPCC also predicts that any changes in precipitation patterns can decrease the productivity of rainfed agriculture by almost 50% by the end of 2020 (*Bates et al.*, 2008). Sustainable water use is a crucial matter, as mentioned by the Roadmap to a Resource Efficient Europe (*Comission*, 2011), as part of the EU 2020 strategy. Water use efficiency in agriculture is stronlgy linked to water scarcity and drought, as reported by EEA Report No. 1/2012.

Thus, on–farm irrigation management is required to be optimized, in this context of increasing demands but diminishing resources. Since a mismatch is registered between irrigation needs and the actual amount of water used for irrigation (*Kharrou et al.*, 2013), optimization should be achieved through adjusting irrigation to crop water requirements during the crop growth season. Over irrigation should be avoided since it entails an increase in exported pollutants, combined with an unnecessary increase of water use. On the other hand, under irrigation entails a dropping in crop yield (*Salvador et al.*, 2011). An increase in crop production yield could be achieved by better managing irrigation and fertilization use (*Mueller et al.*, 2012).

In this context, hydrology plays an important role, as a discipline that searches to understand the effects of natural and man–made changes brought upon variables such as soil moisture, precipitation, drainage, runoff. Since water is a renewable resource in the hydrological cycle, a better understanding of all the fluxes that intervene in the hydrological cycle could improve monitoring capabilities and provide an insight to a better management.

Monitoring water resources is therefore strongly linked with modeling hydrological fluxes such as: soil infiltration, runoff and evapotranspiration. Currently, a multitude of large scale modeling approaches exist, ranging from simple to more complex, that take into account other factors such as vegetation. Historically, the first global scale modeling of hydrological fluxes is based on a simple model, known as the "bucket" model (*Manabe*, 1969). This approach considers the surface as a simple reservoir, where runoff and evapotranspiration take place above a certain threshold value, estimated in connection with the amount of soil moisture present. Following models (*Deardorff*, 1978; *Mahrt and Pan*, 1984; *Sellers et al.*, 1986; *Noilhan and Planton*, 1989) have taken into consideration the presence of vegetation in order to separately account for soil and vegetation fluxes. Further approaches have considered the soil as a multi-layer system (*Chanzy and Bruckler*, 1993) that model fluxes based on the Richards equation (*Viterbo and Beljaars*, 1995) and a dynamical representation of the vegetation (*Calvet et al.*, 1998).

Evapotranspiration (ET) is one of the most important hydrological fluxes, having a high importance especially over semi–arid regions, where it is by far the main outward flux. Besides predicting soil water availability as previously mentioned, it plays a central role in flood forecasting (*Bouilloud et al.*, 2010), in rainfall forecasting (*Findell et al.*, 2011), in projecting changes in the manifestation of droughts (*Sheffield and Wood*, 2008) and heatwaves (*Seneviratne et al.*, 2006). In particular, it is the only hydrological flux linking both the energy and the water balance. It does so through two main variables essential in its estimation: the land surface temperature (the signature of ET, that intervenes in the energy balance) and the soil moisture (a state variable that intervenes in the water cycle). Partitioning ET into its two main components, soil evaporation (E) and plant transpiration (T) is essential in modeling land–atmosphere interactions and vegetation water uptake. In the first stages of a growing crop, water is mostly lost through E; however, once the crop is well developed and has a full soil coverage, T becomes the main process. So, E is the part which is not used for crop productivity (*Wallace*, 2000), whereas T is linked to the root zone water availability (*Allen et al.*, 1998). Furthermore, since this partitioning is currently highly uncertain and model–dependent, more in–depth knowledge on this process could contribute to better constraining the physics of land surface models (*Gutmann and Small*, 2007). Unfortunately, field measurements of E and T are quite scarce. Having accurate estimates at multiple scales (crop, catchment and mesoscale) and with a good temporal resolution are undoubtedly essential in a better monitoring and management of resources

(*Williams et al.*, 2004; *Kool et al.*, 2014). Partitioning ET into E and T is especially crucial in the monitoring of plant water uptake and water stress (*Porporato et al.*, 2001; *Er-Raki et al.*, 2010), in the framework of limited water resources. Evaporation in particular accounts for 20–40% of global ET (*Lawrence et al.*, 2007; *Schlesinger and Jasechko*, 2014) and it mostly comes from soil moisture (*Good et al.*, 2015). Since it is an essential boundary condition between soil and atmosphere, a better representation of soil evaporation could not only help with a better monitoring and management of water resources, but also with a better understanding of land–atmosphere interactions (*Feddes et al.*, 2001).

In this respect, remote sensing is of particular interest, being cost effective in monitoring and mapping E and E–related variables (such as vegetation cover, land surface temperature and soil moisture) over large areas. Moreover, remote sensing is a way of estimating E at muliple resolutions. This is important since different scales are needed, depending on the application domain of the variable of interest. More notably, crop yield and water consumption are evaluated at field scale, surface runoff and flood risks at the watershed scale and atmospheric circulation is dictated by surface fluxes at mesoscale. In this context, remote sensing is a way of coupling available data at multiple resolutions with hydrological models. On the other hand, a deep understanding of hydrological processes at a range of scales could help improve the quality of available remote sensing data.

An overview of existing E modeling approches, remote sensing techniques and links with other variables such as vegetation cover, soil moisture and land surface temperature are discussed below.

1.2 Modeling approaches of soil evaporation

Modeling approaches of E can be roughly classified in three categories: mechanistic approaches, simplified formulations and phenomenological approaches. An overview of each approach is presented below.

1.2.1 Mechanistic models

These approaches consider the soil as a multi-layer system and proceed to theoretically represent the coupled mass and heat exchange between the soil layer and the atmosphere. Some examples of mechanistic models can be found in *Chanzy and Bruckler* (1993), *Bittelli et al.* (2008). Vapor fluxes at the interface between soil and atmosphere can be estimated by solving the coupled mass and heat equation for all layers. Bare soil evaporation can then be computed at a given height by reducing the top soil layer vapor flux using an aerodynamic resistance for water vapor transfer. However, solving these equations demands knowledge on the soil hydraulic properties on every soil layer. These soil hydraulic properties consist of:

- saturated hydraulic conductivity ($\mathrm{kg{\cdot}s{\cdot}m^{-3}}$)
- residual moisture content ($\mathrm{m^3{\cdot}m^{-3}}$)
- saturated moisture content ($\mathrm{m^3{\cdot}m^{-3}}$)
- various empirical shape parameters (*Brooks and Corey*, 1964; *Campbell*, 1974; *Van Genuchten*, 1980)

The main limitations of mechanistic models based on soil hydraulic properties are their sensitivity to initial and boundary conditions as noted in *Chanzy et al.* (2008). Initial conditions refer

in particular to the soil moisture profile, which cannot be measured over large areas with the help of remote sensing, since remote sensing data are mostly provided for a surface layer only. Boundary conditions involve the climatic conditions at the soil surface and the change in water content or water potential with depth. Additionally, the computation time of such models is quite long, and the numerical solution can sometimes be ill-conditioned, and therefore unstable.

The downsides of using soil hydraulic parameters are that they are very hard to measure, they require quite expensive techniques, and most importantly, it is hard to apply them at integrated spatial scales. To remove such limitations, a work–around has been introduced: pedotransfer functions. Pedotransfer functions (*Wösten*, 1997; *Cosby et al.*, 1984; *Rawls and Brackensiek*, 1985; *Vereecken et al.*, 1989; *Twarakavi et al.*, 2009) translate readily available soil texture data into soil hydraulic properties.

The main advantage of mechanistic models is their contribution in better understanding the physical processes linked with E, such as vapor diffusion, capillary rise and gravity drainage. However, they present a number of disadvantages that restrain their application over large areas. Pedotransfer functions are very uncertain and have large variations in space (*Soet and Stricker*, 2003). This is mainly due to soil structure (pore size, distribution, connectivity), presence of soil aggregates and presence of biomass. These pedotransfer functions vary more within a texture class than in–between classes (*Gutmann and Small*, 2007). Additionally, a high number of parameters which cannot be measured directly at the application scale (*Demaria et al.*, 2007) are taken into account in mechanistic models, often leading to an overparameterization. An additional empirical soil resistance can sometimes be required in order to accurately estimate E (*Bittelli et al.*, 2008).

To summarize, the high uncertainty of pedotransfer functions, the sensitivity to initial and boundary conditions, and the overparameterization render mechanistic models impractical for large scale applications. Therefore, simpler models have been developed.

1.2.2 Simplified models

Because of the difficulties and constraints that come with the implementation of mechanistic models over large areas, an alternative approach has been developed, that estimates E as a function of soil moisture. These simplified models are the most widely used approaches to estimate E, with the E module of most land surface schemes being based on these approaches.

The first model, also known as the "bucket" model (*Manabe*, 1969; *Robock et al.*, 1995), calculated evaporation by using the following equation:

$$LE_s = SEE * LE_p \tag{1.1}$$

where LE_s ($\mathrm{W{\cdot}m^{-2}}$) is the latent heat flux, LE_p ($\mathrm{W{\cdot}m^{-2}}$) the potential heat flux, and SEE (-) the soil evaporative efficieny. SEE is defined as the ratio of actual to potential evaporation and in the "bucket" model it is expressed as:

$$SEE = \begin{cases} 1, & \text{if } SM \geqslant SM_c \\ \frac{SM}{SM_c}, & \text{if } SM < SM_c \end{cases} \tag{1.2}$$

with SM the soil moisture, and SM_c being equal to 75% of SM_{fc}, where SM_{fc} is the soil moisture at field capacity.

Improvements have been made to the above model, by separating the two distinct components: soil and vegetation, thus using a dual–source formulation (*Shuttleworth and Wallace*, 1985). Among these methods, four approaches are mostly used:

- The r_{ss} (soil surface resistance) method:

$$LE_s = \frac{\rho C_p}{\gamma} \frac{e_{sat}(T_s) - e_a}{r_{ss} + r_{ah}} \tag{1.3}$$

- The α method:

$$LE_s = \frac{\rho C_p}{\gamma} \frac{\alpha e_{sat}(T_s) - e_a}{r_{ah}} \tag{1.4}$$

- The β method:

$$LE_s = \beta \frac{\rho C_p}{\gamma} \frac{e_{sat}(T_s) - e_a}{r_{ah}} \tag{1.5}$$

- The "threshold" method:

$$LE_s = min(LE_p, LE_{max}) \tag{1.6}$$

with r_{ss} (s·m^{-1}) the resistance to vapor diffusion in large soil pores, α (-) a unitless factor scaling the saturated vapor pressure down to the actual vapor pressure at the soil surface, β (-) a unitless factor that scales potential evaporation down to actual evaporation, ρ (kg·m^{-3}) the air density, C_p (J·kg^{-1}·K^{-1}) the specific heat of air at constant pressure, γ (Pa·K^{-1}) the psychrometric constant, $e_{sat}(T_s)$ (Pa) the saturated vapor pressure at soil temperature, e_a (Pa) the vapor pressure of air, r_{ah} (s·m^{-1}) the aerodynamic resistance to heat transfer and LE_{max} (W·m^{-2}) the maximum soil-limited water flux from below the surface.

In literature, the "threshold" method is also known as the "Priestley–Taylor" or the "demand-supply" method. LE_p is calculated using the aerodynamic, Penman, or Priestley–Taylor methods. Various studies in literature are based on these models (*Mahfouf and Noilhan*, 1991; *Lee and Pielke*, 1992; *Ye and Pielke*, 1993; *Mihailovic et al.*, 1995; *Dekic et al.*, 1995; *Cahill et al.*, 1999). There are two ways of deriving the r_{ss}, α , β and LE_{max} parameters: using a physical approach, or an empirical approach. In the physical approach, the expressions are obtained either by deriving Fick's law of diffusion (*Dickinson et al.*, 1986; *Wetzel and Chang*, 1988; *Sakaguchi and Zeng*, 2009), or by using thermodynamical considerations (*Philip and de Vries*, 1957). Additional empirical parameters are introduced in addition to soil hydraulic parameters, introduced by simplifications made to the theoretical diffusion equation. In empirical approaches, ad–hoc expressions (*Manabe*, 1969; *Noilhan and Planton*, 1989) are introduced or curve–fittings are made based on experimental data (*Sellers et al.*, 1992). Even though studies have fostered the appearance of many formulations, no general agreement regarding the best way to parameterize the evaporation process at extended spatial scales has been reached (*Desborough et al.*, 1996; *Sakaguchi and Zeng*, 2009). There are however three aspects that are generally agreed upon: i) better soil moisture and E estimates might be obtained by using rather simple empirical approaches rather than physical approaches (*Dekic et al.*, 1995; *Mihailovic et al.*, 1995; *Yang et al.*, 1998), ii) the β formulation appears to be more robust than the α formulation (*Cahill et al.*, 1999; *Van den Hurk et al.*, 2000) and iii) even though all approaches are based on surface soil moisture, the effects of soil surface state (e.g.: soil roughness, presence of stubble or mulch, shrinkage cracks associated with clayey soils, soil heterogeneity) have not been thoroughly investigated (*Sakaguchi and Zeng*, 2009).

Most of the evaporation modules implemented in the existing land surface models are based on one of the four formulas presented above. Information regarding the simplified E modules present in state–of–the–art land surface models is presented in Table 1.1.

Table 1.1: A list of simplified E modules used by some of the most well-known land surface models (Merlin 2016, *Mémoire HDR*).

Nature	**Name**	**Evaporation module**	**Formulation**	**Model description**
Empirical	BUCKET	*Manabe* (1969)	β	*Robock et al.* (1995)
Empirical	CABLE	*Wang et al.* (2011)	threshold	*Kowalczyk et al.* (November 2006)
Empirical	CLASS V3	*Lee and Pielke* (1992)	β	*Verseghy* (1996)
Empirical	CLSM	*Koster and Suarez* (1996)	r_{ss}	*Koster et al.* (2000)
Empirical	HTESSEL	*Balsamo et al.* (2011)	r_{ss}	ECMWF 2006
Empirical	IAP94	*Sellers et al.* (1992)	r_{ss}	*Yongjiu and Qingcun* (1997)
Empirical	ICARE	*Sellers et al.* (1992)	r_{ss}	*Merlin et al.* (2006a)
Empirical	ISBA	*Noilhan and Planton* (1989)	α	*Noilhan and Mahfouf* (1996)
Empirical	LaD GFDL	*Manabe* (1969)	β	*Milly and Shmakin* (2002)
Empirical	LAPS	*Sun* (1982)	α,r_{ss}	*Mihailović* (1996)
Empirical	MOSAIC	*Koster and Suarez* (1996)	r_{ss}	*Koster and Suarez* (1996)
Empirical	Noah LSM	*Sellers et al.* (1992)	α,r_{ss}	*Niu et al.* (2011)
Empirical	SECHIBA	*Ducoudré et al.* (1993)	r_{ss}	*Ducoudré et al.* (1993)
Empirical	SEWAB	*Noilhan and Planton* (1989)	α	*Mengelkamp et al.* (1999)
Empirical	SiB2	*Sellers et al.* (1992)	α,r_{ss}	*Sellers et al.* (1996)
Empirical	SSIB	*Camillo and Gurney* (1986)	α,r_{ss}	*Xue et al.* (1991)
Empirical	SWB	*Manabe* (1969)	β	*Schaake et al.* (1996)
Empirical	VIC	*Wood et al.* (1992)	β	*Wood et al.* (1992)
Physical	BASE	*Dickinson et al.* (1986)	threshold	*Desborough and Pitman* (1998)
Physical	BATS	*Dickinson et al.* (1986)	threshold	*Dickinson et al.* (1993)
Physical	BEST	*Dickinson et al.* (1986)	threshold	*Pitman et al.* (1991)
Physical	CAPS	*Mahrt and Pan* (1984)	threshold	*Chang et al.* (1999)
Physical	CLM	*Sakaguchi and Zeng* (2009)	α,r_{ss}	*Oleson et al.* (2010)
Physical	GISS	*Gardner and Hillel* (1962)	threshold	*Rosenzweig and Abramopoulos* (1996)
Physical	MIT	*Entekhabi and Eagleson* (1989)	β	*Entekhabi and Eagleson* (1989)
Physical	PLACE	*Wetzel and Chang* (1988)	threshold	*Wetzel and Boone* (1995)
Physical	SEtHyS	*Soarès et al.* (1988)	threshold	*Coudert et al.* (2006)
Physical	SiSPAT	*Philip and de Vries* (1957)	α	*Braud et al.* (1995)
Physical	SWAP	*Gusev and Nasonova* (1998)	β	*Gusev and Nasonova* (1998)
Physical	TOPLATS	*Milly* (1986)	threshold	*Famiglietti* (1994)

1.2.3 Phenomenological models

Phenomenological models differentiate themselves through the fact that they are neither built on ad–hoc assumptions, nor derived from theory. Instead, they are based on observational data (*Sivapalan et al.*, 2003), still providing a physical or semi–physical interpretation of the model's parameters.

By modeling SEE instead of soil evaporation, one separates the two main factors controlling E: the evaporative demand LE_p and the soil water availability. This is only partially fulfilled, since the LE_p is linked to soil water availability (*Merlin et al.*, 2011). However, this normalization of LE_s removes the first order effects of the evaporative demand on actual evaporation, while fixing the physical limits between ~ 0 and 1. SEE reaches the ~ 0 value when the soil water availability is insignificant, and the 1 value when the soil water availability is maximum.

SEE also helps with distinguishing the different evaporation cycles, which are linked either to daily/seasonal/climatic variations in the evaporative demand, or to variations in soil water availability due to rain/irrigation.

A first attempt to associate an experimental parameter of SEE to soil texture and aerodynamic conditions was made by Komatsu in (*Komatsu*, 2003):

$$SEE = 1 - exp(\frac{SM}{SM_c}) \tag{1.7}$$

with SM_c being a soil parameter whose values depend on the soil type and wind speed. The depth of the surface layer on which the analysis was performed was of several milimeters, which is considerably thinner than the normal thickness of the top surface layer used in most land surface models, which is of several centimeters. Finding a general way to parameterize SEE is oftenly hindered by the drying (around noon) of the top few mm of soil, which prohibits E, independently of the availability of soil water that is present underneath the surface (*Mahrt and Pan*, 1984; *Dickinson et al.*, 1986; *Soarès et al.*, 1988; *Wetzel and Chang*, 1988; *Van de Griend and Owe*, 1994; *Heitman et al.*, 2008; *Shahraeeni et al.*, 2012). Therefore, a new way of deriving SEE with a shape that adjusts with the depth of soil moisture readings was developed. *Merlin et al.* (2011) have looked into the possibility of taking into consideration the soil moisture gradient in the topsoil by introducing a simple parameterization that depends on potential evaporation. The model was tested at two sites located in the south of France, at a daily time scale. Moreover, in their latest study (*Merlin et al.*, 2016), they developed a meta–analysis approach to represent SEE over a large range of soil and atmospheric conditions, by using data collected at over 30 sites worldwide. A significant relationship has been observed between the semi–empirical E model parameter and soil texture. The model was then compared to the evaporation module of several land surface models such as ISBA (Interaction Sol–Biosphère–Atmosphère, *Noilhan and Planton* (1989)), CLM (Community Land Model, *Oleson et al.* (2013)) and H–TESSEL (Hydrology–Tiled ECMWF Scheme for Surface Exchange over Land, *ECMWF* (2014)), and the model was proven to perform relatively well.

The main advantage of phenomenological models is their ability to use readily available data, their potential coupling with remote sensing data making them adapted for large scale applications, all the while being consistent with theoretical assumptions. However, the main drawback is linked to their semi-empirical nature, combined with the use of effective parameters, which makes it difficult to derive evolution scenarios.

1.3 Remote sensing approaches for soil evaporation

Amongst all variables that can be derived from remote sensing, there are three in particular that are pertinent to the estimation of E: vegetation index, land surface temperature and near surface soil moisture. The methodologies based on each variable are described below.

1.3.1 Vegetation index–based approaches

Complementary to mechanistic and physical models, there is an extensive array of methods to estimate E by using remote sensing data. One of the most popular methods is the FAO-56 (*Allen et al.*, 1998; *Allen*, 2000; *Allen et al.*, 2005; *Simonneaux et al.*, 2008) approach, forced by NDVI (Normalized Difference Vegetation Index) data (*Bausch and Neale*, 1989). The approach uses meteorological measurements, a crop phenology that has been computed from NDVI–derived fractional green vegetation cover (f_{vg}), and a crop–specific coefficient in order to derive a potential ET. The actual ET is then estimated as a function of this potential ET. In its dual–coefficient form, the ET is constrained by soil moisture, based on a soil water budget model that requires *a priori* knowledge of irrigation volumes and dates. So, even though a soil module can be integrated in order to separate E and T, and express vegetation stress as a function of soil water availability, the main method is still based on the following assumptions (*Er-Raki et al.*, 2010):

- irrigation volumes and precipitation data are supposed to be known;
- T is considered either as its value at potential rate, or it is calculated via root zone soil moisture (estimated using a soil model forced by available irrigation data);
- E is either set to a constant value, or it is calculated via near surface soil moisture (estimated using a soil model forced by available irrigation data);
- drainage is often neglected in deep soil layers.

The compromise between the intricate parameterization of the model and the availability of the input data has rendered this methodology very attractive from an operational stand point. The downside of the FAO-56 method is its applicability over very large areas, since the coefficients it uses are calibrated from field measurements, which are mostly impacted from a temporal and spatial heterogeneity point of view (*Long and Singh*, 2010).

1.3.2 Temperature–based approaches

Land surface temperature (LST) is a pertinent variable to ET estimation, that can be derived from remote sensing measurements. In non–energy–limited conditions, LST is a signature of ET and soil water availability. It is available at different spatial resolutions, ranging from 90 m to 100 m and 1 km, from thermal sensors such as ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer), Landsat-8 and MODIS (Moderate-Resolution Imaging Spectroradiometer). Most of the operational applications in hydrology and water resources management use thermal resolution data at a kilometric scale (*Stisen et al.*, 2008; *Tang et al.*, 2010). The main reason behind this is data availability, which for the 1 km resolution MODIS data is one or two times per day, whereas the revisit period for ASTER and Landsat is 16 days. This is a rather long time period in comparison with the quick soil moisture changes that take place in irrigated semi–arid regions.

There are two categories of ET models that use LST data: (1) the residual method and (2) the evaporative fraction (EF) method. Residual methods consider ET as the residual term of

the surface energy balance which is the available energy at the surface minus the sensible heat flux. They can be either single–source, or double–source (in which case the sensible heat flux is considered as a sum of two components - one corresponding to soil, and one to vegetation). Some notable models that use this approach are TSEB (Two–Source Energy Balance, *Norman et al.* (1995)), SEBS (Surface Energy Balance System, *Su* (2002)), TSTIM (Two Source Time Integrated Model, *Anderson et al.* (1997)). One advantage of single–source residual approaches is that they are computationally timesaving and that they require less ground–based measurements than the dual–source counterpart, which require ancillary meteorological data and other surface measurements, such as surface roughness, vegetation height and structure, atmospheric stability (*Li et al.*, 2009). Single–source algorithms offer good estimates of ET, with a relatively high accuracy over homogeneous areas. However, over partially vegetated areas, dual–source models are often required to separately model the heat and water exchanges between soil, vegetation and atmosphere. The main disadvantage associated to residual methods (both single– or dual–source) is having a sufficiently detailed parameterization of surface soil and vegetation properties and the availability of input data such as ground–based measurements (wind speed, surface roughness, vegetation height, fractional vegetation cover, humidity, albedo, air temperature, *Li et al.* (2009)).

The second method considers ET as the available energy times EF (which, by definition, is the ratio of latent heat to available energy). There are various models based on this method, such as SEBI (Surface Energy Balance Index, *Menenti et al.* (1989)), S-SEBI (Simplified Surface Energy Balance Index, *Roerink et al.* (2000)), SEB-1S (Monosource Surface Energy Balance model, *Merlin* (2013)). The approach consists in determining the dry and wet edges of LST, which, depending on the observation resolution, may or may not be present within the observation scene. This has lead scientists to often refer to this approach as "contextual". More about the contextual approach can be found in (*Kalma et al.*, 2008; *Li et al.*, 2009), who offer a summary of various existing contextual models. The physical borderlines of LST are determined as follows: the extreme temperatures (T_{ext}) corresponding to fully dry and fully wet conditions for both soil and vegetation are established in the space defined by LST and f_{vg} (*Long and Singh*, 2012; *Tang et al.*, 2010; *Merlin et al.*, 2013a) and/or the space defined by LST and α (*Roerink et al.*, 2000; *Merlin et al.*, 2013a), where α is the surface albedo. The EF of a given data point located in the $LST - f_{vg}$ or in the $LST - \alpha$ space is calculated as the ratio of two distances: the distance separating the point from the line identified as the dry edge to the distance separating the dry edge and the wet edge. From an operational point of view, contextual methods are very alluring thanks to their image–based nature, making them easily transferable to different areas. One of the most important advantages that these models present is that they do not require any calibration, unlike residual methods (*Jiang and Islam*, 1999). The drawbacks, however, are linked to several conditions that these models require: (1) a relative flat area of study, (2) uniform atmospheric conditions, (3) the extreme temperatures be observed within the study domain at the thermal sensor resolution (*Long et al.*, 2011; *Long and Singh*, 2012; *Long et al.*, 2012; *Long and Singh*, 2013; *Timmermans et al.*, 2007; *Yang and Shang*, 2013). To summarize, image–based models used for deriving T_{ext} are well adapted over semi–arid irrigated areas (*Stefan et al.*, 2015), when using high resolution data such that the high heterogeneity is resolved at the observation scale (*de Tomás et al.*, 2014). They are less adequate when using medium to low ($\sim$1 km) resolution thermal data, acquired over less heterogeneous rainfed areas.

However, the main problem with the majority of the models refers to ET partitioning, which is currently highly uncertain and model–dependent: E is estimated to account for 28% to 56% of the total ET globally (*Lawrence et al.*, 2007). Both residual and EF–based approaches are able to provide E estimates over bare soil. However, the retrieval issue is more complex over mixed pixels. In reality, none of the temperature–based methods allows for an explicit decoupling

between E and T over mixed pixels using only LST and vegetation index data. The reason behind this is that there is an equifinality between the E and T rates, and the pixel LST. However, some temperature–based methods provide supplementary assumptions on the modeling of T (e.g. TSEB) in order to address the ET partitioning. Including information on soil temperature (*Moran et al.*, 1994) or on SEE (*Merlin et al.*, 2012a) would be one possibility to remove the equifinality issue. Since the spatial resolution of current thermal sensors ranges between 100 m and 1 km, most pixels contain a fraction of bare soil and a fraction of vegetation. Therefore, any improvement in E monitoring could contribute to further developments regarding the partitioning of ET.

1.3.3 Soil moisture–based approaches

Soil moisture (SM) in the first 5 cm of the surface is one of the main variables controlling evaporation, with many land surface models assimilating SM data in a corresponding surface layer. For instance, the α and β simplified models of E (see Table 1.1) use a parameter to represent the availability of SM for evaporation (*Desborough et al.*, 1996). Soil moisture is an essential hydrologic variable that not only controls evaporation, but also infiltration and runoff. It also plays an important role in energy and carbon exchanges (*Daly and Porporato*, 2005). Compared to *in situ* measurements, SM observations derived from remote sensing can be retrieved at a global scale, in a cost effective manner. Amongst all the electromagnetic spectra that are sensitive to variations in soil moisture, the most encouraging is the microwave domain. The low frequency band it possesses is especially sensitive to the water content present in both the first few centimeters of soil and in vegetation (*Schmugge et al.*, 2002), all the while being independent of atmospheric conditions.

Two approaches based on microwave data can be distinguished: using passive and active sensors, respectively. Active sensors (radars) emit a microwave energy pulse and measure the intensity of the signal re-emitted by the target surface. The intensity of this signal is directly linked to surface conditions, namely soil and vegetation water content and to the surface roughness. Passive sensors (radiometers) measure the microwave radiation that is naturally emitted by surfaces. Radiative transfer models link the energy absorbed by passive sensors to soil water content.

Microwave observations have a resolution ranging from several tens of meters for active sensors to several tens of kilometers for passive sensors. Notably, the high resolution achieved by active sensors (ERS, ALOS, Sentinel–1) is due to the intensity of the signal being much higher than the energy emitted by land surfaces that is measured by radiometers. ESA's Sentinel–1 mission (*Torres et al.*, 2012), launched in 2012, is a C-band SAR (Synthetic Aperture Radar, operating at a centre frequency of 5.405 GHz) that provides data at resolutions of $\sim$ 20 m, with a repeat cycle of 6 days combining both ascending and descending overpasses. The main downside to estimating SM from radar data is that currently, no operational SM product is available at such a high resolution. SM data could be modeled using active remote sensing (*Balenzano et al.*, 2011), but the difficulty resides in modeling the impact of vegetation cover and surface roughness on the backscattered signal (*Satalino et al.*, 2014), in time and over extended areas. Therefore, they often need site-specific calibration (*Zribi et al.*, 2011). These effects are however diminished with resolution. SM can be for example estimated from scatterometers, in particular the C-band ASCAT (Advanced SCATterometer, *Naeimi et al.* (2009)) at a resolution of 50 km.

Nowadays, radiometers are the microwave sensors most widely used at a global scale in order to derive SM. Two types of radiometers can be identified: L–band (functioning at a frequency of 1.4 GHz) and C–band. Notable L–band radiometers include SMOS (Soil Moisture and Ocean Salinity, *Kerr et al.* (2010)) and SMAP (Soil Moisture Active Passive, *Entekhabi et al.* (2010)),

with AMSR–E (Advanced Microwave Scanning Radiometer–EOS, *Njoku et al.* (2003)) a C–band radiometer. Because of the great size required for the antenna in order to have sufficient sensitivity to capture signals, combined with a fixed weight for the payload, the resolution of L–band data is limited to 30–60 km (*Kerr et al.*, 2010; *Njoku and Entekhabi*, 1996; *Schmugge*, 1998). However, this resolution is too coarse for the majority of hydrological and agricultural applications.

For the past ten years, approaches on combining active and passive remote sensing data have been proposed in order to provide high resolution SM (*Zhan et al.*, 2006; *Das et al.*, 2014). NASA's SMAP mission was expected to combine L–band brightness temperatures and L–band backscatter data in order to derive SM estimates (*Entekhabi et al.*, 2010). However, a radar failure occurred several months after its launch in 2015, and the predicted resolution of 9 km (*Das et al.*, 2014) wasn't achieved.

On the other hand, optical (visible/near–infrared/thermal–infrared) sensors have the advantage of providing data at high and medium resolutions. Optical data include LST and information on vegetation cover, which are connected to soil water content (*Fang et al.*, 2013). Even though optical data could be used to derive SM, the main drawback in deriving a retrieval methodology is given by the sensors' sensitivity to meteorological conditions (cloud presence) and vegetation cover. However, a synergy between the low resolution microwave and high resolution optical data (*Zhan et al.*, 2002) can be used in order to derive SM at various spatial scales.

There are mainly two distinct categories of methods based on the synergy between microwave and optical data: purely empirical methods and methods based on the triangle/trapezoid approach. Empirical algorithms use a polynomial fitting between SM, LST and NDVI. The brightness temperature can also be taken into account (*Piles et al.*, 2011) to estimate SM at different resolutions (10 km and 1 km) from SMOS data. The bias is overall reduced at the expense of reduced spatio–temporal correlation between SMOS SM and *in situ* measurements.

Triangle– (*Carlson et al.*, 1994) and trapezoid– (*Moran et al.*, 1994) based approaches are contextual ET models that regroup semi–physical evaporation–based methods. The polynomial function is replaced by a physical model that interprets the SM variability through evaporation (*Merlin et al.*, 2008a,b): a spatial link is created between optical–derived SEE and surface SM. In other words, they link variations in LST (signature of ET) to variations in soil water content and vegetation cover (*Carlson*, 2007; *Petropoulos et al.*, 2009). Compared to empirical algorithms, their main assets reside in their self–calibrating nature and in the fact that the average of estimated SM at high resolution is equal to the SM observations at low resolution.

An example of a trapezoid–based approach is the DISPATCH (DISaggregation based on a Physical and Theoretical scale CHange) algorithm (*Merlin et al.*, 2008b, 2012b, 2013b). It combines optical and microwave data in order to derive SM at high resolution, by using a first order Taylor series expansion of a SEE model. It employs two different SEE models: i) a temperature–based one (in order to estimate SEE at high resolution from optical data) and ii) a linear SM–based model, which connects the temperature–based SEE to SM. It could be a tool to estimate the SEE at multiple scales by combining available SM and LST data.

1.4 Objectives

Looking at all the different methodologies that have been developed since the 1960s to estimate soil evaporation, no consensus has been made with respect to the best approach of modeling this variable at integrated spatial scales. On the other hand, taking into consideration the monitoring capabilities of E–related quantities through remote sensing means, several studies have shown the possibility to combine shortwave–derived vegetation cover, thermal–derived LST and

microwave–derived SM in order to partition ET into soil evaporation and plant transpiration and therefore successfully retrieve soil evaporation. Unfortunately, given the different spatial resolution of SM and LST products, there is currently no quasi–instantaneous E model that consistently combines these data.

The objective of this PhD thesis is to improve the spatio–temporal representation of soil evaporation by deriving a model that is valid over a wide range of soils and atmospheric conditions, and that can be coupled with readily available remote sensing data. This is a multi–scale modeling approach based on a synergy between remote sensing data and existing phenomenological models. Three sites are used to validate the modeling approach developed in this thesis, all representative of semi–arid conditions, located in Mexico, Spain and Morocco.

In a first step, improvements to the temperature–based contextual modeling approaches are sought. Taking into account the limitations that arise when deriving the extreme temperatures corresponding to fully dry and fully wet conditions (which may or may not be present at the observation resolution) from satellite images, coupled with the necessity of a flat study area and uniform atmospheric conditions, a new model is developed, that seeks to limit the impact of these factors. The model computes the extreme temperatures independently of the spatial resolution of LST data, by using an energy balance model forced by available meteorological data. A validation in terms of extreme temperatures is performed in Morocco, Mexico and Spain. The impact of the two different approaches for deriving (image– or model–based) extreme temperatures is then assessed in term of ET estimates over the Mexican site.

Since in the first step efforts have been made to improve ET estimates from a spatial point of view, in a second step, attempts have been made to characterize the temporal behavior of soil evaporation (through the soil evaporative efficiency) for different soil types. Characterizing its dynamics can be a difficult task. A new approach is used to do so, using the global modeling technique (*Mangiarotti et al.*, 2012a), which relies on the theory of nonlinear dynamical systems. Instead of analytically solving differential or discrete equations, it uses the phase space in order to represent, analyze and understand the temporal evolution of the observed behaviors. The technique is applied for the first time to synthetic SEE series and it seeks to reproduce the daily cycle of the SEE.

In a third step, the energy balance model developed at the first step has been introduced in DISPATCH, and thus a new algorithm, DISPATCH–E, has been developed. It seeks to characterize both the temporal and spatial behavior of the SEE, by combining both the LST– and SM–based approaches, validating the technique through disaggregated SM estimates. In this part, an indirect validation of the improvement of SEE models is performed in terms of DISPATCH/DISPATCH–E outputs: downscaled SM products are compared against *in situ* measurements over the Spanish site.

The following manuscript is structured in 6 parts:

- Chapter 2, that offers a detailed description of the three sites, *in situ* measurements and remote sensing data
- Chapter 3, that presents the modifications made to a temperature–based model of evapotranspiration and validation results in terms of ET estimates
- Chapter 4, that presents the global modeling technique applied to a soil evaporative efficiency series, in an attempt to characterize its temporal behavior
- Chapter 5, that presents an improved evaporation–based disaggregation scheme of soil moisture data and validation results in terms of disaggregated SM products
- Chapter 6, that offers general conclusions and future perspectives

- Annex, where additional papers to which I have contributed as co–author can be found

This PhD thesis plays a central role in the ANR JC MIXMOD–E project, project which started in 2013 and will be finishing in 2017. The goal of this project is to improve the spatio–temporal representation of E at multiple scales by deriving a new model with robust regionalization capabilities from readily available data, while also developing the first multi–resolution remote sensing algorithm dedicated to E monitoring. It aims at developing a multi–scale modeling approach based on the synergy between remote sensing, mechanistic and mathematical approaches, using station–based or field campaign measurements and multi–sensor remote sensing data.

An important collaboration to be mentioned is with the REC (Root zone soil moisture Estimates at the daily and agricultural parcel scales for Crop irrigation management and water use impact - a multi–sensor remote sensing approach) project. It is a H2020 European project, which started in 2015 and will be finishing in 2019. It is a collaboration between three institutions: CESBIO (Centre d'Etudes Spatiales de la Biosphère) Toulouse, isardSAT Barcelona and Université Cadi Ayyad Marrakech. The project aims at developing a new innovative operational algorithm that would allow i) to obtain daily estimates of root zone soil moisture at the crop field scale and ii) to quantitatively evaluate the different components of the water budget at the field scale from readily available multi–sensor remote sensing data. The project will combine active SAR data from Sentinel–1 with DISPATCH disaggregated SM data in order to derive SM estimates at high temporal and spatial resolutions. My main role within this project is with improving the DISPATCH algorithm. I have spent a total of 11 months at isardSAT in Barcelona and 1 month at the Cadi Ayyad University in Marrakech as part of a one year secondment for this project.

INTRODUCTION (Français)

1.5 Contexte

Dans le contexte actuel de croissance de la population et de développement de l'économie, de plus en plus des ressources en eau sont utilisées par les agriculteurs afin d'irriguer leurs terres, mais aussi par différentes industries, augmentant toujours la demande en énergie. Des préoccupations se posent avec les ressources limitées en eau, compte tenu des circonstances de la croissance démographique, du changement climatique et de l'augmentation des problèmes environnementaux. À l'échelle mondiale, environ 10% de l'eau douce disponible est nécessaire chaque an afin d'assouvir les besoins de la population (*Oki and Kanae*, 2006). Cela peut paraître faible en terme de pourcentage, mais il faut garder à l'esprit qu'il existe des zones où la pénurie d'eau est un problème réel, affectant plus de deux milliards de personnes dans le monde entier. Cela implique un réel problème, lorsque l'on considère que, pour répondre aux besoins alimentaires croissants (liés à une population croissante), une augmentation de 70 à 100% de la production agricole est nécessaire au cours des trois prochaines décennies, selon plusieurs organisations environnementales. L'agriculture est donc une contrainte notable sur les ressources en eau, vu que la relation entre la disponibilité de l'eau et la production alimentaire est cruciale (*FAO*, 2009). 70% de l'eau totale est utilisée pour l'agriculture, dont 40% de l'offre totale de nourriture produite par l'agriculture irriguée.

L'importance des ressources en eau dans le secteur agricole est d'un intérêt autant plus important pour les pays méditerranéens, où jusqu'à 80% de l'eau totale est utilisée pour l'irrigation (*FAO*, 2015; *Garrido et al.*, 2010). Ces zones sont aussi les plus sujettes au changement climatique (*IPCC*, 2013; *Stocker et al.*, 2014), tout en ayant seulement 1% des ressources mondiales en eau douce. Avec les prévisions de précipitations annuelles au cours des prochaines décennies étant largement incertaines en termes de magnitude et de la distribution spatiale (*Solomon*, 2007), le IPCC prévoit également que toute modification des régimes des précipitations peut diminuer la productivité de l'agriculture irriguée de jusqu'à 50% à la fin de 2020 (*Bates et al.*, 2008). L'utilisation durable de l'eau est une question cruciale, comme mentionné par Roadmap to a Resource Efficient Europe (*Comission*, 2011), dans le cadre de la stratégie de l'UE 2020. L'utilisation efficace de l'eau dans l'agriculture est fortement liée à la pénurie d'eau et à la sécheresse, tel que rapportée par EEA Report No. 1/2012.

Ainsi, la gestion de l'irrigation des fermes doit être optimisée, dans ce contexte où la demande est croissante mais les ressources diminuent. Du fait de l'inadéquation observée entre les besoins d'irrigation et la quantité réelle d'eau utilisée pour l'irrigation (*Kharrou et al.*, 2013), l'optimisation doit être obtenue en ajustant l'irrigation aux besoins en eau des cultures pendant la saison de croissance des cultures. La sur–irrigation doit être évitée car elle entraîne une augmentation des polluants exportés, combinée à une augmentation inutile de l'utilisation de l'eau. D'autre part, la sous-irrigation entraîne une chute du rendement des cultures (*Salvador et al.*, 2011). Une augmentation du rendement de la production agricole pourrait être atteinte par une meilleure gestion de l'irrigation et de l'utilisation de la fertilisation (*Mueller et al.*, 2012).

Dans ce contexte, l'hydrologie joue un rôle majeur, en tant que discipline qui cherche à comprendre les effets de la nature et les changements crées par l'homme apportées sur des variables telles que l'humidité du sol, la précipitation, le drainage, le ruissellement. Puisque l'eau est une ressource renouvelable dans le cycle hydrologique, une meilleure compréhension de tous les flux qui interviennent dans le cycle hydrologique pourrait améliorer les capacités de surveillance et de fournir un aperçu pour une meilleure gestion.

La surveillance des ressources en eau est donc fortement liée à la modélisation des flux hydrologiques telles que: l'infiltration dans le sol, le ruissellement et l'évapotranspiration. Actuellement, une multitude d'approches de modélisation à grande échelle existent, allant du plus simple au plus complexe, qui prennent en compte d'autres facteurs tels que la végétation. Historiquement, la première modélisation à l'échelle mondiale des flux hydrologiques est basée sur un modèle simple, connu sous le nom de modèle "bucket" (*Manabe*, 1969). Cette approche considère la surface comme un réservoir simple, où le ruissellement et l'évapotranspiration prennent place au–dessus d'une certaine valeur de seuil, estimée à partir de l'humidité du sol présente. D'autre modèles futurs (*Deardorff*, 1978; *Mahrt and Pan*, 1984; *Sellers et al.*, 1986; *Noilhan and Planton*, 1989) ont pris en considération la présence de la végétation afin de tenir compte séparément des flux du sol et de la végétation. Enfin, d'autres approches ont considéré le sol comme un système multi–couche (*Chanzy and Bruckler*, 1993) et modélisent ainsi les flux à partir de l'équation de Richards (*Viterbo and Beljaars*, 1995) et d'une représentation dynamique de la végétation (*Calvet et al.*, 1998).

L'évapotranspiration (ET) est l'un des plus importants flux hydrologiques, ayant une grande importance en particulier dans les régions semi–arides, où elle est de loin le principal flux vers l'extérieur. Au–delà de prévoir la disponibilité en eau du sol tel que mentionné précédemment, elle joue un rôle central dans la prévision des innondations (*Bouilloud et al.*, 2010) et des précipitations (*Findell et al.*, 2011), dans la projection des changements dans la manifestation des sécheresses (*Sheffield and Wood*, 2008) et de canicules (*Seneviratne et al.*, 2006). En particulier, l'évapotranspiration est le seul flux hydrologique reliant à la fois les bilans d'énergie et d'eau. Elle le fait à travers deux variables principales qui sont essentielles à son estimation: la température de surface (la signature de l'ET, qui intervient dans le bilan d'énergie) et l'humidité du sol (une variable d'état qui intervient dans le bilan d'eau). La partition de l'ET dans ses deux principales composantes, l'évaporation du sol (E) et la transpiration des plantes (T) est essentielle dans la modélisation des interactions entre la terre et l'atmosphère et de l'absorption de l'eau par la végétation. Dans les premières étapes de la croissance d'une culture, l'eau est la plupart du temps perdue par l'E; cependant, une fois que la récolte est bien développée et a une couverture complète du sol, T devient le principal processus. Ainsi, l'E est la partie qui ne sert pas à la productivité des cultures (*Wallace*, 2000), alors que T est liée à la disponibilité de l'eau en zone racinaire (*Allen et al.*, 1998). De plus, étant donné que cette partition est actuellement très incertaine et dépendante du modèle utilisé, une connaissance plus approfondie sur ce processus pourrait contribuer à mieux contraindre la physique des modèles de surface (*Gutmann and Small*, 2007). Malheureusement, des mesures *in situ* de l'E et T sont assez rares. Avoir des estimations précises à des échelles multiples (de la culture, de bassins et méso–échelle) et avec une bonne résolution temporelle est sans aucun doute essentiel pour une meilleure surveillance et gestion des ressources (*Williams et al.*, 2004; *Kool et al.*, 2014). La partition de l'ET dans E et T est particulièrement cruciale dans le suivi de la consommation en eau des plantes et du stress hydrique (*Porporato et al.*, 2001; *Er-Raki et al.*, 2010), dans le cadre des ressources en eau limitées. L'évaporation en particulier compte pour 20 à 40 % de l'ET globale (*Lawrence et al.*, 2007; *Schlesinger and Jasechko*, 2014) et elle provient principalement de l'humidité du sol (*Good et al.*, 2015). Comme elle est une condition limite essentielle entre le sol et l'atmosphère, une meilleure représentation de l'évaporation du sol pourrait non seulement aider à une meilleure surveillance

et gestion des ressources en eau, mais aussi à une meilleure compréhension des interactions entre la terre et l'atmosphère (*Feddes et al.*, 2001).

A cet égard, la télédétection présente un intérêt tout particulier, étant rentable dans le suivi et la cartographie de l'E et variables liées à l'E (telles que la couverture de la végétation, la température de surface et l'humidité du sol) sur de grandes surfaces. De plus, la télédétection est un moyen d'estimer l'E à des résolutions multiples. Ceci est important car des échelles différentes sont nécessaires, en fonction du domaine d'application de la variable. Plus particulièrement, le rendement des cultures et la consommation d'eau sont évalués à l'échelle du champ, le ruissellement et les risques d'inondation à l'échelle des bassins versants et la circulation atmosphérique est dictée par les flux de surface à méso–échelle. Dans ce contexte, la télédétection est un moyen de couplage des données disponibles à plusieurs résolutions avec des modèles hydrologiques. D'autre part, une compréhension profonde des processus hydrologiques à différentes échelles pourrait contribuer à améliorer la qualité des données de télédétection disponibles.

Un aperçu des approches existantes pour la modélisation de l'E, des techniques basées sur la télédétection et des liens avec d'autres variables telles que la couverture de la végétation, l'humidité du sol et la température de surface sont discutés ci–dessous.

1.6 Approches de modélisation de l'évaporation du sol

Les approches de modélisation de l'E peuvent être grossièrement classées en trois catégories: les approches mécanistes, des formulations simplifiées et des approches phénoménologiques. Un aperçu de chaque approche est présentée ci–dessous.

1.6.1 Modèles mécanistes

Ces approches considèrent le sol comme un système multi–couches et représentent théoriquement les échanges couplés de masse et chaleur entre la couche de sol et l'atmosphère. Quelques exemples de modèles mécanistes peuvent être trouvés dans *Chanzy and Bruckler* (1993), *Bittelli et al.* (2008). Les flux de vapeur à l'interface entre le sol et l'atmosphère peuvent être estimés en résolvant l'équation couplée de masse et chaleur pour toutes les couches. L'évaporation du sol nu peut alors être calculée à une hauteur donnée par la réduction du flux de vapeur de la couche de sol supérieure à l'aide d'une résistance aérodynamique pour le transfert de vapeur d'eau. Cependant, la résolution de ces équations exige des connaissances sur les propriétés hydrauliques du sol sur chaque couche de sol. Ces propriétés hydrauliques du sol se composent de:

- la conductivité hydraulique saturée ($kg{\cdot}s{\cdot}m^{-3}$)
- le contenu d'humidité résiduelle ($m^3{\cdot}m^{-3}$)
- le contenu d'humidité saturée ($m^3{\cdot}m^{-3}$)
- divers paramètres empiriques de forme (*Brooks and Corey*, 1964; *Campbell*, 1974; *Van Genuchten*, 1980)

Les principales limites des modèles mécanistes basés sur les propriétés hydrauliques du sol sont leur sensibilité aux conditions initiales et limites comme indiqué dans *Chanzy et al.* (2008). Les conditions initiales se réfèrent notamment au profil de l'humidité du sol, qui ne peut être mesuré sur de grandes surfaces à l'aide de la télédétection, puisque les données de télédétection sont principalement prévues pour une couche de surface seulement. Les conditions aux limites portent sur les conditions climatiques à la surface du sol et le changement de teneur en eau ou le

potentiel de l'eau avec la profondeur. En outre, le temps de calcul de ces modèles est assez long, et la solution numérique peut être parfois mal conditionnée, et par conséquent instable.

Les inconvénients de l'utilisation des paramètres hydrauliques du sol résident en leur difficulté à être mesurés, le cout des techniques à utiliser, et surtout, en leur inadaptabilité à être appliqués à des échelles spatiales intégrées. Pour supprimer ces limitations, un contournement a été introduit: les fonctions de pédotransfert. Les fonctions de pédotransfert (*Wösten*, 1997; *Cosby et al.*, 1984; *Rawls and Brackensiek*, 1985; *Vereecken et al.*, 1989; *Twarakavi et al.*, 2009) traduisent les données de texture du sol facilement disponibles dans des propriétés hydrauliques du sol.

Le principal avantage des modèles mécanistes est leur contribution à une meilleure compréhension des processus physiques liés à l'E, tels que la diffusion de vapeur, la remontée capillaire et le drainage. Cependant, ils présentent un certain nombre d'inconvénients qui limitent leur application sur de grandes surfaces. Les fonctions de pédotransfert sont très incertaines et ont des grandes variations dans l'espace (*Soet and Stricker*, 2003). Ceci est principalement dû à la structure du sol (taille des pores, distribution, connectivité), à la présence d'agrégats et à la présence de la biomasse dans le sol. Ces fonctions de pédotransfert varient plus dans une classe de texture qu'entre classes (*Gutmann and Small*, 2007). En outre, un grand nombre de paramètres qui ne peuvent pas être mesurés directement à l'échelle de l'application (*Demaria et al.*, 2007) sont pris en compte dans les modèles mécanistes, qui conduit souvent à une sur–paramétrisation. Une résistance du sol empirique supplémentaire peut parfois être nécessaire afin d'estimer avec précision l'E (*Bittelli et al.*, 2008).

Pour résumer, la grande incertitude des fonctions de pédotransfert, la sensibilité aux conditions initiales et limites, et la sur–paramétrisation rendent impraticables les modèles mécanistes pour les applications à grande échelle. Par conséquent, des modèles plus simples ont été développés.

1.6.2 Modèles simplifiés

En raison des difficultés et des contraintes qui viennent avec l'implémentation des modèles mécanistes sur de grandes surfaces, une autre approche a été développée, qui estime l'E en fonction de l'humidité du sol. Ces modèles simplifiés sont les approches plus largement utilisées pour estimer l'E, avec le module de l'E de la plupart des modèles de surface en se basant sur ces approches.

Le premier modèle, également connu sous le nom de modèle "bucket" (*Manabe*, 1969; *Robock et al.*, 1995), calcule l'évaporation en utilisant l'équation suivante:

$$LE_s = SEE * LE_p \tag{1.8}$$

où LE_s ($\mathrm{W{\cdot}m^{-2}}$) est le flux de chaleur latente, LE_p ($\mathrm{W{\cdot}m^{-2}}$) est le flux de chaleur potentielle, et SEE (-) l'efficacité évaporative du sol (Soil Evaporative Efficiency). SEE est définie comme étant le rapport entre l'évaporation réelle et celle potentielle. Dans le modèle "bucket" , elle est exprimée comme:

$$SEE = \begin{cases} 1, & \text{if } SM \geqslant SM_c \\ \frac{SM}{SM_c}, & \text{if } SM < SM_c \end{cases} \tag{1.9}$$

où SM est l'humidité du sol, et SM_c est 75% de SM_{fc}, où SM_{fc} l'humidité du sol à la capacité du champ.

Des améliorations ont été apportées au modèle ci–dessus, en séparant les deux éléments distincts: le sol et la végétation, utilisant ainsi une formulation double source (*Shuttleworth and*

Wallace, 1985). Parmi ces méthodes, quatre approches sont principalement utilisés:

- La méthode basée sur la r_{ss} (la résistance du sol en surface):

$$LE_s = \frac{\rho C_p}{\gamma} \frac{e_{sat}(T_s) - e_a}{r_{ss} + r_{ah}} \tag{1.10}$$

- La méthode α:

$$LE_s = \frac{\rho C_p}{\gamma} \frac{\alpha e_{sat}(T_s) - e_a}{r_{ah}} \tag{1.11}$$

- La méthode β:

$$LE_s = \beta \frac{\rho C_p}{\gamma} \frac{e_{sat}(T_s) - e_a}{r_{ah}} \tag{1.12}$$

- La méthode du "seuillage":

$$LE_s = min(LE_p, LE_{max}) \tag{1.13}$$

avec r_{ss} ($\mathrm{s \cdot m^{-1}}$) la résistance à la diffusion des vapeurs dans les grands pores du sol, α (-) un facteur de mise à l'échelle sans unités de la pression de vapeur saturante jusqu'à la pression de vapeur réelle à la surface du sol, β (-) un facteur de mise à l'échelle de l'évaporation potentielle jusqu'à l'évaporation réelle, ρ ($\mathrm{kg \cdot m^{-3}}$) la densité de l'air, C_p ($\mathrm{J \cdot kg^{-1} \cdot K^{-1}}$) la chaleur spécifique de l'air à pression constante, γ ($\mathrm{Pa \cdot K^{-1}}$) la constante psychrométrique, $e_{sat}(T_s)$ (Pa) la pression de vapeur saturée à la température du sol, e_a (Pa) la pression de vapeur de l'air, r_{ah} ($\mathrm{s \cdot m^{-1}}$) la résistance aérodynamique au transfert de chaleur et LE_{max} ($\mathrm{W \cdot m^{-2}}$) le flux d'eau maximum limitée par le sol en dessous de la surface.

Dans la littérature, la méthode du "seuillage" est également connue sous le nom "Priestley–Taylor" ou la méthode "demande et offre". LE_p est calculée en utilisant les méthodes aérodynamiques de Penman ou Priestley–Taylor. Diverses études dans la littérature sont basées sur ces modèles (*Mahfouf and Noilhan*, 1991; *Lee and Pielke*, 1992; *Ye and Pielke*, 1993; *Mihailovic et al.*, 1995; *Dekic et al.*, 1995; *Cahill et al.*, 1999). Il y a deux façons de dériver les paramètres r_{ss}, α, β et LE_{max}: en utilisant une approche physique, ou une approche empirique. Dans l'approche physique, les expressions sont obtenues soit en dérivant la loi de diffusion de Fick (*Dickinson et al.*, 1986; *Wetzel and Chang*, 1988; *Sakaguchi and Zeng*, 2009), soit en utilisant des considérations thermodynamiques (*Philip and de Vries*, 1957). Des paramètres empiriques supplémentaires sont introduits en plus des paramètres hydrauliques du sol, obtenus grâce après simplification de l'équation théorique de diffusion. Dans les approches empiriques, des expressions ad–hoc (*Manabe*, 1969; *Noilhan and Planton*, 1989) sont introduites ou des ajustements aux courbes sont faits en se basant sur des données expérimentales (*Sellers et al.*, 1992). Même si des études ont favorisé l'apparition de nombreuses formulations, aucun accord général sur la meilleure façon de paramétrer le processus d'évaporation à des échelles spatialement étendues a été atteint (*Desborough et al.*, 1996; *Sakaguchi and Zeng*, 2009). Il y a cependant trois aspects qui sont généralement convenus: i) des meilleurs estimation de l'humidité du sol et de l'E peuvent être obtenues en utilisant des approches empiriques simplifiées plutôt que des approches physiques (*Dekic et al.*, 1995; *Mihailovic et al.*, 1995; *Yang et al.*, 1998), ii) la formulation β semble être plus robuste que la formulation α (*Cahill et al.*, 1999; *Van den Hurk et al.*, 2000) et iii), même si toutes les approches sont basées sur l'humidité du sol en surface, les effets de l'état de surface du sol (par exemple: la rugosité du sol, présence de chaume ou de paillis , les fissures de retrait associés à des sols argileux, hétérogénéité du sol) n'ont pas été complètement étudiées (*Sakaguchi and Zeng*, 2009).

La plupart des modules d'évaporation implémentés dans les modèles de surface existants sont basées sur l'une des quatre formulations présentées ci–dessus. Les informations concernant les modules simplifiés de l'E présents dans l'état d'art des modèles de surface sont présentées dans le Tableau 1.2.

Table 1.2: Une liste des modules simplifiés de l'E utilisés par certains modèles de surface les plus connus (Merlin 2016, *Mémoire HDR*).

Caractère	**Nom**	**Module d'évaporation**	**Formulation**	**Description du modèle**
Empirique	BUCKET	*Manabe* (1969)	β	*Robock et al.* (1995)
Empirique	CABLE	*Wang et al.* (2011)	seuillage	*Kowalczyk et al.* (November 2006)
Empirique	CLASS V3	*Lee and Pielke* (1992)	β	*Verseghy* (1996)
Empirique	CLSM	*Koster and Suarez* (1996)	r_{ss}	*Koster et al.* (2000)
Empirique	HTESSEL	*Balsamo et al.* (2011)	r_{ss}	ECMWF 2006
Empirique	IAP94	*Sellers et al.* (1992)	r_{ss}	*Yongjiu and Qingcun* (1997)
Empirique	ICARE	*Sellers et al.* (1992)	r_{ss}	*Merlin et al.* (2006a)
Empirique	ISBA	*Noilhan and Planton* (1989)	α	*Noilhan and Mahfouf* (1996)
Empirique	LaD GFDL	*Manabe* (1969)	β	*Milly and Shmakin* (2002)
Empirique	LAPS	*Sun* (1982)	α,r_{ss}	*Mihailović* (1996)
Empirique	MOSAIC	*Koster and Suarez* (1996)	r_{ss}	*Koster and Suarez* (1996)
Empirique	Noah LSM	*Sellers et al.* (1992)	α,r_{ss}	*Niu et al.* (2011)
Empirique	SECHIBA	*Ducoudré et al.* (1993)	r_{ss}	*Ducoudré et al.* (1993)
Empirique	SEWAB	*Noilhan and Planton* (1989)	α	*Mengelkamp et al.* (1999)
Empirique	SiB2	*Sellers et al.* (1992)	α,r_{ss}	*Sellers et al.* (1996)
Empirique	SSIB	*Camillo and Gurney* (1986)	α,r_{ss}	*Xue et al.* (1991)
Empirique	SWB	*Manabe* (1969)	β	*Schaake et al.* (1996)
Empirique	VIC	*Wood et al.* (1992)	β	*Wood et al.* (1992)
Physique	BASE	*Dickinson et al.* (1986)	seuillage	*Desborough and Pitman* (1998)
Physique	BATS	*Dickinson et al.* (1986)	seuillage	*Dickinson et al.* (1993)
Physique	BEST	*Dickinson et al.* (1986)	seuillage	*Pitman et al.* (1991)
Physique	CAPS	*Mahrt and Pan* (1984)	seuillage	*Chang et al.* (1999)
Physique	CLM	*Sakaguchi and Zeng* (2009)	α,r_{ss}	*Oleson et al.* (2010)
Physique	GISS	*Gardner and Hillel* (1962)	seuillage	*Rosenzweig and Abramopoulos* (1996)
Physique	MIT	*Entekhabi and Eagleson* (1989)	β	*Entekhabi and Eagleson* (1989)
Physique	PLACE	*Wetzel and Chang* (1988)	seuillage	*Wetzel and Boone* (1995)
Physique	SEtHyS	*Soarès et al.* (1988)	seuillage	*Coudert et al.* (2006)
Physique	SiSPAT	*Philip and de Vries* (1957)	seuillage	*Braud et al.* (1995)
Physique	SWAP	*Gusev and Nasonova* (1998)	β	*Gusev and Nasonova* (1998)
Physique	TOPLATS	*Milly* (1986)	seuillage	*Famiglietti* (1994)

1.6.3 Modèles phénoménologiques

Les modèles phénoménologiques se distinguent par le fait qu'ils ne sont ni construits sur des hypothèses ad–hoc , ni dérivés de la théorie. Au lieu de cela, ils sont basés sur l'observation des données (*Sivapalan et al.*, 2003), en offrant une interprétation physique ou semi–physique des paramètres du modèle.

En modélisant la SEE au lieu de l'évaporation du sol, on sépare les deux principaux facteurs qui contrôlent l'E: la demande évaporative LE_p et la disponibilité en eau du sol. Ceci est seulement partiellement satisfait, puisque la LE_p est liée à la disponibilité en eau du sol (*Merlin et al.*, 2011). Cependant, cette normalisation de LE_s supprime les effets de premier ordre de la demande évaporative sur l'évaporation réelle, tout en fixant les limites physiques entre $\sim$ 0 et 1. SEE atteint la valeur $\sim$0 lorsque la disponibilité en eau du sol est insignifiante, et la valeur 1 lorsque la disponibilité en eau du sol est maximale.

La SEE aide aussi à distinguer les différents cycles d'évaporation, qui sont liés soit à des variations quotidiennes/saisonnières/climatiques de la demande évaporative, ou à des variations de la disponibilité en eau du sol en raison de la pluie/irrigation.

Une première tentative d'associer un paramètre expérimental de la SEE à la texture du sol et aux conditions aérodynamiques a été faite par Komatsu dans (*Komatsu*, 2003):

$$SEE = 1 - exp(\frac{SM}{SM_c}) \tag{1.14}$$

avec SM_c étant un paramètre du sol dont les valeurs dépendent du type de sol et de la vitesse du vent. La profondeur de la couche de surface sur laquelle l'analyse a été réalisée était de plusieurs millimètres, ce qui est beaucoup plus petit que l'épaisseur normale de la couche de surface supérieure utilisée dans la plupart des modèles de surface, qui est de plusieurs centimètres. Trouver une façon générale de paramétrer la SEE est pour la plupart du temps entravée par le séchage (vers midi) de premières mm du sol, qui empêche l'E, indépendamment de la disponibilité en eau du sol qui est présente sous la surface (*Mahrt and Pan*, 1984; *Dickinson et al.*, 1986; *Soarès et al.*, 1988; *Wetzel and Chang*, 1988; *Van de Griend and Owe*, 1994; *Heitman et al.*, 2008; *Shahraeeni et al.*, 2012). Par conséquent, une nouvelle méthode a été développé afin de dériver la SEE avec une forme qui s'ajuste à la profondeur à laquelle des mesures d'humidité sont effectuées. *Merlin et al.* (2011) se sont penchés sur la possibilité de prendre en considération le gradient d'humidité du sol dans la couche supérieure en introduisant une paramétrisation simple qui dépend de l'évaporation potentielle. Le modèle a été testé sur deux sites situés dans le sud de la France, à une échelle de temps journalière. En outre, dans leur dernière étude (*Merlin et al.*, 2016), ils ont développé une méta–approche d'analyse pour représenter la SEE pour une large gamme de sols et de conditions atmosphériques, en utilisant des données recueillies dans plus de 30 sites dans le monde. Une relation significative a été observée entre les paramètres semi–empiriques du modèle de l'E et la texture du sol. Le modèle a ensuite été comparé au module d'évaporation de plusieurs modèles de surface, comme *Noilhan and Planton* (1989)), CLM (Community Land Model, *Oleson et al.* (2013)) and H–TESSEL (Hydrology–Tiled ECMWF Scheme for Surface Exchange over Land, *ECMWF* (2014)), et le modèle a été prouvé comme fonctionnant de manière satisfaisante.

Le principal avantage des modèles phénoménologiques est leur capacité à utiliser les données facilement disponibles et leur couplage potentiel avec des données de télédétection ce qui les rend adaptés pour des applications à grande échelle, tout en étant compatible avec les hypothèses théoriques. Cependant, le principal inconvénient est lié à leur caractère semi–empirique, combiné avec l'utilisation des paramètres effectifs, ce qui rend difficile d'estimer des scénarios d'évolution.

1.7 Approches pour l'évaporation du sol basées sur la télédétection

Parmi toutes les variables qui peuvent être dérivées de la télédétection, il y en a trois qui sont pertinentes pour l'estimation de l'E: l'indice de végétation, la température de surface et l'humidité du sol en surface. Les méthodes basées sur chacune de ces variables sont décrites ci–dessous.

1.7.1 Approches basées sur l'indice de la végétation

En complément des modèles mécanistes et physiques, il existe une vaste gamme de méthodes pour estimer l'E en utilisant des données de télédétection. Une des méthodes les plus populaires est l'approche FAO–56 (*Allen et al.*, 1998; *Allen*, 2000; *Allen et al.*, 2005; *Simonneaux et al.*,

2008), forcée par des données de NDVI (Normalized Difference Vegetation Index) (*Bausch and Neale*, 1989). L'approche utilise des mesures météorologiques, une phénologie des cultures qui a été calculé à partir du fractionnaire du couvert végétal vert (f_{vg}) dérivé de NDVI , et un coefficient spécifique au cultures afin d'en tirer une ET potentielle. L'ET réelle est alors estimée en fonction de cette ET potentielle. Dans sa forme en double coefficient, l'ET est limitée par l'humidité du sol, sur la base d'un modèle du bilan d'eau du sol qui nécessite la connaissance *a priori* des volumes d'irrigation et les dates. Donc, même si un module de sol peut être intégré afin de séparer E et T, et d'exprimer le stress de la végétation en fonction de la disponibilité en eau du sol, la principale méthode est toujours basée sur les hypothèses suivantes (*Er-Raki et al.*, 2010):

- les volumes d'irrigation et les données de précipitations sont censés être connus;
- T est considérée soit comme sa valeur au taux potentiel, soit elle est calculée via l'humidité du sol en zone racinaire (estimée en utilisant un modèle de sol contraint par des données d'irrigation disponibles);
- E est soit fixée à une valeur constante, soit elle est calculée par l'humidité du sol près de la surface (estimée en utilisant un modèle de sol contraint par des données d'irrigation disponibles);
- le drainage est souvent négligé dans les couches profondes du sol.

Le compromis entre le paramétrage complexe du modèle et la disponibilité des données d'entrée a rendu cette méthodologie très intéressante d'un point de vue opérationnel. L'inconvénient de la méthode FAO–56 est son applicabilité sur de très grandes surfaces, puisque les coefficients qu'elle utilise sont calibrés à partir de mesures sur le terrain, qui sont principalement impactés d'un point de vue temporel et d'hétérogénéité spatiale (*Long and Singh*, 2010).

1.7.2 Approches basées sur la température

La température de surface (Land Surface Temperature - LST) est une variable pertinente pour l'estimation de l'ET, qui peut être obtenue à partir des mesures de télédétection. Dans des conditions illimitées en énergie, la LST est une signature de l'ET et de la disponibilité en eau du sol. Elle est disponible à des résolutions spatiales différentes, allant de 90 m à 100 m et 1 km, à partir de capteurs thermiques tels que ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer), Landsat–8 et MODIS (Moderate-Resolution Imaging Spectroradiometer). La plupart des applications opérationnelles dans la gestion de l'hydrologie et des ressources en eau utilisent des données de résolution thermique à une échelle kilométrique (*Stisen et al.*, 2008; *Tang et al.*, 2010). La principale raison derrière cela est la disponibilité de données. Pour une résolution de 1 km, il est possible d'obtenir les données MODIS une ou deux fois par jour, alors que la période de revisite pour ASTER et Landsat n'est que de 16 jours. Ceci est une période de temps longue en comparaison avec les changements rapides d'humidité du sol qui ont lieu dans les zones irriguées des régions semi–arides.

Il existe deux catégories de modèles de l'ET qui utilisent des données LST: (1) la méthode résiduelle et (2) la méthode basée sur la fraction évaporative (Evaporative Fraction - EF). Les méthodes résiduelles considèrent ET comme le terme résiduel du bilan d'énergie de la surface, qui est l'énergie disponible à la surface moins le flux de chaleur sensible. Elles peuvent être soit source unique, soit double source (dans ce cas, le flux de chaleur sensible est considéré comme une somme de deux composantes - l'une correspondant au le sol, et l'autre à la végétation). Certains modèles notables qui utilisent cette approche sont TSEB (Two–Source Energy Balance,

Norman et al. (1995)), SEBS (Surface Energy Balance System, *Su* (2002)), TSTIM (Two Source Time Integrated Model, *Anderson et al.* (1997)). Un avantage des approches résiduelles source unique est qu'elles sont rapides d'un point de vue informatique et qu'elles nécessitent moins de mesures *in situ* que les approches double source, qui nécessitent des données météorologiques auxiliaires et d'autres mesures de surface, tels que la rugosité de la surface, la hauteur et la structure de la végétation, ainsi que la stabilité atmosphérique (*Li et al.*, 2009). Les algorithmes source unique offrent des bonnes estimations de l'ET, avec une précision relativement élevée sur les zones homogènes. Cependant, sur les zones partiellement couvertes par la végétation, les modèles double source sont souvent nécessaires pour modéliser séparément les échanges de chaleur et d'eau entre le sol, la végétation et l'atmosphère. Le principal inconvénient associé aux méthodes résiduelles (soit source unique, soit double source) est d'avoir une paramétrisation suffisamment détaillée des propriétés de surface des sols et de la végétation, et la disponibilité des données d'entrée comme les mesures *in situ* (la vitesse du vent, la rugosité de surface, l'hauteur de la végétation, la couverture végétale fractionnaire, l'humidité, l'albédo, la température de l'air, *Li et al.* (2009)).

La seconde méthode considère l'ET comme l'énergie disponible fois EF (qui, par définition, est le rapport entre la chaleur latente et l'énergie disponible). Il existe différents modèles basés sur cette méthode, comme SEBI (Surface Energy Balance Index, *Menenti et al.* (1989)), S–SEBI (Simplified Surface Energy Balance Index, *Roerink et al.* (2000)), SEB–1S (Monosource Surface Energy Balance model, *Merlin* (2013)). L'approche consiste à déterminer les bords secs et humides de la LST, qui, en fonction de la résolution de l'observation, peuvent ou non être présents dans la scène d'observation. Cela a amené les scientifiques à se référer à cette approche comme "contextuelle". Des détails sur l'approche contextuelle peuvent être trouvée dans (*Kalma et al.*, 2008; *Li et al.*, 2009), qui offrent un résumé des différents modèles contextuels existants. Les limites physiques de la LST sont déterminées comme suit: les températures extrêmes (T_{ext}) pour des conditions complètement sèches et entièrement humides associées au sol et à la végétation sont établies dans l'espace défini par LST et f_{vg} (*Long and Singh*, 2012; *Tang et al.*, 2010; *Merlin et al.*, 2013a) et/ou l'espace défini par LST et α (*Roerink et al.*, 2000; *Merlin et al.*, 2013a), où α est l'albédo de surface. L'EF d'un point donné situé dans l'espace $LST - f_{vg}$ ou l'espace $LST - \alpha$ est calculé comme le rapport de deux distances: la distance séparant le point de la ligne identifiée comme étant le bord sec et la distance séparant le bord sec et le bord humide. D'un point de vue opérationnel, les méthodes contextuelles sont très séduisantes grâce à leur nature basée sur les images, ce qui les rend facilement transférables à différents domaines. Un des avantages les plus importants que ces modèles présentent est qu'ils ne nécessitent pas de calibration, contrairement aux méthodes résiduelles (*Jiang and Islam*, 1999). Les inconvénients, cependant, sont liés aux plusieurs conditions que ces modèles exigent: (1) une zone d'étude relativement plate, (2) des conditions atmosphériques uniformes, (3) que les températures extrêmes soient observées sur la zone d'étude à la résolution du capteur thermique (*Long et al.*, 2011; *Long and Singh*, 2012; *Long et al.*, 2012; *Long and Singh*, 2013; *Timmermans et al.*, 2007; *Yang and Shang*, 2013). Pour résumer, les modèles basés sur les images utilisés pour dériver T_{ext} sont bien adaptés pour les zones irriguées de régions semi–arides (*Stefan et al.*, 2015), lors de l'utilisation des données de haute résolution pour que la forte hétérogénéité soit résolue à l'échelle d'observation (*de Tomás et al.*, 2014). Ils sont moins adéquats lors de l'utilisation des données thermiques à des moyennes résolutions ($\sim$ 1 km), acquises sur des zones pluviales moins hétérogènes.

Toutefois, le principal problème avec la majorité des modèles se réfère à la partition de l'ET, qui est actuellement très incertaine et dépendante du modèle choisi: l'E est estimée pour tenir compte globalement de 28 % à 56 % de l'ET totale (*Lawrence et al.*, 2007). Les deux approches, résiduelles et basées sur EF, sont en mesure de fournir des estimations de l'E sur le sol nu. Toutefois, la question de récupération est plus complexe sur les pixels mixtes. En

réalité, aucune des méthodes basées sur la température permet un découplage explicite entre E et T sur des pixels mixtes en utilisant uniquement les données LST et l'indice de végétation. La raison derrière cela est qu'il y a une équifinalité entre les taux de l'E et de la T et la LST du pixel. Cependant, certaines méthodes basées sur la température fournissent des hypothèses supplémentaires sur la modélisation de T (par exemple TSEB) afin de répondre à la partition de l'ET. En intégrant des informations sur la température du sol (*Moran et al.*, 1994) ou sur la SEE (*Merlin et al.*, 2012a), il serait possible de supprimer la question de l'équifinalité. Etant donné que la résolution spatiale des capteurs thermiques actuels est comprise entre 100 m et 1 km, la plupart des pixels comprennent une fraction du sol nu et une fraction de la végétation. Par conséquent, toute amélioration du suivi de l'E pourrait contribuer à de nouveaux développements en ce qui concerne la partition de l'ET.

1.7.3 Approches basées sur l'humidité du sol

L'humidité du sol (Soil Moisture - SM) dans les 5 premiers cm de la surface est l'une des principales variables qui contrôlent l'évaporation, avec de nombreux modèles de surface assimilant les données SM dans une couche de surface correspondante. Par exemple, les modèles simplifiés de l'E α et β (voir Tableau 1.2) utilisent un paramètre pour représenter la disponibilité de SM pour l'évaporation (*Desborough et al.*, 1996). L'humidité du sol est une variable hydrologique essentielle qui ne contrôle pas seulement l'évaporation, mais aussi l'infiltration et le ruissellement. Elle joue également un rôle important dans les échanges d'énergie et de carbone (*Daly and Porporato*, 2005). Par rapport aux mesures *in situ*, les observations SM obtenues par télédétection peuvent être récupérées à l'échelle mondiale, d'une manière rentable. Parmi tous les spectres électromagnétiques qui sont sensibles aux variations de l'humidité du sol, le plus encourageant est le domaine des micro–ondes. Sa bande basse fréquence est particulièrement sensible à la teneur en eau présente dans les premiers quelques centimètres du sol et dans la végétation (*Schmugge et al.*, 2002), tout en étant indépendant des conditions atmosphériques.

Deux approches basées sur les données micro–ondes peuvent être distingués: en utilisant des capteurs soit passifs soit actifs. Les capteurs actifs (radars) émettent une impulsion d'énergie micro–onde et mesurent l'intensité du signal ré–émis par la surface cible. L'intensité de ce signal est directement liée à la surface, à savoir à la teneur en eau du sol et de la végétation et à la rugosité de surface. Les capteurs passifs (radiomètres) mesurent le rayonnement des micro–ondes qui est naturellement émis par les surfaces. Les modèles de transfert radiatif relient l'énergie absorbée par les capteurs passifs à la teneur en eau du sol.

Les observations micro–ondes ont une résolution allant de plusieurs dizaines de mètres pour les capteurs actifs à plusieurs dizaines de kilomètres pour les capteurs passifs. Notamment, la haute résolution obtenue par des capteurs actifs (ERS, ALOS, Sentinel–1) est en raison de l'intensité du signal étant beaucoup plus élevé que l'énergie émise par les surfaces terrestres qui est mesurée par les radiomètres. La mission Sentinel–1 de l'ESA (*Torres et al.*, 2012), lancée en 2012, est un SAR en bande C (Synthetic Aperture Radar, fonctionnant à une fréquence centrale de 5.405 GHz) qui fournit des données à des résolutions de $\sim$ 20 m, avec un cycle de répétition de 6 jours combinant à la fois les passages ascendante et descendante. Le principal inconvénient de l'estimation de SM à partir des données radar est qu'à l'heure actuelle, aucun produit SM opérationnel est disponible à une telle haute résolution.

Les données SM peuvent être modélisées en utilisant la télédétection active (*Balenzano et al.*, 2011), mais la difficulté réside dans la modélisation de l'impact de la couverture végétale et de la rugosité de surface sur la rétrodiffusion du signal (*Satalino et al.*, 2014), dans le temps et sur des zones étendues. Par conséquent, elles ont souvent besoin d'étalonnage sur des sites spécifiques (*Zribi et al.*, 2011). Ces effets sont cependant diminués avec la résolution. SM peut

être par exemple estimée à partir de diffusiomètres, en particulier ASCAT en bande C (Advanced SCATterometer, *Naeimi et al.* (2009)) à une résolution de 50 km.

Actuellement, les radiomètres sont les capteurs micro–ondes les plus largement utilisés à l'échelle mondiale afin d'en tirer la SM. Deux types de radiomètres peuvent être identifiés: en bande L (fonctionnant à une fréquence de 1,4 GHz) et en bande C. Quelques radiomètres notables en bande L comprennent SMOS (Soil Moisture and Ocean Salinity, *Kerr et al.* (2010)) et SMAP (Soil Moisture Active Passive, *Entekhabi et al.* (2010)), avec AMSR–E (Advanced Microwave Scanning Radiometer–EOS, *Njoku et al.* (2003)) un radiomètre en bande C. En raison de la grande taille requise par l'antenne afin d'avoir une sensibilité suffisante pour capter les signaux, combinée à un poids fixe pour la charge utile, la résolution de données en bande L est limitée à 30–60 km (*Kerr et al.*, 2010; *Njoku and Entekhabi*, 1996; *Schmugge*, 1998). Cependant, cette résolution est trop grossière pour la majorité des applications hydrologiques et agricoles.

Au cours des dix dernières années, des approches basées sur la combinaison de données de télédétection active et passive ont été proposées dans le but de fournir SM à une haute résolution (*Zhan et al.*, 2006; *Das et al.*, 2014). La mission SMAP de la NASA avait pour projet de combiner les températures de brillance en bande L et les données de rétrodiffusion en bande L afin d'en tirer des estimations de SM (*Entekhabi et al.*, 2010). Cependant, une panne de radar a eu lieu plusieurs mois après son lancement en 2015, et la résolution prévue de 9 km (*Das et al.*, 2014) n'a pas été atteinte.

D'autre part, les capteurs optiques (visible/proche–infrarouge/infrarouge thermique) ont l'avantage de fournir des données à des résolutions élevées et moyennes. Les données optiques comprennent la LST et des informations sur la couverture végétale, qui sont reliées à la teneur en eau du sol (*Fang et al.*, 2013). Même si les données optiques pourraient être utilisées pour dériver la SM, l'inconvénient principal de dériver une méthode de récupération provient de la sensibilité des capteurs aux conditions météorologiques (présence de nuages) et la couverture végétale. Cependant, une synergie entre les données micro–ondes à basse résolution et les données optiques à haute résolution (*Zhan et al.*, 2002) peut être utilisée afin de dériver la SM à différentes échelles spatiales.

Il existe principalement deux catégories distinctes de méthodes basées sur la synergie entre les données micro–ondes et les données optiques: méthodes purement empiriques et méthodes basées sur l'approche du triangle/trapèze. Les algorithmes empiriques utilisent un raccord à partir des polynômes entre la SM, la LST et le NDVI. La température de brillance peut également être prise en compte (*Piles et al.*, 2011) pour estimer SM à différentes résolutions (10 km et 1 km) à partir de données SMOS. Le biais est globalement réduit au détriment de la réduction de la corrélation spatio–temporelle entre le SM SMOS et les mesures *in situ*.

Les approches basées sur le triangle (*Carlson et al.*, 1994) et le trapèze (*Moran et al.*, 1994) sont des modèles contextuels de l'ET qui regroupent les méthodes semi–empiriques basée sur l'évaporation. La fonction polynôme est remplacée par un modèle physique qui interprète la variabilité de la SM par l'évaporation (*Merlin et al.*, 2008a,b): un lien spatial est créé entre la SEE dérivée de l'optique et la SM de surface. En d'autres termes, ils relient les variations de la LST (signature de l'ET) aux variations de la teneur en eau du sol et du couvert végétal (*Carlson*, 2007; *Petropoulos et al.*, 2009). Par rapport à des algorithmes empiriques, leurs principaux atouts résident dans leur auto–calibration et dans le fait que la moyenne de SM estimée à haute résolution est égale aux observations SM à faible résolution.

Un exemple d'une méthode basée sur le trapèze est l'algorithme DISPATCH (DISaggregation based on a Physical and Theoretical scale CHange, *Merlin et al.* (2008b, 2012b, 2013b)). Il combine des données optiques et micro–ondes afin de dériver la SM à haute résolution, en utilisant un développement de Taylor de premier ordre d'un modèle de la SEE. Il emploie deux modèles SEE différents: i) un basé sur la température (afin d'estimer la SEE à haute résolution à partir des

données optiques) et ii) un modèle linéaire de la SM, qui relie la SEE dérivée de la température avec la SM. Il pourrait être un outil pour estimer la SEE à des échelles multiples en combinant les données SM et LST disponibles.

1.8 Objectifs

En regardant les différentes méthodes qui ont été développées depuis les années 1960 pour estimer l'évaporation du sol, aucun consensus n'a été fait par rapport à la meilleure approche de modélisation de cette variable à des échelles spatiales intégrées. D'autre part, en tenant compte des capacités de surveillance des quantités dérivées de l'E par des moyens de télédétection, plusieurs études ont montré la possibilité de combiner le couvert végétal dérivé des courtes–ondes, la LST dérivée du thermique et la SM dérivée de micro–ondes afin de partitionner l'ET dans l'évaporation du sol et la transpiration des plantes afin de récupérer avec succès l'évaporation du sol. Malheureusement, compte tenu de la résolution spatiale différente de la SM et des produits LST, il n'y a actuellement aucun modèle quasi–instantané de l'E qui combine systématiquement ces données.

L'objectif de cette thèse est d'améliorer la représentation spatio–temporelle de l'évaporation du sol en dérivant un modèle qui est valable pour une large gamme de sols et de conditions atmosphériques, et qui peut être couplé avec les données de télédétection facilement disponibles. Ceci est une approche de modélisation multi–échelle basée sur une synergie entre les données de télédétection et des modèles phénoménologiques existants. Trois sites sont utilisés pour valider l'approche de modélisation développée dans cette thèse, tous représentatifs des conditions semi–arides, situés au Mexique, en Espagne et au Maroc.

Dans une première étape, une amélioration des méthodes contextuelles basées sur la température est recherchée. Compte tenu des limitations qui apparaissent lors de la dérivation des températures extrêmes à partir des images satellitaires dans des conditions entièrement sèches et entièrement humides (qu'elles soient ou non présentes à la résolution d'observation), couplé à la nécessité d'avoir une zone d'étude plate et des conditions atmosphériques uniformes, un nouveau modèle est développé, afin de limiter l'impact de ces facteurs. Le modèle calcule les températures extrêmes indépendamment de la résolution spatiale des données LST, en utilisant un modèle de bilan d'énergie forcée par des données météorologiques disponibles. Une validation en termes de températures extrêmes est effectuée au Maroc, au Mexique et en Espagne. L'impact des deux approches différentes (basées sur les images et basées sur le modèle) pour dériver les températures extrêmes est ensuite évalué en terme d'estimations de l'ET sur le site mexicain.

Comme dans un premier temps une étude a été menée afin d'améliorer les estimations de l'ET d'un point de vue spatial, dans une seconde étape, des tentatives ont été faites pour caractériser le comportement temporel de l'évaporation du sol (par l'efficacité évaporative du sol) pour différents types de sols. Caractériser sa dynamique peut être une tâche difficile. Une nouvelle approche est utilisée pour le faire, en utilisant la technique de modélisation globale (*Mangiarotti et al.*, 2012a), qui repose sur la théorie des systèmes dynamiques non linéaires. Au lieu de résoudre analytiquement des équations différentielles ou discrètes, elle utilise l'espace de phase afin de représenter, d'analyser et de comprendre l'évolution temporelle des comportements observés. La technique est appliquée pour la première fois à une série synthétique de la SEE et elle cherche à reproduire le cycle journalier de la SEE.

Dans un troisième temps, le modèle du bilan d'énergie développé à la première étape a été introduit dans DISPATCH, et donc un nouvel algorithme, DISPATCH–E, a été développé. Il cherche à caractériser à la fois le comportement temporel et spatial de la SEE, en combinant les deux approches basées sur la LST et la SM, validant la technique par les estimations de SM

désagrégées. Dans cette partie, une validation indirecte de l'amélioration des modèles de la SEE est réalisée en termes de sorties de DISPATCH/DISPATCH–E: les produits SM désagrégés sont comparés avec des mesures *in situ* sur le site espagnol.

Le manuscrit suivant est structuré en 6 parties:

- Chapitre 2, qui offre une description détaillée des trois sites, des mesures *in situ* et des données de télédétection
- Chapitre 3, qui présente les modifications apportées à un modèle de l'évapotranspiration basé sur la température et des validations des résultats en termes d'estimations de l'ET
- Chapitre 4, qui présente la technique de modélisation globale appliquée à une série d'efficacité d'évaporation du sol, dans le but de caractériser son comportement dans le temps
- Chapitre 5, qui présente une méthode améliorée de la désagrégation des données d'humidité du sol, basée sur l'évaporation, et les résultats de la validation en termes de produits SM désagrégées
- Chapitre 6, qui offre des conclusions générales et perspectives d'avenir
- Annexe, où des articles supplémentaires dont j'ai contribué en tant que co–auteure peuvent être trouvés

Cette thèse de doctorat joue un rôle central dans le projet ANR JC MIXMOD–E, projet qui a commencé en 2013 et sera terminé en 2017. Le but de ce projet est d'améliorer la représentation spatio–temporelle de l'E à des échelles multiples en dérivant un nouveau modèle avec des capacités solides de régionalisation à partir de données facilement disponibles, tout en développant le premier algorithme dédié au suivi de l'E à des résolutions multiples à partir de la télédétection. Il vise à développer une approche de modélisation multi–échelle basée sur la synergie entre la télédétection, les approches mécanistes et mathématiques, en utilisant des mesures issues des stations ou des campagnes de terrain et des données de télédétection multi–capteur.

Une collaboration importante à mentionner a été réalisée avec le projet REC (Root zone soil moisture Estimates at the daily and agricultural parcel scales for Crop irrigation management and water use impact - a multi–sensor remote sensing approach). C'est un projet européen H2020, qui a commencé en 2015 et sera terminé en 2019. C'est une collaboration entre trois institutions: CESBIO (Centre d'Etudes Spatiales de la Biosphère) Toulouse, isardSAT Barcelona et Université Cadi Ayyad de Marrakech. Le projet vise à développer un nouvel algorithme opérationnel innovant qui permettrait i) d'obtenir des estimations journalières de l'humidité du sol en zone racinaire à l'échelle du champ et ii) d'évaluer quantitativement les différentes composantes du bilan d'eau à l'échelle du champ en utilisant des données de télédétection multi–capteur facilement disponibles. Le projet combinera les données actives SAR de Sentinel–1 avec les données désagrégées de DISPATCH afin de dériver la SM à des résolutions temporelles et spatiales élevées. Mon rôle principal au sein de ce projet est d'améliorer l'algorithme de DISPATCH. J'ai passé un total de 11 mois à isardSAT Barcelone et 1 mois à l'Université Cadi Ayyad de Marrakech dans le cadre d'une mission longue durée d'un an pour ce projet.

Chapter 2

DATA

Contents

2.1 Introduction 29
2.2 Sites and *in situ* data description 30
2.2.1 Yaqui site 30
2.2.2 Haouz plain (R3 perimeter) 31
2.2.3 Urgell site 35
2.3 Spatial data 36
2.3.1 Optical data 37
2.3.2 Microwave data 38
2.3.3 SAFRAN data 39
2.3.4 DEM data 40
2.3.5 Data pre–processing 40
2.4 Conclusion 42

2.1 Introduction

The modeling approaches developed in this thesis have been tested over three sites, reflective of the same climatic conditions: semi–arid areas, where irrigation is also present. Although ET and SM estimates were validated over different sites, the common point between all three sites is reflected by the analysis of extreme soil temperatures. Using three different study sites allowed for an independent assessment of the methodology for computing extreme soil temperatures using an energy balance model over semi–arid areas.

The three areas used as validation sites are:

- the Yaqui valley situated in the Sonora State (north–west Mexico), used for the comparison of extreme soil temperatures issued using either satellite data or the energy balance model. It also serves for the validation of evapotranspiration (ET) estimates;
- the R3 perimeter situated in the Haouz plain (central Morocco), used for the validation of extreme soil temperatures estimated using the energy balance model with *in situ* measurements;
- the Urgell region located in Catalonia (north–east Spain), used for the comparison of extreme soil temperatures estimated using satellite data or the energy balance model.

It also serves for the validation of disaggregated soil moisture (SM) products with *in situ* measurements.

Different data have been used as an input to the different models present in this study. In particular, remote sensing data have been used to i) derive ET estimates from a contextual temperature–based model, and ii) disaggregate SMOS SM using a soil evaporative efficiency (SEE) method and a similar contextual ET model. Meteorological data available over three study sites have been used in order to force an energy balance model. *In situ* measurements collected during different field campaigns were used as validation data of either ET, extreme temperatures (T_{ext}), or SM products.

A detailed description of the study sites, along with the remote sensing and meteorological data used (including data pre–processing) is offered below.

2.2 Sites and *in situ* data description

2.2.1 Yaqui site

The Yaqui valley, situated in the Sonora state in north–west Mexico, is a typical semi–arid region, having a mean annual potential evapotranspiration of 2233 mm and a mean annual precipitation of 290 mm (*Chirouze et al.*, 2014). Most of the precipitation (approximately 240 mm) occurs during the monsoon season (June–September). The valley represents the largest agricultural area in Sonora, with a total surface used for cultivation spanning 225 000 ha. It is bounded by the Sierra Madre Mountains in the north–east and by the Sea of Cortez in the south–west. The main crops consist of wheat (50 %), chilli pepper, beans, chickpea, potatoes, broccoli, corn, safflower, and fruit trees (orange). Irrigation represents almost 90 % of the total water consumption, with the Alvaro Obregon Reservoir (located on the Yaqui river) as the main contributor.

2.2.1.1 *In situ* flux data

An experiment was performed, spanning a total of six months (December 2007 till May 2008), on an irrigated area of 4 km by 4 km situated within the Yaqui perimeter. Its goal was to characterize, from a spatial variability point of view, the surface fluxes at different spatial scales, ranging from the hectometric field scale to the kilometric scale, (*Merlin et al.*, 2010a; *Fieuzal et al.*, 2011; *Chirouze et al.*, 2014; *Merlin*, 2013). The flux data used in the evapotranspiration validation analysis were provided by six micro–meteorological stations equipped with an eddy–covariance flux measurement system that were set up in different fields (*Merlin*, 2013). Data from these six stations were provided simultaneously with at least four ASTER overpasses. The latent and sensible heat fluxes were evaluated using KH20 fast response hydrometers at a frequency of 10 Hz, which were afterwards averaged on a 30 minute time period. The net radiation was measured using CNR1 and Q7.1 radiometers, while the ground heat flux was acquired using HUKSEFLUX HFP–01 plates (buried at a 0.05 m depth).

A visual representation of the Yaqui valley is presented in Figure 2.1.

2.2.1.2 Meteorological data

Apart from flux measurements, meteorological data, such as air temperature, relative air humidity, solar radiation and wind speed, were acquired every half hour, by a station located in the center of the region. The acquisitions were made at a 10 m height, from December 27th,

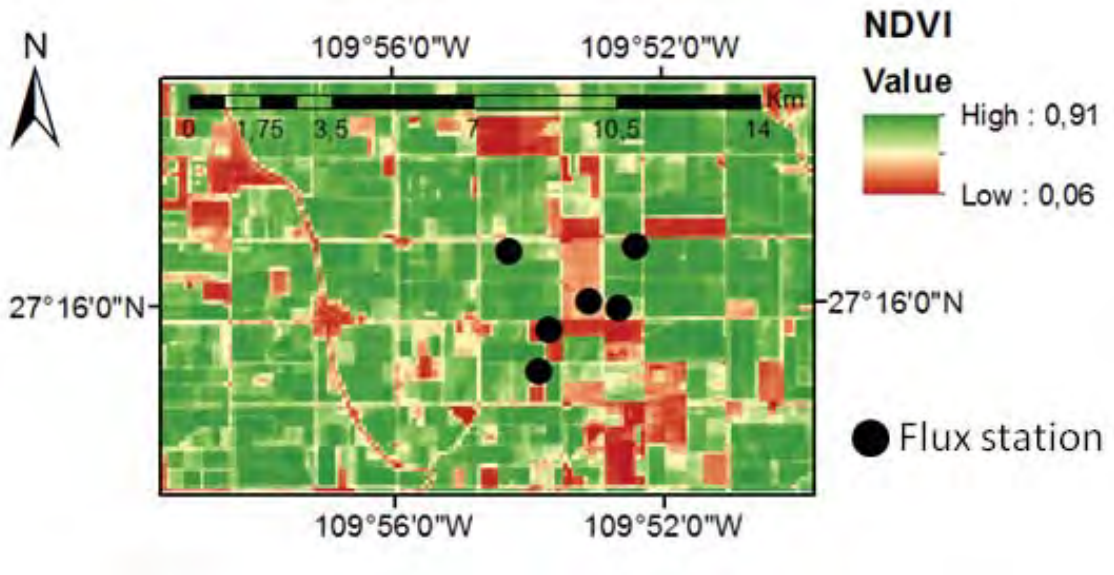


Figure 2.1: The Yaqui site along with the six flux stations situated in the study area.

2007, till May 17th, 2008. Further details on the experiment can be found in *Merlin* (2013), *Chirouze et al.* (2014), *Fieuzal et al.* (2011).

The area of interest is taken as a 16 km by 10 km area encompassing the 4 km by 4 km perimeter in which measurements were taken.

2.2.2 Haouz plain (R3 perimeter)

The Haouz plain is situated in central Morocco (40 km east of Marrakech) and covers about 6000 km^2 of almost flat surface. It presents a typical semi–arid Mediterranean climate, with a mean annual precipitation of 250 mm, ocurring mostly from autumn till spring. The average annual potential evapotranspiration is of 1600 mm. Approximately 85 % of available water is used for agriculture, with flood irrigation being the most widely used method (*Er-Raki et al.*, 2007). The main crops in the R3 irrigated perimeter consist of wheat (generally sown between mid November and mid Januray), with maize, beet and olive trees making up the rest. The typical crop surface is of about 3–4 ha.

2.2.2.1 *In situ* temperature data

A field experiment was conducted in a maize parcel of the R3 perimeter, on the following days in 2014: April 14th, April 22nd, April 30th, May 8th and May 16th. The experiment seeked to perform measurements of the minimum and maximum soil temperatures ($T_{s,wet}$ and $T_{s,dry}$) that would further serve as reference for the direct validation of T_{ext} modeling approaches using meteorological data. Since maize is sown in early April, the crop field was mostly representative of bare soil conditions during the whole experiment.

Various atmospheric conditions were met during the experiment, the weather generally being sunny, with one cloudy day (May 16th) and one post–rainy day (April 22nd). This allowed testing the approach in different conditions.

Temperature measurements were performed over two 2 m by 2 m soil plots. In order to simulate the two extreme (fully dry and fully wet) conditions, one soil plot was kept fully dry, while the soil moisture of the other soil plot was maintained equal to the soil moisture at saturation (*Stefan et al.*, 2015). The dry soil plot was kept dry by isolating it with a plastic sheeting, burried at 15 cm. This was done in order to avoid water infiltrations from capillary rising. In order to provide shelter against potential rainfalls, it was always covered by plasting sheeting between sampling days. The wet soil plot was kept at saturation by irrigating it (continuously and manually) from about 10–20 minutes before the start of the measurements until the time of the last measurement, on each sampling day. In order to prevent flooding and runoff, water was poured in sufficient quantity to reach soil saturation, but as slow as possible (*Stefan et al.*, 2015). An illustration of the 2 m by 2 m dry soil plot where the experiment took place is shown in Figure 2.2.



Figure 2.2: The dry soil plot located in the R3 perimeter. A pit of 15 cm depth was dug **(a)** and was kept fully dry by using a plastic sheeting to cover it in between experiments **(b)**. iButton sensors **(c)** and the handheld infrared thermometer **(d)** used to perform temperature measurements are also illustrated.

Continuous measurements of $T_{s,wet}$ and $T_{s,dry}$ were taken between 10:00 AM and 12:00 AM UTC (Coordinated Universal Time). The observation time was chosen to be compatible with the overpass times of several thermal sensors: MODIS, ASTER, Landsat-7/Landsat-8. The actual soil temperature measurements were taken by using two different means: DS1921G–F5 Ther-

mochron iButton sensors (`http://www.ibuttonlink.com/products/ds1921g`) and OS532E–W8 handheld infrared thermometer.

The iButton sensors are a low–cost, self–sufficient system that keep track of the temperature and store the readings in time. Aiming at having a high confidence in the experimental data, and thus a better estimate of the measurements' uncertainty, several Thermochron iButtons were installed at each soil plot. Basically, iButtons were set up on acrylic glass plates (7 cm by 7 cm), mounting up to three iButtons per plate. Two plates (six iButtons) were used to monitor $T_{s,dry}$, since more variability was found in the dry soil temperature measurements. Only one plate was used to estimate $T_{s,wet}$. All plates were planted very close to the soil surface, without having direct exposure to the Sun. Thermochron data were registered every 5 minutes. Entries logged by all Thermochron iButtons were averaged per soil plot, and per 30 minute bins, which were then cumulated at 10:30 AM, 11:00 AM and 11:30 AM UTC (*Stefan et al.*, 2015). Thus, a total of three datasets (one for $T_{s,wet}$ and two for $T_{s,dry}$) were obtained every measurement day. The two distinct datasets obtained for $T_{s,dry}$ were also averaged in order to have just one set of measurements, as for $T_{s,wet}$.

The infrared thermometer provides non contact temperature measurements with an absolute accuracy of 2–3 °C. It provides temperature reading at a glance, on a digital LCD display, displaying both the current, the minimum and maximum, average or differential temperatures. The range of the measurable target varies between $\sim$ 13 cm and $\sim$ 30 m. Emissivity can be manually adjusted with a 0.01 step. Measurements were taken at a 10 minute time interval, with three different readings for each soil plot, that were consequently averaged to obtain a single set of $T_{s,dry}$ and $T_{s,wet}$ values. A visual representation of the (mean) measurements provided using both sensors (iButton and infrared thermometer) over the dry and wet soil plots on April, 30th, is presented in Figure 2.3.

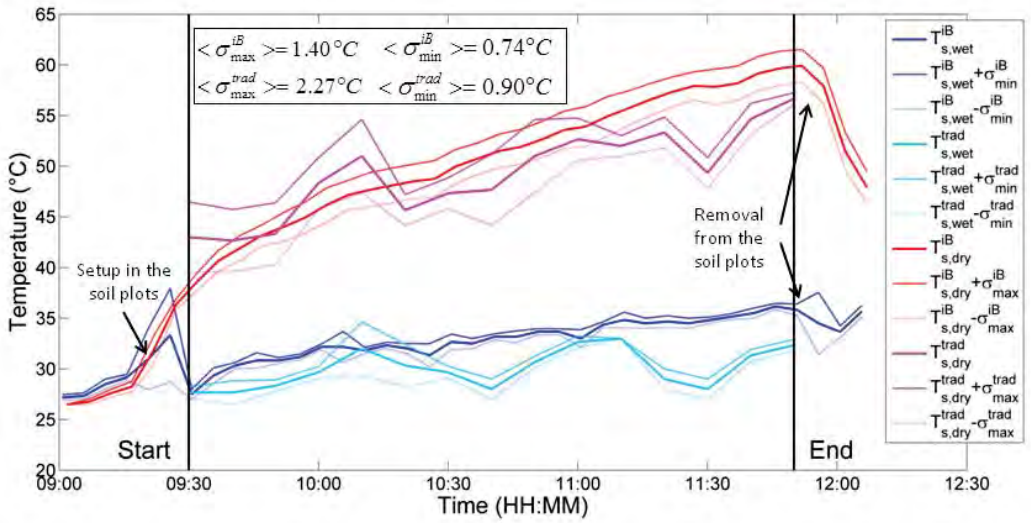


Figure 2.3: Mean and standard deviation of the *in situ* measurements acquired over the wet and dry soil plots in the R3 perimeter (April 30th, 2014). Measurements were acquired using either iButton (here abbreviated as "iB") sensors or the infrared thermometer (here abbreviated as "trad"). Vertical lines mark the begining and ending of the experiment. The average of the standard deviation of the instantaneous measurements is also computed.

By looking at results of this preliminary analysis, one can see that the infrared thermometer

readings are generally underestimated as compared to iButton readings, for both $T_{s,dry}$ and $T_{s,wet}$, which is to be expected since the emisivitty for the thermometer was set to 1. The point is that they present larger variations. This behavior was observed for all the days where the measurements were performed. The good confidence in the iButton sensors' embedded calibration is reflected by the soil T_{ext} measurements matching until 30 minutes before the start of the experiment. The uncertainty in $T_{s,dry}$ and $T_{s,wet}$ was estimated by computing the average on each soil plot of the standard deviation of instantaneous measurements acquired either with the iButton sensors or with the infrared thermometer. The uncertainty for the iButton sensors was found to be 1.40 °C and 0.74 °C for $T_{s,dry}$ and $T_{s,wet}$, respectively. The uncertainty for the infrared thermometer was found to be 2.27 °C and 0.90 °C for $T_{s,dry}$ and $T_{s,wet}$, respectively. Therefore, the temperature readings registered by the iButton sensors were used as a reference when comparing model–derived extreme temperatures with *in situ* measurements, because less uncertainty was found in the iButton readings than in the infrared thermometer readings.

2.2.2.2 Meteorological data

Meteorological data (wind speed, relative air humidity, air temperature, solar radiation) were collected every 30 minutes at a 2 m height by a permanent wheather station located in an alfalfa crop field. The station is located approximately 200 m away from the dry/wet soil plots (see Figure 2.4).

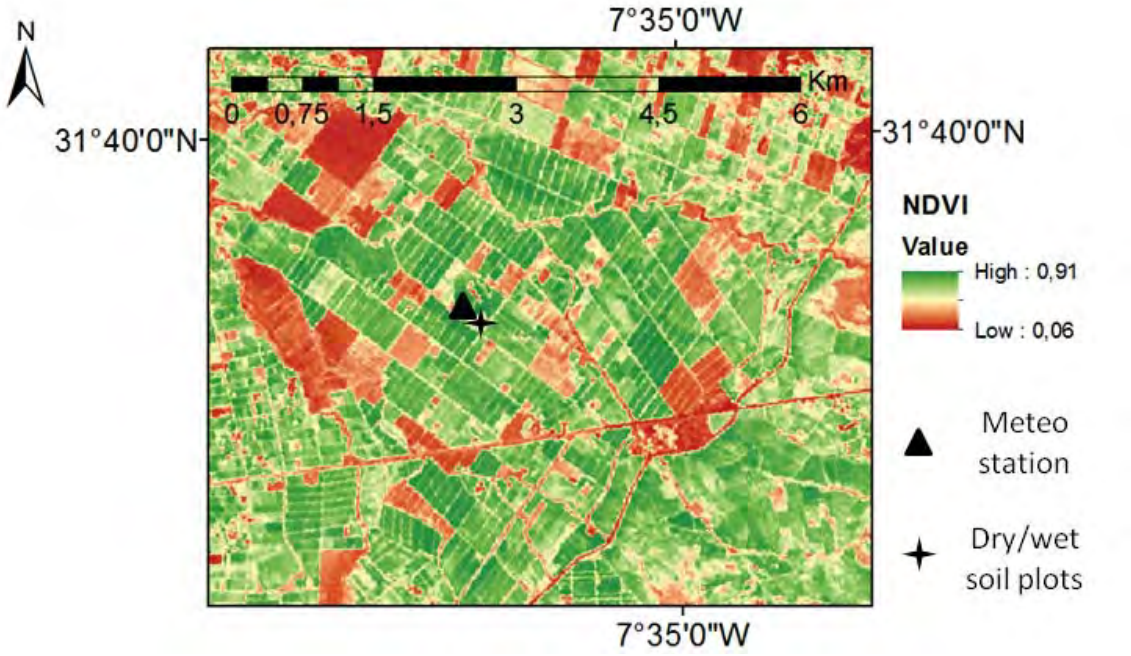


Figure 2.4: The wet and dry soil plots along with the meteorological station located in the R3 perimeter.

2.2.3 Urgell site

The Urgell region in Catalonia (north–east Spain) is a typical semiarid Mediterranean region, with mild winters and warm summers. It presents two rainy seasons (autumn and spring) and a very dry season (summer). The average yearly temperature is 16 °C, while the average yearly precipitation is about 400 mm. The main irrigated crops found within the area are wheat, maize, alfalfa and fruit trees (apple and pear). The dryland crops include barley, vineyards, olive and almond trees.

2.2.3.1 *In situ* soil moisture data

A field experiment was conducted in 2011 and 2012 on a 20 km by 20 km area to gather the 0–5 cm soil moisture using the gravimetric technique. Ten days were sampled in 2011: on day of year (doy) 97, doy 98, doy 146, doy 147, doy 165, doy 196, doy 228, doy 229, doy 244 and doy 277, and four days in 2012: doy 87, doy 103, doy 151 and doy 167. Measurements were taken over four areas of 3 km by 3 km: two located in a dryland area and two in an irrigated area, with ten points per sampling area. Three separate measurements per sampling point were performed, which mounts to a total of 120 daily measurements within the entire area.

In order to derive soil texture, soil particle analysis was done at each sampling point location (*Merlin et al.*, 2013b). The data post-processing showed a mean clay fraction of 0.24 and a mean sand fraction of 0.37. Since gravimetric measurements were performed, the approach described in (*Saxton et al.*, 1986) was then used to convert them to volumetric values.

The *in situ* measurements serve as a validation dataset for the downscaled soil moisture products (see Chapter 5). Since they are point measurements, they were upscaled to a 3 km resolution by using a simple average.

2.2.3.2 Meteorological data

Two distinct meteorological datasets were used over the Urgell area: provided by weather stations and issued by an interpolation model. The meteorological variables that are of interest are air temperature, relative air humidity, net radiation and wind speed. This section provides a description of the dataset provided by weather stations. The description of the data provided by the interpolation model can be found in Section 2.3.3.

The meteorological data provided by weather stations concern stations integrated in the XEMA (Xarxa d'Estacions Meteorològiques Automàtiques i.e. Automatic Weather Station Network), run by the Meteorological Service of Catalonia (SMC). As of December 31st, 2011, it engulfs a total of 169 automatic weather stations. A list of the acquired data, along with the individual acquisition height is provided in Table 2.1. Measurements are conducted on either an hourly basis, or a 30 minutes basis, depending on each station.

The data used for implementation of the soil energy balance model over the Urgell site (see Chapter 3) was provided by the Mollerussa (XI) weather station. This station is favorably situated within the irrigated area and far enough from the mountains in order to rule out as much as possible the topographic effects. It provides meteorogological data at a 2 m height, every half hour. A visual representation of the study area, along with the *in situ* sampling areas and the location of the XEMA meteorological stations is shown in Figure 2.5.

Table 2.1: List of variables and acquisition heights, as measured by the XEMA meteorological stations

Variable	Acquisition height (m)
Air temperature	1.5
Precipitation	1.4 (except stations situated over 1800 m altitude)
Relative air humidity	1.5
Wind speed and directionality	2/ 6 / 10
Radiance	2
Atmospheric pressure	-
Snow thickness	-

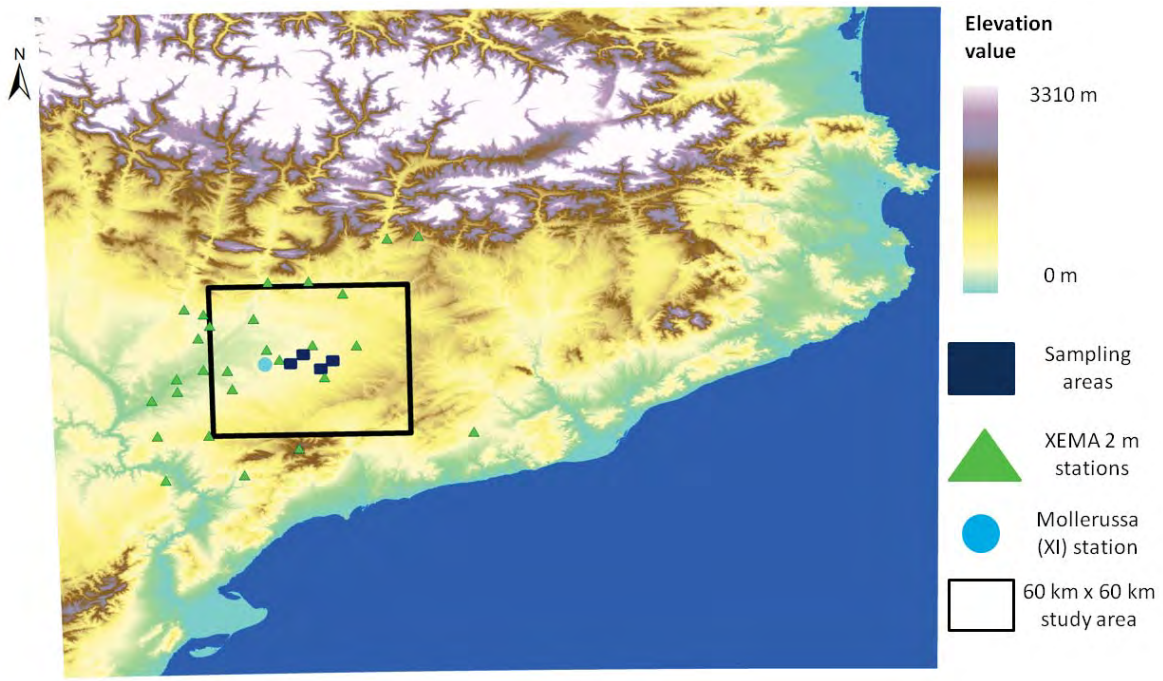


Figure 2.5: The Urgell 60 km by 60 km (encompassing the 20 km by 20 km area where *in situ* measurements where performed) study area, alongside the *in situ* sampling areas and the XEMA meteorological stations (with a 2 m acquisition height). The Mollerussa (XI) station whose data was used in this study is shown in light blue.

2.3 Spatial data

Different spatial datasets are used in this thesis: remote sensing data, meteorological data issued from an interpolation model and digital elevation data. First of all, an in–depth description of the remote sensing data is offered. Two distinct remote sensing data types are used: issued from optical (thermal and short–wave) and microwave sensors, which are both discussed below. Then, an overview of the interpolation model (SAFRAN) used to derive meteorological data over the Spanish site is presented. Digital elevation data is then described. In a final step, remarks regarding the data processing are covered.

2.3.1 Optical data

Short–wave data

Satellite surface reflectances (red and near–infrared bands) collected from ASTER and Formosat–2 have been used in order to derive NDVI, fractional green (photosynthetically active) vegetation cover (f_{vg}) and surface albedo (α) over the Yaqui and Urgell regions. These data will be specifically used as an input to the temperature–based model of ET presented in Chapter 3, in order to derive extreme temperatures and ET estimates.

The Formosat–2 satellite is an Earth observation satellite belonging to the National Space Organization of Taiwan, launched in 2004. It has an equator crossing time of 9:30 AM, providing images at 8 m resolution on a daily basis.

ASTER is an imaging instrument onboard NASA's Terra satellite. It has an equator crossing time at 11 AM, with a 16–day revisit cycle. Its short–wave infrared sensor has a resolution of 30 m. ASTER data are extracted over the 16 km by 10 km Yaqui and 60 km by 60 km (encompassing the 20 km by 20 km area where *in situ* measurements were performed) Urgell areas. Two cloud–free ASTER images were obtained over the Urgell site at approximately 11:00 AM UTC on August 16th and October 3rd, 2011. Seven cloud-free ASTER images were acquired over the Yaqui area during the agricultural season of 2007–2008, at around 11:30 AM local solar time on December 30th, February 23rd, March 10th , April 11th, April 27th, May 6th and May 13th.

The fractional green vegetation cover is obtained by using the *Gutman and Ignatov* (1998) formula:

$$f_{vg} = \frac{\mathrm{NDVI} - \mathrm{NDVI}_s}{\mathrm{NDVI}_{vg} - \mathrm{NDVI}_s} \tag{2.1}$$

with NDVI_{vg} corresponding to fully–covering green vegetation and NDVI_s to bare soil or to bare soil partially covered by senescent (non–photosynthetically active) vegetation. The values of NDVI_{vg} and NDVI_s are set to the maximum and minimum value of the NDVI observed during the agricultural season within the study domain. NDVI endmembers are estimated as 0.93 and 0.018, and 0.93 and 0.18 for the Urgell and Yaqui areas, respectively. The reason behind the small value registered for NDVI_s for Urgell is the presence of open water (reservoir of irrigation water) in the region. NDVI is computed as the ratio of the difference between (re–sampled) near–infrared and red reflectances to their sum.

Surface albedo (α) is computed as a weighted sum of (re–sampled) red and near–infrared reflectances, using the coefficients provided by *Weiss et al.* (1999) and corroborated by *Bsaibes et al.* (2009) and in *Chirouze et al.* (2014). It is derived over Urgell using ASTER data, while over the Yaqui area from Formosat–2 data. The reason behind using Formosat–2 derived albedo resides in the unusability of the ASTER shortwave infrared data on its seven overpass days over the Yaqui site (*Chirouze et al.*, 2014). Its high revisit cycle made it possible to use data collected on the nearest date from each of the seven ASTER overpass dates in order to estimate both NDVI and α from the red and near–infrared reflectances, aggregated at ASTER sensor resolution.

MODIS is another sensor which provides NDVI data that were used over Urgell, but as an input to the SMOS SM downscaling algorithm (see Chapter 5). MODIS is the payload scientific instrument onboard NASA's Terra satellite (launched in 1999) and Aqua satellite (launched in 2002) as part of the Earth Observing System (EOS) Programme. Data are captured in 36 spectral bands with wavelenghts ranging from 0.4 μm to 14.4 μm. The spatial resolution of the products ranges from 250 m to 500 m and to 1 km, depending on the bands used, with a daily time coverage.

The MODIS version–5 normalized difference vegetation index NDVI (MOD13A) was down-

loaded using the NASA Land Processes Distributed Active Archive Center (LP DAAC). The NDVI products are re–sampled on a 1 km grid, with a sinusoidal projection (*Solano et al.*, 2010; *Wan*, 2006, 1999). The NDVI product has a 16-day time coverage.

Thermal

ASTER also has a thermal sensor, consisting of five thermal bands, with a spatial resolution of 90 m. ASTER official surface kinetic temperature products (AST_08) were downloaded from the Earth Observing System Data Gateway. The "temperature and emissivity separation" algorithm (*Gillespie et al.*, 1998; *Schmugge et al.*, 1998) was used in order to retrieve the LST from ASTER over the Yaqui and Urgell regions.

MODIS has a thermal sensor as well, which is used to derive the LST. The MODIS version–5 LST products onboard Terra (MOD11A1) and Aqua (MYD11A1) were downloaded using the NASA Land Processes Distributed Active Archive Center (LP DAAC). The LST products, which have a daily coverage, are also re–sampled on a 1 km grid, like the NDVI products.

2.3.2 Microwave data

SMOS soil moisture data were used in order to obtain high resolution soil moisture products using DISPATCH (see Chapter 5). The primary goal of ESA's SMOS satellite, launched in November 2009, is to observe the soil moisture over Earth's landmasses and the salinity over Earth's oceans. It is the second Earth Explorer opportunity mission designed as part of ESA's Living Planet Programme. The novelty of the SMOS mission is a new measuring technique in the Earth observation domain: its payload incorporates for the first time in history a 2D interferometric radiometer (operating at the lowest band in the microwave domain), on a polar orbit. Microwave radiation is measured in L–band at a 1.4 GHz frequency (*Kerr et al.*, 2001, 2010), with the help of its Microwave Imaging Radiometer using Aperture Synthesis (MIRAS) instrument. SMOS has a global coverage and a 3–day equatorial revisiting period. It has two overpasses at (local solar time) 6 AM (ascending orbit) and 6 PM (descending orbit). Depending on the incidence angle, the spatial resolution of its products ranges from 35 to 55 km. The top soil layer (5 cm) SM is retrieved with an accuracy of 0.04 $m^3 \cdot m^{-3}$ (*Kerr et al.*, 2012).

The SMOS Level–3 1–day global SM product (MIR CLF31A/D), version 2.72 (in 220 reprocessing mode RE02) product is used in this study. They are available in a NetCDF format on the EASE (Equal Area Scalable Earth) grid, with a grid spacing of approximately 25 km by 25 km. The Level–3 SM products are computed directly from Level–1 products by the CATDS (Centre Aval de Traitement des Données SMOS). The approach for deriving SM from brightness temperature observations is developed from the Level–2 retrieval algorithm (*Wigneron et al.*, 2007; *Kerr et al.*, 2012). Both processing chains (Level–2 and Level–3) take into account multiangular observations of brightness temperatures in order to simultaneously retrieve SM and vegetation optical depth at nadir, by using a standard iterative minimization of a cost function. The main term of the cost function is given by the sum of squared weighted differences between measured and modeled brightness temperatures, at a variety of incidence angles. The algorithm then chooses the best set of parameters (SM and vegetation optical depth) that minimizes the cost function (*Kerr et al.*, 2012). The main difference between the two processing chains is that while Level–2 takes into consideration multiangular observations retrieved the same day and by the same orbit (ascending/descending), Level–3 takes into account multiangular observations of several overpasses (a maximum of 3), over a 7–day time frame. The Level–3 products thus present a better robustness (*Al-Yaari et al.*, 2014). A detailed description of the Level–3 processor can be found in the Algorithm Theoretical Baseline Document (*Kerr et al.*, 2013) and in the Level–3 product data description (*Kerr et al.*, 2014).

2.3.3 SAFRAN data

The second meteorological dataset (used in the downscaling of SMOS SM products, see Chapter 5) is provided by the SAFRAN (Système d'analyse fournissant des renseignements atmosphériques à la neige) model. It is a mesoscale atmospheric analysis system for surface variables, originally developed to offer an investigation of the atmospheric forcing in mountainous landscapes to be used in avalanche forecasting (*Durand et al.*, 1993, 1999). Its application was later extended to all France. A list of variables and the associated height at which they were predicted is provided in Table 2.2.

Table 2.2: List of variables and acquisition heights, as estimated by SAFRAN

Variable	**Height (m)**
Air temperature	2
Relative air humidity	2
Wind speed	10
Incoming solar radiation	-
Incoming atmospheric radiation	-
Cloudiness	-
Snowfall	-
Rainfall	-

A homogeneity of the atmospheric variables over well defined areas is assumed in the model, and the only possible variations are inflicted by topography. A homogeneous climate zone is considered as an area containing at least two rain gauges and one meteorologic station. The model takes into consideration all of the observed data and, for each variable (with the exception of downward visible and infrared radiation and precipitation), it uses an optimal interpolation method to estimate values at the considered altitudes. The observations used by SAFRAN are available at 0, 6, 12 and 18 hours UTC. The data is then interpolated in time at an hourly time step. The estimates are on a gridded dataset, at a 5 km resolution.

The incoming solar and atmospheric radiation and the precipitation are calculated in a different manner. A radiation model (*Ritter and Geleyn*, 1992) and previously analyzed variables are used to estimate radiation. Precipitation rate is computed using daily rain gauges; it is then interpolated on an hourly time step based on the evolution of relative air humidity: if the relative air humidity is sufficiently high, then precipitation is expected to occur. A partitioning based on air temperature is then used in order to separate snowfall from rainfall as follows: if air temperature is higher than 0.5 °C, then the precipitation is considered as rainfall. Otherwise, it is considered as snowfall.

SAFRAN was recently implemented over Urgell by meteorological analysis for screen level variables (*Quintana-Seguí et al.*). The study showed that in north–east Spain, SAFRAN emulates estimations in terms of temperature well, with no biases on the annual means. The mean RMSE of temperature is of about 1.5 °C. The relative air humidity is also estimated with no bias and with a mean RMSE of 7 %. There is a systematic negative bias of -0.2 $m \cdot s^{-1}$ and 1.3 $m \cdot s^{-1}$ RMSE detected in the wind speed estimations. In terms of radiation estimates, in north–east Spain, visible radiation presents a 10 $W \cdot m^{-2}$ bias and a 114 $W \cdot m^{-2}$ RMSE. Infrared radiation has a negative bias of -12 $W \cdot m^{-2}$ and a RMSE of 33 $W \cdot m^{-2}$. The results are equivalent with results validated over France (*Quintana-Seguí et al.*, 2008), proving that the model is robust. However, statistics were performed over large areas. In specific areas (like the the chosen study area in Urgell), results might be better or worse. The area of study considered in this analysis,

which is flat, should be favorable to SAFRAN.

The data used in this study can be found at the HyMex database (`http://mistrals.sedoo.fr/10.14768/MISTRALS-HYMEX.1388`).

The reason behind using SAFRAN data within DISPACTH is because they are spatialized data, and the aim of this part is to characterize the soil evaporaive efficiency (through its link with soil moisture) from a spatial point of view.

2.3.4 DEM data

Elevation data extracted from the GTOPO30 digital elevation model (DEM) are also required. It is a model developed by the United States Geological Survey and it has a resolution of ~ 1 km. It is split in 33 tiles, which are stored in the DEM file format. DEM data is used as an input to DISPATCH and DISPATCH–E (see Chapter 5) to correct the MODIS LST for topography effects.

2.3.5 Data pre–processing

This part deals with the pre–procesing applied to the SMOS, MODIS and SAFRAN datasets that were used as an input to DISPATCH.

DISPATCH was applied to the SMOS Level–3 1–day global SM product (both ascending and descending overpasses), MODIS version–5 LST products onboard Terra and Aqua – from one day before up to one day after the SMOS overpass, and normalized difference vegetation index NDVI – available each 16 days.

SMOS original datasets are downsampled in order to work at the radiometer resolution. SMOS L3 products are provided on a 25 km grid, which can be up to half of the original SMOS resolution (35–50 km). Following the SMOS re–sampling strategy described in (*Merlin et al.*, 2010b), re–sampled 40 km resolution SMOS data overlap four times over the area. The four SM datasets that are obtained at a 40 km resolution are assumed to be independent.

A maximum of 24 SM–LST input datasets can be generated, depending on the user's preferences. Four resampled SMOS datasets (for each ascending/descending orbit) can be used with six MODIS LST datasets (3 consecutive days of MODIS acquisitions for both Terra and Aqua overpasses). The downscaling relationship is applied to each input datasets, with 24 corresponding disaggregated products, which are then combined to produce 3 final outputs: the disaggregated SM dataset, the STD dataset, which is the standard deviation of the up to 24 disaggregated SM fields, and the CNT dataset, which is the size of the number of fields or size of the aggregation ensemble. The aggregation is done if at least 3 SM fields are generated, so the CNT values range from 3 to 24. The final disaggregated soil moisture product is generated on the intersection of these four SMOS grids.

All input data are extracted over MODIS tile(s) that encompass the user's area of interest. A visual representation of the 40 km SMOS SM data extracted over 1 MODIS tile (namely h18v04) that encloses our study area is shown in Figure 2.6. SAFRAN 5 km resolution data is also shown. In this study, SMOS data are extracted over a 200 km by 240 km area.

The pre–processing chain that is applied to the input dataset, that reprojects, resamples and extracts data over MODIS tiles before DISPATCH is applied, is known as DISPATCH Input Interface (*Molero et al.*, 2016).

From an operational point of view, considering the small extent of the study area, modifications were made with respect to the input dataset. In order to save computational time, new smaller grids were defined, and SMOS, MODIS, DEM and SAFRAN data were extracted over these new grids. A visual representation of a new grid is shown in Figure 2.7.

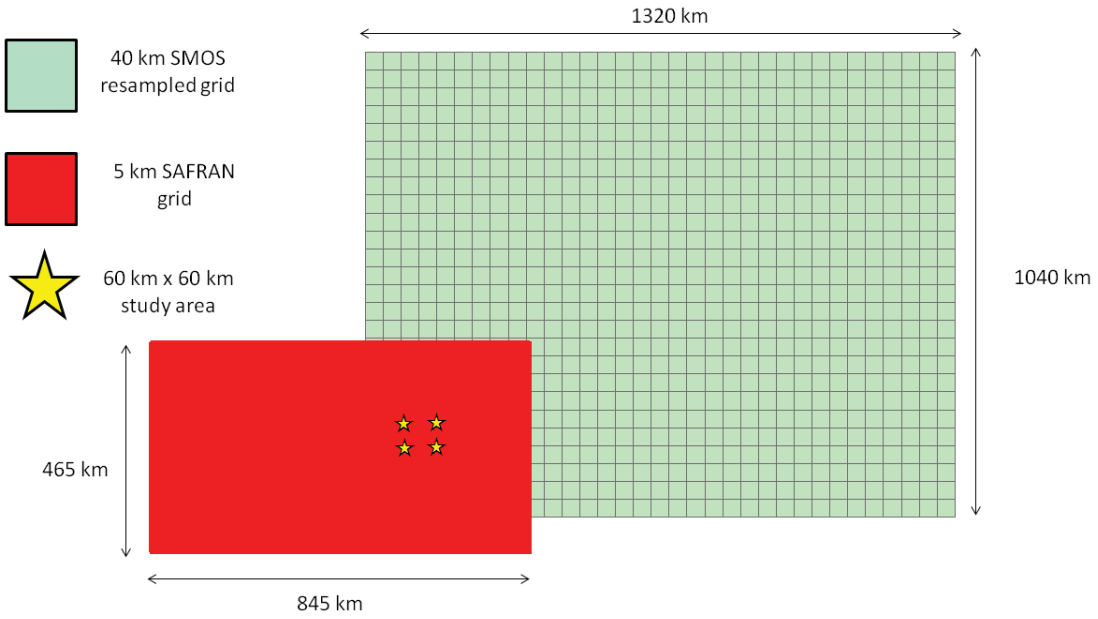


Figure 2.6: 40 km SMOS SM data corresponding to one MODIS tile that encompasses the 60 km by 60 km Urgell study area. SAFRAN 5 km resolution data is also shown.



Figure 2.7: An example of one of the new, smaller SMOS grids (amongst a total of four).

A pre–processing of the SAFRAN data was also independently performed. Firstly, data close to the two distinct MODIS overpasses were extracted: 10:00 AM and 11:00 AM, and 1:00 PM and 2:00 PM, respectively. Data were then averaged (simple mean) in order to have estimates at the corresponding MODIS overpass times: 10:30 AM (Terra) and 1:30 PM (Aqua).

The second step in the pre–processing consists in extracting the data corresponding to the SMOS grids that cover the study area. Finally, the data were upscaled from 5 km (SAFRAN resolution) to 40 km (SMOS resolution) by simple averaging.

2.4 Conclusion

Three different semi–arid regions have been used as study sites for the implementation of different models, which provide estimates in terms of ET (Chapter 3) and SM (Chapter 5). The common point between the three sites is given by estimates of the extreme soil temperatures, which are derived from two different models: image–based (which uses solely remote sensing data), and an energy balance model (which uses solely meteorological data). Using more than one site in the comparison of extreme temperatures is a measure of the consistency in the estimates provided by the two models, in different surface and atmospheric conditions. *In situ* measurements of E are missing, so a direct evaluation of the SEE models cannot be performed. The approach in this thesis was to validate extreme soil temperatures as a proxy for the SEE and to evaluate the DISPATCH output (SM) as an assessement of the SEE model.

Chapter 3

Improving the wet and dry boundaries of a contextual evapotranspiration model

Contents

3.1 Introduction 43
3.2 A contextual method 45
3.2.1 SEB–1S overview 45
3.2.2 Image–based extreme temperatures 46
3.2.3 Adjustments to the estimation of the image–based extreme temperatures 48
3.3 Modeling extreme temperatures 49
3.3.1 Different aerodynamic resistance formulations 49
3.3.2 Vegetation extreme temperatures 50
3.4 Results 51
3.4.1 Evaluation of extreme soil temperatures 52
3.4.2 Application to evapotranspiration estimation at multiple resolutions . . 55
3.5 Summary and Conclusions 57
3.6 Article 58

3.1 Introduction

The objective of this PhD is to improve the representation of soil evaporation from a spatio–temporal point of view, in a range of different soil and atmospheric conditions, by making use of available remote sensing data. However, what is observed from remote sensing data is rather the soil evaporative efficiency (SEE), defined as the ratio of soil evaporation to potential evaporation. One way of deriving SEE by remote sensing means is from LST data, using the following formula (*Shuttleworth et al.*, 1989; *Nichols and Cuenca*, 1993):

$$SEE = \frac{T_{s,dry} - T_s}{T_{s,dry} - T_{s,wet}} \tag{3.1}$$

where $T_{s,dry}$ and $T_{s,wet}$ are the soil temperatures corresponding to fully dry and fully wet conditions, and T_s is the soil temperature derived from remote sensing data. The linearity between SEE and T_s in Equation 3.1 was recently verified in (*Merlin et al.*, 2016).

However, currently there are very few available ground measurements of E to be used for validation purposes. Therefore, remote sensing approaches are mostly validated in terms of total evapotranspiration (ET), whose ground measurements are more accesible via eddy covariance techniques (see Chapter 2). There are two methodologies used in deriving ET through LST: residual and contextual methods. Contextual methods such as SEB–1S (*Merlin*, 2013), S–SEBI (*Roerink et al.*, 2000), which calculate ET as the available energy times the evaporative fraction (EF), are appealing because unlike residual methods, they do not require calibration. EF is defined as the ratio of latent heat to available energy and can be estimated using the formula (*Roerink et al.*, 2000; *Merlin*, 2013):

$$EF = \frac{T_{dry} - T}{T_{dry} - T_{wet}} \tag{3.2}$$

where T is the surface temperature corresponding to a given point in the space defined by LST and the fractional vegetation cover (f_{vg}) and/or the space defined by LST and the soil albedo (α), and T_{dry} and T_{wet} are the surface temperatures corresponding to fully dry and fully wet conditions.

Note that Equations 3.1 and 3.2 are semi-empirical; they are not based on analitical computations that derive from physical equations.

An important step in contextual methods consists in determining the wet and dry edges of the LST. These edges can be determined by computing the exreme temperatures (T_{ext}) that correspond to fully dry and wet conditions for both soil and vegetation components in the space defined by LST and f_{vg} and/or the space defined by LST and α. The vertices of the obtained polygons are subsequently joined by straight lines or curves in order to interpolate the wet and dry edges. Contextual methods are very attractive because they can be easily transferred to different areas thanks to their image–based nature. However, one of their main limitation is that T_{ext} be actually observed at the thermal sensor resolution within a study domain characterized by uniform atmospheric conditions.

By developing algorithms that estimate T_{ext} independently of the surface conditions (so independently of the contextual nature) within the study area, some of the limitations associated to contextual methods could be dismissed. These models would estimate T_{ext} independently of the spatial resolution of the LST data (*Moran et al.*, 1994; *Long and Singh*, 2012) by running a surface energy balance model forced by meteorological data.

This part mainly aims at developing a theoretical approach to estimate T_{ext} independently of LST data, by using an energy balance model forced by available meteorological data. Estimates in terms of T_{ext} (as calculated by the energy balance model) will then be compared with *in situ* measurements (over the Moroccan site) and with image–based estimates provided by a contextual model (over the Mexican and Spanish sites). T_{ext} estimates given either by the energy balance model or by the contextual model will then be used as input in the same EF model in order to derive ET estimates over the Mexican site.

First of all, a description of the contextual model (SEB–1S) used, along with the modifications made to it are presented in Section 3.2. The soil energy balance model is then described in Section 3.3. Results in terms of T_{ext} and ET estimates are then presented in Section 3.4. A summary and general conclusions can be found in Section 3.5. The article corresponding to this analysis can be found in Section 3.6, providing a more detailed description of the soil energy balance model, additional results in terms of ET validation with *in situ* measurements, as well as a

mixed modeling approach that combines the energy balance model with the image-based model in deriving T_{ext}.

3.2 A contextual method

3.2.1 SEB–1S overview

The contextual model chosen to derive ET is the SEB–1S model (*Merlin et al.*, 2013a). Being a contextual method as *Gómez et al.* (2005), (*Sobrino et al.*, 2005), *Sobrino et al.* (2007), *Fan et al.* (2007), *Galleguillos et al.* (2011a), *Galleguillos et al.* (2011b), SEB–1S derives the evaporative fraction EF as the ratio of two distances: the distance separating a pixel in the $LST - \alpha$ space from the dry edge, and the distance separating the dry and the wet edges. The next step after the estimation of EF is to compute ET as EF multiplied by the available energy $(Rn - G)$, where Rn (W·m^{-2}) is the surface net radiation and G (W·m^{-2}) the ground heat flux. A visual representation of the computation behind EF, as well as the interpretation behind the edges of the polygon in the $LST - \alpha$ space is presented in Figure 3.1.

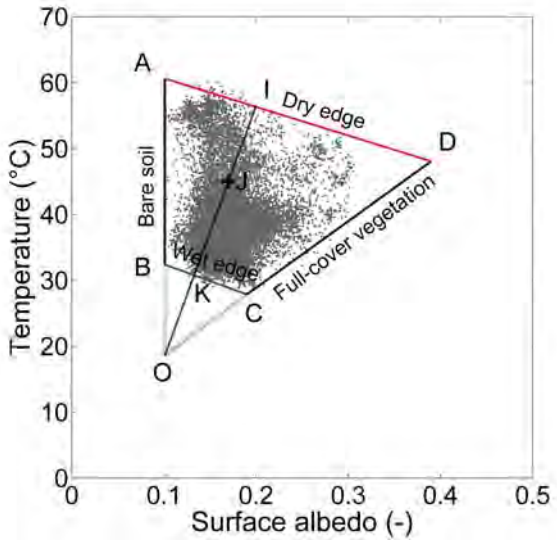


Figure 3.1: For a given point J in the $LST - \alpha$ space, the EF is calculated as the ratio of IJ to IK. Underlying grey points correspond to 90 m resolution LST data acquired over Yaqui, on April 27th, 2008.

The reason behind using SEB–1S to derive ET estimates is twofold: i) it provides a consistent interpretation of the edges in both spaces and ii) it is based on a synergy between the two $LST - f_{vg}$ and $LST - \alpha$ spaces with respect to the determination of extreme temperatures. Therefore, SEB–1S builds upon the two commonly used contextual approaches, namely the trapezoid method of *Moran et al.* (1994) and the S–SEBI method of *Roerink et al.* (2000). Since the determination of exteme temperatures in either space is restricted if fully wet and dry conditions are not met at the optical sensor resolution, a synergistic combination of the two spaces can help improve the estimates (*Merlin*, 2013).

3.2.2 Image–based extreme temperatures

Based on a synergy between the $LST - \alpha$ and the $LST - f_{vg}$ spaces, SEB–1S has a submodule (hereby referred to as Text_RS) that provides estimates of the extreme soil and vegetation temperatures. A visual representation of the meaning behind the edges of the polygons in both the $LST - \alpha$ and in the $LST - f_{vg}$ spaces is provided in Figure 3.2, for data gathered over the Yaqui site on April 11th, 2008.

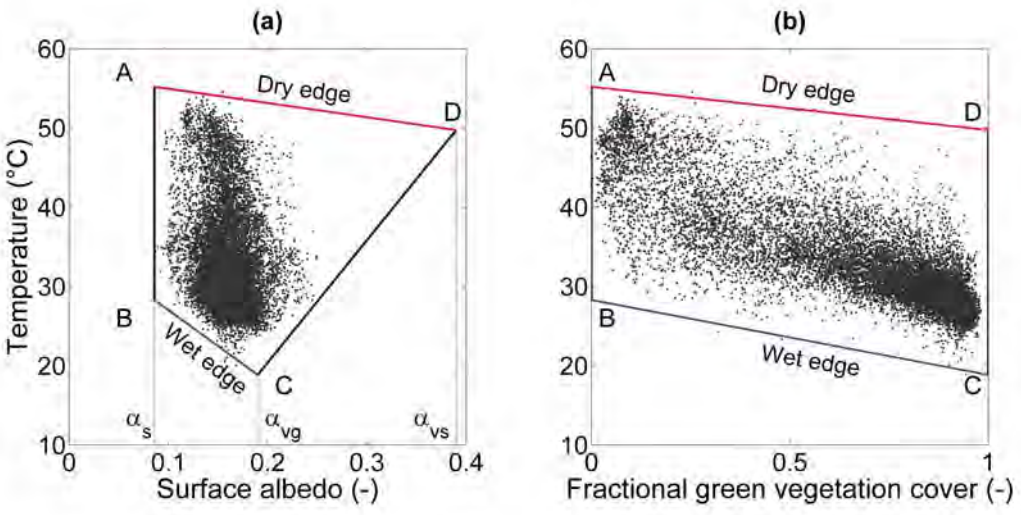


Figure 3.2: Meaning behind the edges of the polygons identified in the $LST - \alpha$ **(a)** and the $LST - f_{vg}$ **(b)** spaces. Underlying grey points correspond to 90 m resolution LST data acquired over Yaqui, on April 11th, 2008.

For each space, SEB–1S determines four T_{ext} and four polygon edges. However, in the $LST - f_{vg}$ space, it identifies two extreme values for f_{vg} (0 and 1), while in the $LST - \alpha$ space, three extreme albedo values are identified: α_s (corresponding to bare soil), α_{vg} (corresponding to green vegetation) and α_{vs} (corresponding to senescent vegetation). The interpretation of the edges thus slightly differs depending on the space.

For the $LST - \alpha$ space, the edges are interpreted as follows:

- (AB) – "bare soil"
- (BC) – "wet surface"
- (CD) – "full-cover vegetation"
- (DA) – "dry surface"

For the $LST - f_{vg}$ space, the edges are considered as:

- (AB) – "mixed soil and senescent vegetation"
- (BC) – "wet surface"
- (CD) – "full-cover green vegetation"

- (DA) – "dry surface"

An important aspect to mention is the negative slope of the dry edge (AD). This is explained by both radiative and convective controls in dry conditions (*Roerink et al.*, 2000; *Merlin et al.*, 2013a). Assuming a constant global radiation and air temperature, there is a formal relationship between surface temperature and albedo. For low albedo values (characteristic of open water and irrigated lands), surface temperature is almost constant with increasing albedo. Beyond a certain threshold, surface temperature decreases with increasing albedo; this is explained by the fact that all energy is now used to heat up the surface (hence radiative and convective controls) since soil moisture is so low that evaporation cannot take place. The reason behind the negative slope corresponding to the wet edge (BC) identified in the $LST - \alpha$ space is the convective control in wet conditions (*Merlin et al.*, 2013a), consistent with the common interpretation of the $LST - f_{vg}$ space.

In the original version of SEB–1S, the minimum vegetation temperature $T_{v,wet}$ is set to the air temperature T_a (*Long and Singh*, 2012; *Jiang and Islam*, 1999). The maximum soil temperature $T_{s,dry}$ is the same for both spaces and is set to the maximum surface temperature T_{max}. The other T_{ext} are estimated by combining the T_{ext} retrieved from the $LST - \alpha$ and $LST - f_{vg}$ spaces separately. The minimum soil temperature $T_{s,wet}$ is computed by extrapolating the wet edge passing through the well-watered full-cover green vegetation index until the bare soil line. The maximum vegetation temperature $T_{v,dry}$ is computed by extrapolating the dry edge passing through the well-watered full-cover green vegetation index until the bare soil line. This process is undertaken for each space separately. $T_{s,wet}$ and $T_{v,dry}$ are then set to the average of the two individual corresponding values retrieved in each space.

More specifically, the extreme soil and vegetation temperatures in the $LST - \alpha$ space are given by:

- $T_{v,wet,1}$ (at $\alpha = \alpha_{vg}$) is set to the air temperature T_a.
- $T_{s,dry,1}$ (at $\alpha = \alpha_s$) is set to the maximum temperature T_{max} observed within the study area.
- $T_{s,wet,1}$ is defined as the intercept at $\alpha = \alpha_s$ of the line passing through the point (α_{vg}, T_a) and the point with $\alpha < \alpha_{th,1}$ such that the slope of the line is maximum (meaning that all the other data points with $\alpha < \alpha_{th,1}$ are located above the wet surface edge). $\alpha_{th,1}$ is set to α_{vg}.
- $T_{v,dry,1}$ is defined as the intercept at $\alpha = \alpha_{vs}$ of the line passing through (α_s, $T_{s,dry,1}$) and the point with $\alpha > \alpha_{vs}$ such that the slope of the line is maximum (meaning that all the other data points with $\alpha > \alpha_{th,1}$ are located below the dry surface edge).

Regarding the $LST - f_{vg}$ space, extreme soil and vegetation temperatures are computes as follows:

- $T_{v,wet,2}$ (at $f_{vg} = 1$) is set to the air temperature T_a.
- $T_{s,dry,1}$ (at $f_{vg} = 0$) is set to T_{max}.
- $T_{s,wet,2}$ is defined as the intercept (at $f_{vg} = 0$) of the line passing through the point (1, T_a) and the point with $f_{vg} < f_{vg,avg}$ such that the slope of the line is maximum (meaning that all the other data points with $f_{vg} < f_{vg,avg}$ are located above the wet surface edge). The threshold value ($f_{vg,avg}$) is set to 0.5.

- $T_{v,dry,2}$ is defined as the intercept (at $f_{vg} = 1$) of the line passing through the point (0, $T_{s,dry,2}$) and the point with $f_{vg} > f_{vg,avg}$ such that the slope of the line is maximum (meaning that all the other data points with $f_{vg} > f_{vg,avg}$ are located below the dry surface edge).

To summarize:

$$T_{s,dry} = T_{s,dry,1} = T_{s,dry,2} = T_{max} \tag{3.3}$$

$$T_{v,wet} = T_{v,wet,1} = T_{v,wet,2} = T_a \tag{3.4}$$

The final step in estimating the extreme T_{ext} consists in averaging the two distinct T_{ext} sets obtained for each space:

$$T_{s,wet} = \frac{(T_{s,wet,1} + T_{s,wet,2})}{2} \tag{3.5}$$

$$T_{v,dry} = \frac{(T_{v,dry,1} + T_{v,dry,2})}{2} \tag{3.6}$$

3.2.3 Adjustments to the estimation of the image–based extreme temperatures

Adjustments to the estimation (as originally implemented in SEB–1S) of the image–based T_{ext} are performed in this study. SEB–1S was developed to be used with remote sensing data at high resolution ($\sim$ 100 m). Particular attention was given to enhancing the robustness of the SEB–1S submodule that estimates extreme temperatures when using remote sensing data with a lower spatial resolution ($\sim$ 1 km). Adjustments primarily concern the image–based thresholds that are taken into consideration when determining the wet and dry edges. An effort was made into defining new thresholds such that the wet and dry edges are estimated with a better precision, in the case when they are not necessarily representative of the extreme soil and vegetation conditions.

Modifications concern the calculation of the minimum soil temperature and maximum vegetation temperature in both spaces. More specifically, the thresholds are now computed differently:

- $\alpha_{th,1}$ is now considered as the average between α_{vg} and α_s.

- a new threshold is defined for $T_{v,dry,1}$: α_{avg}, which is the average of all α values within the study area.

- $f_{vg,avg}$ is now computed for each day separately.

Choosing different threshold values was done in order to render the methodology more robust when using 1 km resolution remote sensing data. In particular, the new thresholds are calculated taking into account the surface conditions observed at a 1 km resolution. The thresholds used in the original version of SEB–1S are suited for high resolution data, when there are enough points in the $LST - \alpha$ and $LST - f_{vg}$ spaces to cover the real surface conditions; the same values might not prove so robust at a low resolution.

3.3 Modeling extreme temperatures

In order to estimate T_{ext} independently of remote sensing data, a different approach is suggested. It consists in applying a soil energy balance (EBsoil) model forced by available meteorological data in order to compute soil T_{ext}. The goal is to use a minimum number of *a priori* input parameters, hence, it is applied in bare soil conditions only. Vegetation T_{ext} are then estimated from soil T_{ext}.

EBsoil runs an iterative loop on the soil temperature T_s (intialized to the air temperature T_a) until thermal equilibrium is reached (*Norman et al.*, 1995):

$$Rn_s - G = H_s + LE_s \tag{3.7}$$

where Rn_s ($\mathrm{W{\cdot}m^{-2}}$) is the surface net radiation, G ($\mathrm{W{\cdot}m^{-2}}$) is the ground heat flux, H_s ($\mathrm{W{\cdot}m^{-2}}$) the soil sensible heat flux and LE_s ($\mathrm{W{\cdot}m^{-2}}$) the soil latent heat flux.

r_{ss} is considered as a function of surface soil moisture (*Sellers et al.*, 1992):

$$r_{ss} = exp(A - B\frac{SM}{SM_{fc}}) \tag{3.8}$$

where the two best fit parameters A (unitless) and B (unitless) are considered as 8 and 5 respectively (*Kustas et al.*, 1993; *Crow et al.*, 2008). SM represents the surface (0 - 5 cm) soil moisture, whereas SM_{fc} the soil moisture at field capacity.

By prescribing a soil evaporation resistance r_{ss} equal to zero and infinity (in practice a very large number), the minimum and maximum soil temperatures can be estimated for a given atmospheric forcing. Note that a full description of the model can be found in the article Annex B.

3.3.1 Different aerodynamic resistance formulations

One of the variables that intervene in the latent and sensible heat fluxes of the energy balance equation is the aerodynamic resistance. Several parameterizations of this variable exist in literature. They fall into two broad categories: i) those based on the Monin–Obukhov similarity theory (*Monteith*, 1973; *Choudhury et al.*, 1986; *Lee*, 1997; *Yang et al.*, 2001) and ii) those based on a semi–empirical approach (*Liu et al.*, 2006).

Several assumptions and simplifications concerning the stability effects can be considered, which have a different impact when comparing estimates to measurements of r_{ah} (*Liu et al.*, 2006). Under stable conditions, the aerodynamic resistance can be effortlessly estimated, thanks to the linearity of the functions of the stability parameters. However, difficulties arise under non–stable conditions, induced by the non–linearity of these functions. In this case, iterative algorithms were developped to predict the aerodynamic resistance, approach that is referred to as the "Monin–Obukhov Similarity" method.

Two different formulations are chosen to be implemented within EBsoil, and their impact on the estimation of T_{ext} is analyzed: one based on the Richardson number and one based on the Monin-Obukhov length. The first formulation of the aerodynamical resistance (herein referred to as $r_{ah,RI}$) is based on the Richardson number R_i (unitless), which represents the importance of natural relative to forced convection. It is given by the formulation in *Choudhury et al.* (1986)

as:

$$r_{ah,RI} = \frac{r_{ah,SI}}{(1+R_i)^{\eta}} \tag{3.9}$$

where η (unitless) being a coefficient equal to 0.75 in unstable conditions ($T_s > T_a$) and to 2 in stable conditions ($T_s < T_a$).

$r_{ah,SI}$ is a simple formulation of the aerodynamical resistance and it symbolizes the aerodynamic resistance that neglects natural convection:

$$r_{ah,SI} = \frac{log^2 \frac{Z_r}{Z_{0m}}}{k^2 u_a} \tag{3.10}$$

where Z_{0m} (m) the roughness length for momentum transfer over bare soil, Z_r (m) the reference height at which the wind speed is measured, k (unitless) the von Kármán constant and u_a (m·s^{-2}) the wind speed.

The bulk Richardson number is a measure of the influence of atmospheric stability on the flux–gradient relationship in the surface layer:

$$R_i = \frac{5gZ_r(T_s - T_a)}{T_a u_a^2} \tag{3.11}$$

where g (m·s^{-2}) is the gravitational constant. This method is referred herein to as the "empirical" method.

The second formulation of the aerodynamial resistance used in this study (herein referred to as $r_{ah,MO}$), which takes into account the Monin–Obukhov length, is expressed as:

$$r_{ah,MO} = \frac{\log \frac{Z_r}{Z_{0m}} - \psi_h}{k u_*} \tag{3.12}$$

where ψ_h (unitless) is the stability correction factor for heat transport and u_* (m·s^{-1}) the friction velocity.

The stability correction factor for heat transport is defined as:

$$\psi_h = 2 \log \frac{1+x^2}{2} \tag{3.13}$$

where x is a function of the Monin–Obukhov length (L_{mo}) and of the reference height for wind speed observations:

$$x = (1 - 16\frac{Z_r}{L_{mo}})^{0.25} \tag{3.14}$$

3.3.2 Vegetation extreme temperatures

Once soil T_{ext} are obtained, two hypotheses are made in order to derive vegetation T_{ext}. As in (*Carlson et al.*, 1995; *Prihodko and Goward*, 1997; *Bastiaanssen et al.*, 1998), the minimum vegetation temperature $T_{v,wet}$ corresponding to well–watered unstressed vegetation is set to T_a, an assumption that is made in several models such as SEBAL (*Bastiaanssen et al.*, 1998), S–

SEBI (*Roerink et al.*, 2000) and SEB–1S (*Merlin et al.*, 2013a). It can also be considered as a fraction of T_a, as in (*Senay et al.*, 2013), who, after observing a systematic bias between T_s and T_a, introduced a correction factor equal to 0.993. Note that by setting the minimum vegetation temperature to T_a, the roughness length for momentum transfer over full–cover unstressed vegetation is implicitly constrained (*Stefan et al.*, 2015).

The assumption made concerning the maximum vegetation temperature $T_{v,dry}$ corresponding to fully water–stressed vegetation is based on the observation that the lines (AD) and (BC) in the $LST - f_{vg}$ space are practically parallel when using high resolution remote sensing data in highly heterogeneous conditions, as noted in *Merlin et al.* (2013a). This leads to $T_{v,dry}$ being calculated as:

$$T_{v,dry} = T_{s,dry} - (T_{s,wet} - T_{v,wet}) \tag{3.15}$$

This paralellity assumption implies that the difference between the aerodynamic resistances corresponding to the water–saturated bare soil and to the full–cover unstressed vegetation is almost equal to the difference between the aerodynamic resistances corresponding to the bone dry bare soil and to the full–cover water stressed vegetation, respectively. Thus, we neglect the influence of water stress on vegetation height. This translates into the decreasing rate in surface temperature with increasing vegetation cover (and the related increase in roughness length for momentum transfer from bare soil to full–cover vegetation) be almost the same for both dry and wet surface edges (*Stefan et al.*, 2015). In a similar manner as for the $T_{v,min} = T_a$ strategy, Equation 3.15 represents a constraint on the roughness length over full–cover water–stressed vegetation.

One of the advantages of estimating $T_{v,wet}$ and $T_{v,dry}$ from T_a and soil T_{ext} is not needing extra information on vegetation height information, nor on the zero plane displacement, on the roughness length for momentum transfer and on the roughness length for heat transfer. The sole parameter to be known is the roughness length for momentum transfer (Z_{0m}) over bare soils, whose range of values reported by literature is between 0.0006 an 0.035 m (*Liu et al.*, 2007; *Allen et al.*, 2007; *Timmermans et al.*, 2007; *Yang et al.*, 2008; *Long and Singh*, 2012). In this study, it is uniformly set to 0.001 m as in *Yang et al.* (2008).

3.4 Results

Results in terms of both soil T_{ext} and ET estimates are presented in this section. First of all, in terms of soil T_{ext}, a validation of the extreme temperatures modeled using the energy balance (Text_EBsoil) is done against *in situ* measurements over the R3 perimeter. Secondly, a comparison between modeled Text_EBsoil and the image–based extreme temeperatures, derived solely from remote sensing data (Text_RS_90m, simulated using 90 m resolution ASTER data) is performed over the Yaqui and Urgell areas. The soil T_{ext} modeled by the soil energy balance model takes into account both formulations of the aerodynamical resistance: $r_{ah,RI}$ and $r_{ah,MO}$.

The influence of model–derived and image-based T_{ext} is then assessed in terms of contextual ET estimates, by separately injecting T_{ext} as inputs into the SEB–1S model. SEB–1S is then applied to both 90 m resolution and 1 km resolution (aggregated) ASTER data.

3.4.1 Evaluation of extreme soil temperatures

3.4.1.1 Validation with *in situ* measurements

A comparison between modeled soil T_{ext} and *in situ* measurements acquired over the R3 perimeter is shown in Figure 3.3. Soil T_{ext} are modeled using either $r_{ah,RI}$ or $r_{ah,MO}$. Results in terms of bias, correlation coefficient (R), slope of the linear regression and root mean square error (RMSE) are presented in Table 3.1.

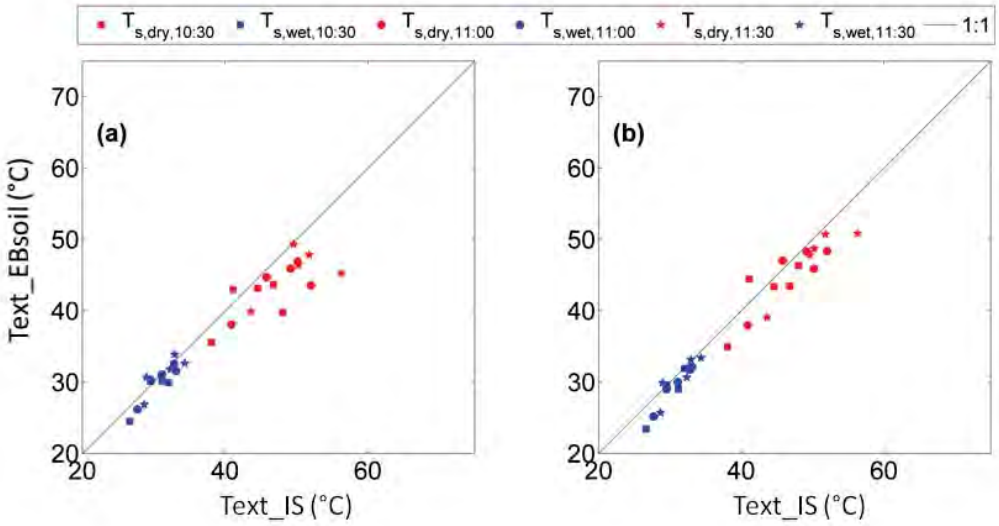


Figure 3.3: Soil T_{ext} simulated by EBsoil are plotted against the *in situ* measured T_{ext}, at the two soil plots in the R3 perimeter. Both RI (**a**) and MO (**b**) formulations of the aerodynamic resistance are taken into account when deriving modeled T_{ext}.

Table 3.1: Root mean square error (RMSE), bias, correlation coefficient (R) and slope of the linear regression between soil T_{ext} (as given by EBsoil) and the soil T_{ext} observed *in situ* at the two soil plots, located in the R3 perimeter. Modeled T_{ext} take into account either the RI or MO aerodynamic resistance.

Time	**RMSE (°C)**		**Bias (°C)**		**R (-)**		**Slope (-)**	
	RI	**MO**	**RI**	**MO**	**RI**	**MO**	**RI**	**MO**
10:30 a.m.	3.2	2.3	−1.8	−1.2	0.93	0.96	0.82	0.99
11:00 a.m.	3.3	2.3	−2.2	−1.7	0.97	0.98	0.79	0.94
11:30 a.m.	4.2	2.7	−2.5	−1.9	0.95	0.98	0.74	0.90
Mean	3.6	2.4	−2.2	−1.6	0.95	0.97	0.78	0.94

One can clearly see that the MO formulation of the aerodynamical resistance performs better. The mean (average of the correlation values obtained at 10:30 AM, 11:00 AM and 11:30 AM) correlation is increased from 0.95 (RI) up to 0.97 (MO) formulation, respectively. The mean RMSE is equal to 2.4 °C for the MO formulation, compared to 3.6 °C for the RI formulation. In terms of mean bias, the lowest value is also obtained for the MO formulation, with a reported value of −1.6 °C, as opposed to −2.2 °C that is obtained for the RI formulation. A separate

analysis showed that more bias is found for $T_{s,dry}$ than for $T_{s,wet}$. It was also found that $T_{s,dry}$ is more weakly correlated with *in situ* measurements than $T_{s,wet}$. A possible explanation behind this could reside in the $T_{s,dry}$ measurements presenting greater variability than the $T_{s,wet}$ measurements, as explained previously in Section 2.2.2.

To sum up, in terms of direct validation with *in situ* measurements, soil T_{ext} simulated using the MO formulation provide good estimates. Overall, the uncertainty in modeled $T_{s,dry}$ is larger than in modeled $T_{s,wet}$, which is to be expected considering the larger range of variation reported for $T_{s,dry}$.

3.4.1.2 Consistency with image–based extreme temperatures

In this section, soil T_{ext} modeled by EBsoil are compared against image–based soil T_{ext}, issued using the SEB–1S Text_RS submodel, using 90 m resolution data. The analysis is performed over the Yaqui and Urgell regions. Figure 3.4 plots model–derived (Text_EBsoil) *versus* image–based (Text_RS_90m) soil T_{ext} for the Yaqui and Urgell areas separately, and using the RI and MO r_{ah} formulations.

Table 3.2 provides the daily bias in $T_{s,wet}$ and $T_{s,dry}$, respectively. The mean and standard deviation of daily biases is also computed over the Yaqui region. The mean of daily biases is also reported over the Urgell segion. The mean bias is systematically lower when using $r_{ah,MO}$, with reported values over Yaqui of -3.4 (for $T_{s,dry}$) and 1.5 °C (for $T_{s,wet}$) compared to -7.7 and 2 °C obtained when using $r_{ah,RI}$. Results are consistent over Urgell, with a mean bias of -5.8 (for $T_{s,dry}$) and 2.8 °C (for $T_{s,wet}$) for the MO formulation, as opposed to -8.0 and 3.9 °C for the RI formulation. The robustness and precision of the image–based Text_RS_90m over highly heterogeneous irrigated areas is reflected by the consistency observed between model–derived and image–derived soil T_{ext}. Both the physical approach based on the MO theory and the remote sensing T_{ext} algorithm (that takes into account both $LST - \alpha$ and $LST - f_{vg}$ spaces) are hence validated by obtaining such a consistency.

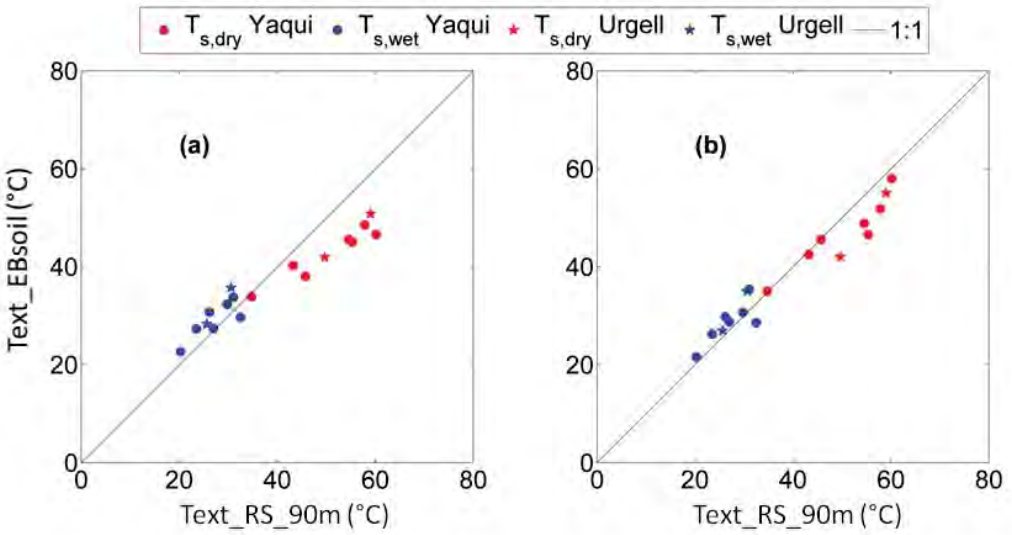


Figure 3.4: Soil T_{ext} simulated by EBsoil using either the RI(**a**) or MO (**b**) aerodynamic resistance are plotted against the image–based T_{ext} for the Yaqui and Urgell regions separately.

Table 3.2: Difference between the soil T_{ext} simulated by EBsoil with the RI and MO aerodynamic resistance and the soil T_{ext} retrieved using 90 m resolution data, for the Yaqui and Urgell regions separately.

Day	**Site**	**Bias $T_{s,dry}$ (°C)**		**Bias $T_{s,wet}$ (°C)**	
		RI	**MO**	**RI**	**MO**
30 December	Yaqui	−0.84	0.23	2.4	1.3
23 February	Yaqui	−3.0	−0.80	3.8	2.8
10 March	Yaqui	−7.7	−0.21	0.40	1.8
11 April	Yaqui	−9.0	−5.7	4.6	3.6
27 April	Yaqui	−14	−2.2	2.8	4.3
6 May	Yaqui	−10	−8.8	−2.8	− 3.9
13 May	Yaqui	−9.3	−6.0	2.6	0.90
16 August	Urgell	−8.3	−4.0	5.2	4.4
3 October	Urgell	−7.8	−7.6	2.6	1.3
All (mean)	Yaqui	−7.7	−3.4	2.0	1.5
All (σ)	Yaqui	4.4	3.5	2.5	2.7
All (mean)	Urgell	−8.0	−5.8	3.9	2.8

As a step further, the methodology is also implemented using 1 km resolution (aggregated) ASTER data over the two sites. Figure 3.5 shows side by side the $LST-\alpha$ and $LST-f_{vg}$ spaces retrieved using either 90 m resolution (grey dots) or 1 km resolution (black dots) data acquired over the Yaqui site on December 30th. The polygons obtained using modeled Text_EBsoil (RI and MO formulations) and image–based Text_RS_90m and Text_RS_1km are also shown.

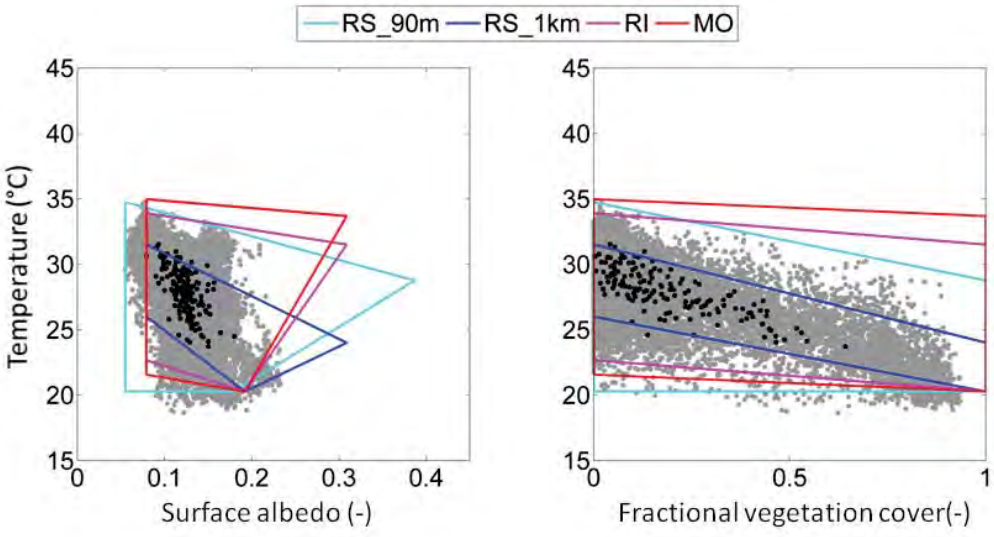


Figure 3.5: Polygons obtained using either image–based Text_RS_1km (blue) and Text_RS_90m (light blue), or model–derived Text_EBsoil for the RI (magenta) and MO (red) aerodynamic resistance are shown in the $LST-\alpha$ and $LST-f_{vg}$ spaces. The spaces correspond to either 1 km resolution data (black dots) or 90 m resolution data (grey dots) acquired over the Yaqui area on December 30th.

A difference can be observed in the boundaries of the two spaces when looking at Figure 3.5. At 90 m resolution, the polygons derived using Text_RS_90m manage to capture quite well the contour of the LST data points. Noticeable differences can be remarked when looking at the other three polygons, especially in the $LST - \alpha$ space. The reason why these three polygons are narrower and closer to the 1 km resolution data is that the albedo endmembers for Text_EBsoil and Text_RS_1km are derived from 1 km resolution ASTER data. Between the two r_{ah} formulations, the MO provides dry and wet edges closer to the polygon obtained with Text_RS_90m, a result expected since a better performance in terms of soil T_{ext} was reported by using the same MO formulation. When compared to the Text_RS_1km algorithm, which systematically underestimates the dry edge and overestimates the wet edge, the $r_{ah,MO}$ significantly improves the estimations. Therefore, the aerodynamic resistance plays a significant role in the computation of physics–based T_{ext}, with the MO formulation improving the dry and wet edge estimation, as compared to the RI formulation.

3.4.2 Application to evapotranspiration estimation at multiple resolutions

The idea behind using EBsoil to derive T_{ext} is to strip T_{ext} of their contextual nature, by applying ET models to remote sensing data retrieved at multiple resolutions and over less heterogeneous areas. To quantitatively assess the impact of T_{ext} on 1 km ET estimates, SEB–1S is applied to ASTER data aggregated at 1 km resolution, by using both the image–based and model–derived T_{ext} as an input. For validation purposes, SEB–1S is first applied to 90 m resolution ASTER data, and the 90 m resolution ET is aggregated at 1 km resolution. The 90 m resolution ET estimates have been validated beforehand with *in situ* measurements (*Merlin*, 2013). The 90 m resolution ET estimates, aggregated at 1 km are chosen as reference to test the impact of T_{ext} in terms of 1 km ET estimates.

The SEB–1S 1 km resolution simulated ET is plotted against the 1 km ET chosen as reference over the Yaqui region, for all seven ASTER overpass dates in Figure 3.6. The reference is obtained by aggregating the 90 m resolution ET estimated by SEB–1S using the Text_RS_90m, given the good results reported when analyzing the 90 m resolution estimates. SEB–1S uses as input either the modeled T_{ext} (both $r_{ah,MO}$ and $r_{ah,RI}$ formulations), or the imaged–based Text_RS_1km.

The mean and standard deviation of the RMSD, slope of the linear regression, bias and correlation coefficient, between simulated and reference ET are shown in Table 3.3, for each T_{ext} algorithm separately. The mean RMSD values are 56 W·m^{-2}, 77 W·m^{-2} and 78 W·m^{-2} for the Text_EBsoil calculated with the MO and RI formulations, and for Text_RS_1km, respectively, with the MO formulation clearly outperforming the other two. However, the lowest mean bias (-15 W·m^{-2}) in ET is obtained for Text_RS_1km compared to 23 W·m^{-2} and -27 W·m^{-2} for Text_EBsoil with the MO and RI formulations, respectively. Both Text_EBsoil algorithms increase the mean correlation coefficient from 0.95 (for Text_RS_1km) to 0.98.

The EBsoil algorithm with the $r_{ah,MO}$ formulation improves ET predictions at 1 km resolution. Nevertheless, a more significant bias can be observed on the ET estimates computed using the MO–derived and RI–derived T_{ext}. Because Text_EBsoil are computed independently of the LST data, this can influence the polygon not to be centered on the cluster points present in the $LST - \alpha$ and $LST - f_{vg}$ spaces, which would introduce a bias in model–derived T_{ext} and, thus on ET. The polygon obtained from Text_RS_1km is however centered on the cluster points, because the computation of T_{ext} is done using the observed LST data points, which lead to the ET estimates derived from Text_RS_1km to present a low bias. On the other hand, the mean slope of the linear regression between modeled and reference ET is estimated as 1.7, explained by the fact that Text_RS_1km systematically overestimates the wet edge and underestimates the

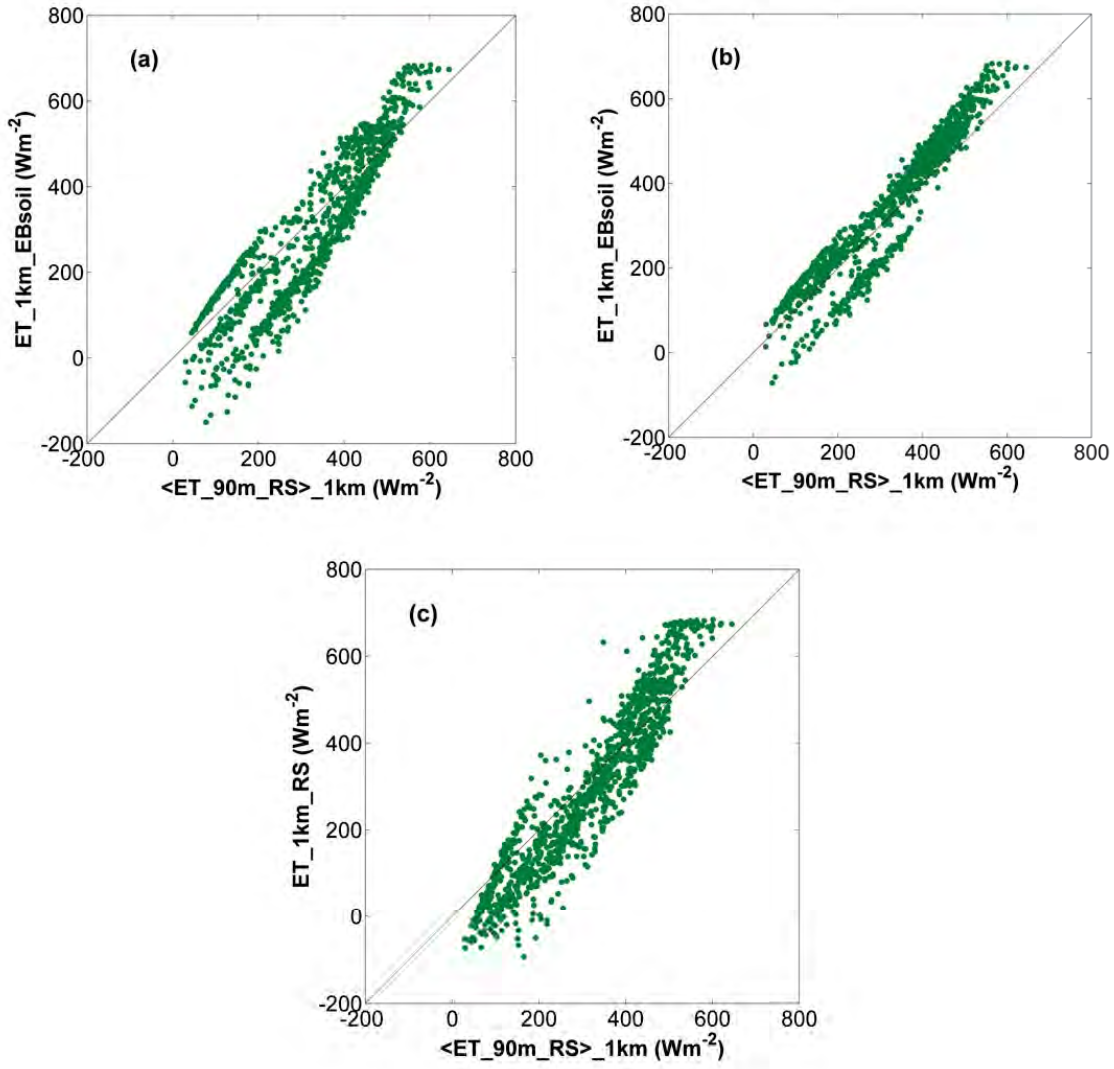


Figure 3.6: 1 km resolution ET estimates *versus* 90 m resolution simulated ET, aggregated at 1 km resolution, for the Yaqui site. The T_{ext} used as input to the SEB–1S consist of the T_{ext} simulated by EBsoil with the RI (**a**) and MO (**b**) aerodynamic resistance, and the T_{ext} retrieved from 1 km resolution data (**c**).

Table 3.3: Mean and standard deviation (σ) of the root mean square difference (RMSD), bias, correlation coefficient (R) and slope of the linear regression between the ET simulated at 1 km resolution and the 90 m resolution simulated ET aggregated at 1 km resolution, for data over the Yaqui region. The T_{ext} used as input to SEB-1S consist of T_{ext} simulated by EBsoil with the RI and MO aerodynamic resistance and the T_{ext} retrieved from 1 km resolution data.

Date	**RMSD ($\mathbf{W{\cdot}m^{-2}}$)**			**Bias ($\mathbf{W{\cdot}m^{-2}}$)**			**R (-)**			**Slope (-)**		
	RI	**MO**	**RS**	**RI**	**MO**	**RS**	**RI**	**MO**	**RS**	**RI**	**MO**	**RS**
All (mean)	77	56	78	−27	23	−15	0.98	0.98	0.95	1.4	1.1	1.7
All (σ)	34	21	17	72	53	44	0.01	0.01	0.01	0.22	0.12	0.19

dry edge, which means that the polygon boundaries are too close to the point cluster. The data points are thus closer to the boundaries at 1 km resolution than at 90 m resolution, influencing the range of variability and leading to slopes significantly larger than 1.

3.5 Summary and Conclusions

A complementary methodology for estimating T_{ext} independently of remote sensing data was developed. Soil T_{ext} are computed using a soil energy balance (EBsoil) model with two different formulations of the aerodynamic resistance: based on the Monin–Obukhov length ($r_{ah,MO}$) and based on the Richardson formulation ($r_{ah,RI}$). Soil T_{ext} are derived from a r_{ah} formulation considered in bare soil conditions only, while the majority of theoretical algorithms take into account r_{ah} formulations for both soil and vegetation *Su* (2002); *Allen et al.* (2007); *Long et al.* (2011); vegetation T_{ext} are derived from soil T_{ext}.

In a first step, a validation of soil T_{ext} calculated by EBsoil with *in situ* measurements made over an irrigated area in Morocco is performed. This is the first time that soil T_{ext} have been evaluated using *in situ* measurements, with results pointing out the $r_{ah,MO}$ formulation as the best choice, with a mean RMSE of 2.4 °C as opposed to 3.6 °C, obtained for the $r_{ah,RI}$ formulation. A important aspect to mention is that during the field experiment that took place in Morocco for the *in situ* validation, by digging the two soil plots, the ground surface structure was changed. This may potentially affect the surface roughness and thus the apparent roughness length, parameters which intervene in the soil energy balance model. However, the soil surface was made as flat as possible to mimic the surrounding bare soil field. As a step further, model–derived soil T_{ext} are then compared against the image–based soil T_{ext} derived directly from SEB-1S. The methodology is applied to 90 m resolution ASTER data over two semi–arid irrigated areas in Mexico and Spain. Results are consistent with those obtained for the *in situ* validation, with the MO formulation outperforming the RI formulation.

A second analysis is undergone regarding ET estimates, by looking at how the two different algorithms of estimating T_{ext} influence ET calculations. SEB–1S is applied to both 90 m resolution and 1 km resolution (aggregated) ASTER data using either the model–derived or image–based T_{ext} as input. The 90 m resolution SEB–1S ET is used as reference against 1 km resolution ET estimates, by simply aggregating it to 1 km. The reported RMSD between 1 km resolution SEB–1S and aggregated 90 m resolution SEB–1S ET is 78 and 56 $\mathrm{W{\cdot}m^{-2}}$, when using the image–based and model–derived T_{ext} (with the $r_{ah,MO}$ formulation), respectively.

A general conclusion that can be drawn by looking at the two different analyses is that T_{ext} should be estimated *a priori* when contextual models are used with low resolution data. Never-

theless, image–based T_{ext} provide accurate ET estimates when high resolution images are used in highly heterogeneous (irrigated) landscapes. Taking this into account, the consistency between model–derived and high resolution image–based T_{ext} provides a meaningful basis for developing mixed modeling observational approaches. The soil energy balance model was successfully implemented in a contextual ET model and it was shown to improve the accuracy in ET estimates at multiple resolutions. Perspective–wise, this could have many potential applications. One could imagine ways of improving the parameterization of the aerodynamical resistance by using remote sensing data, since a good coherence was found between model–derived and image–based T_{ext}. By increasing the accuracy in soil T_{ext}, since the $T_{s,wet}$ and $T_{s,dry}$ are the boundary conditions used for deriving the LST–based SEE, this could provide a means for monitoring SEE by using optical/thermal remote sensing data. In that perspective, SEB–1S can be integrated within DISPATCH, a disaggregation methodology of SMOS soil moisture (see Chapter 5).

3.6 Article

Remote Sens. **2015**, *7*, 10444-10479; doi:10.3390/rs70810444

OPEN ACCESS

remote sensing

ISSN 2072-4292

www.mdpi.com/journal/remotesensing

Article

Consistency between *In Situ*, Model-Derived and High-Resolution-Image-Based Soil Temperature Endmembers: Towards a Robust Data-Based Model for Multi-Resolution Monitoring of Crop Evapotranspiration

Vivien Georgiana Stefan [1,*], Olivier Merlin [1,2], Salah Er-Raki [3], Maria-José Escorihuela [4] and Said Khabba [2]

[1] Centre d'Etudes Spatiales de la Biosphère (CESBIO), 18 avenue Edouard Belin, Toulouse 31401, France; E-Mail: olivier.merlin@cesbio.cnes.fr

[2] Faculté des Sciences Semlalia Marrakech (FSSM), Université Cadi Ayyad, Avenue Prince Moulay Abdellah, BP 2390, Marrakech 40000, Morocco; E-Mail: khabba@uca.ma

[3] Faculté des Sciences et Techniques (FST), Université Cadi Ayyad, Avenue Abdelkarim Khettabi, BP 549, Marrakech 40000, Morocco; E-Mail: s.erraki@uca.ma

[4] isardSAT, Parc Tecnologic BCNord, Marie Curie 8-14, Barcelona 08042, Spain; E-Mail: mj.escorihuela@isardsat.cat

* Author to whom correspondence should be addressed; E-Mail: vivien.stefan@cesbio.cnes.fr; Tel.: +33-5-6155-8829.

Academic Editors: Soe Myint and Prasad S. Thenkabail

Received: 21 April 2015 / Accepted: 31 July 2015 / Published: 14 August 2015

Abstract: Due to their image-based nature, "contextual" approaches are very attractive to estimate evapotranspiration (ET) from remotely-sensed land surface temperature (LST) data. Their application is however limited to highly heterogeneous areas where the soil and vegetation temperature endmembers (Tends) can be observed at the thermal sensor resolution. This paper aims to develop a simple theoretical approach to estimate Tends independently from LST images. Soil Tends are simulated by a soil energy balance model forced by meteorological data. Vegetation Tends are obtained from soil Tends and air temperature. Model-derived soil Tends are first evaluated with *in situ* measurements made over an irrigated area in Morocco. The root mean square difference (RMSD) between modeled and ground-based soil Tends is estimated as 2.4 °C. Model-derived soil Tends are next compared with the soil Tends retrieved from 90-m resolution ASTER (Advanced

Spaceborne Thermal Emission and Reflection Radiometer) data collected over two irrigated areas in Mexico and Spain. Such a comparison reveals a strong consistency between model-derived and high-resolution image-based soil Tends. A recent contextual ET model (SEB-1S) is then applied to 90-m resolution and to 1-km resolution (aggregated) ASTER data using the model-derived or image-based Tends as the input. The RMSD between 90-m resolution SEB-1S and *in situ* ET is estimated as 65 and 82 W·m^{-2}, and the RMSD between 1-km resolution SEB-1S and aggregated SEB-1S ET is estimated as 78 and 56 W·m^{-2} for the image-based and model-derived Tends, respectively. In light of the above results, Tends should be estimated *a priori* when contextual models are applied to low resolution images. Moreover, the consistency over highly heterogeneous areas between model-derived and high-resolution image-based Tends provides a meaningful basis for developing mixed modeling observational approaches.

Keywords: evapotranspiration; energy balance; temperature endmembers; SEB-1S; ASTER; multi-resolution

1. Introduction

Crop evapotranspiration (ET) is by far the main outward water flux over semi-arid irrigated areas. In the Mediterranean countries, for instance, ET can represent up to 80% of the consumptive uses of water [1]. In those regions, accurate estimates of spatially-averaged ET are hence required for efficient crop irrigation and water resources management. Remote sensing is in this regard one of the most promising and cost-effective techniques for mapping and monitoring ET over broad areas. Especially multi-sensor/multi-resolution remote sensing data have a strong potential to provide ET data at multiple scales: the crop field scale where irrigation volumes are allocated and the regional scale where long-term management decisions are made.

Among the considerable variety of existing approaches to estimate ET from remote sensing data, the most widely-used approach is to force the FAO-56 method [2–4] with NDVI (Normalized Difference Vegetation Index) data [5]. It consists of computing ET as a function of the potential ET derived from meteorological measurements, a crop-specific coefficient and the crop phenology estimated from NDVI-derived fractional green vegetation cover (f_{vg}). The FAO-56 method forced by NDVI time series has excellent operational application capabilities, notably due to a remarkable coherence between the complexity of the model parameterization and the availability of input (meteorological and NDVI) data. However, one main limitation of the FAO-56 method is that the coefficient(s) used (either in its single-coefficient or dual-coefficient form) is calibrated using field measurements, which, given the spatial and temporal heterogeneity in these coefficients, restricts its application to large areas [6]. Moreover, the soil moisture constraint on evapotranspiration in the dual-coefficient form is based on a soil water budget model and *a priori* knowledge of irrigation volumes and dates.

Another remote sensing variable relevant to ET monitoring is the land surface temperature (LST). LST is a signature of both ET and soil water availability, especially under non-energy-limited conditions. It

is currently observed by thermal sensors, such as ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer), Landsat-8 and MODIS (Moderate-Resolution Imaging Spectroradiometer) at a spatial resolution of 90 m, 100 m and 1 km, respectively. The operational use of thermal remote sensing for hydrological and water resource management studies has focused on regional-scale applications at the kilometric spatial resolution [7,8]. The reason is that the revisit time of high-spatial resolution sensors (16 days for ASTER and Landsat) is relatively long compared to the quick soil moisture changes that generally occur in irrigated semi-arid regions. The temporal frequency of 1-km resolution MODIS data (one or two times per day) is instead very relevant for monitoring ET on an operational basis.

LST-based ET models fall into two broad ranges of models based on (i) the residual method and (ii) the evaporative fraction (EF, defined as the ratio of latent heat to available energy) method. The first category of models evaluates ET as the residual term of the surface energy balance, which is the available energy at the surface minus the sensible heat flux, e.g., TSEB (Two-Source Energy Balance) [9] and SEBS (Surface Energy Balance System) [10]. The second category evaluates ET as the available energy times EF. The latter models are often called "contextual" because the methodology involves determining the wet and dry boundaries of LST, which may or may not be present at the observation resolution within the observation scene. An overview of these contextual models is presented in [11,12]. In practice, physical boundaries of LST are estimated in a two-step procedure. First, the temperature endmembers (Tends) corresponding to fully dry and wet conditions for both soil/vegetation components are located within the space defined by LST and α [13,14] and/or the space defined by LST and f_{vg} [8,14,15]. Then, the vertices of the obtained polygons are connected by straight lines [16] or curves [17] to interpolate the dry and wet boundaries over the full range of vegetation cover. Once the LST boundaries have been determined, for any data point in the $LST - \alpha$ or in the $LST - f_{vg}$ space, EF is derived as the ratio of the distance separating the point from the line identified as the dry edge to the distance separating the dry edge and the line identified as the wet edge. In general, both submodels are fully independent, meaning that the different Tends algorithms can be implemented using the same EF submodel. The contextual nature of EF-based models is hence attributed to the Tends submodel.

EF-based ET models are very attractive for operational applications. Due to their image-based nature, they are easily transferable to different areas. In particular, they do not require calibrating the, often site-specific, surface resistances of residual-based methods [18]. The downside is the applicability of the image-based Tends submodel, which requires: (i) uniform atmospheric conditions; (ii) a relatively flat area; and above all; (iii) that the temperature endmembers be actually observed within the study domain at the thermal sensor resolution [15,19–23]. In a nutshell, image-based Tends submodels should be quite well adapted for semi-arid irrigated areas provided that the inherent high heterogeneity be resolved at the observation scale, meaning that high-spatial resolution data are used [24]. In contrast, image-based Tends submodels are likely to be inadequate for less heterogeneous rainfed areas and when using medium to low ($\sim$1 km) resolution thermal remote sensing data. Long and Singh offer an in-depth study of how the determination of the image-based Tends impact the EF retrievals in terms of performance and error propagation [21].

To remove some of the above limitations, attempts have been made to de-contextualize Tends submodels, notably by developing Tends algorithms independent of the (contextual) surface conditions within the study area and, hence, of the spatial resolution of input LST data [15,16]. Tends are

thus modeled by running a surface energy balance model forced by meteorological data in both dry and wet conditions and for soil and vegetation components separately. This involves implementing aerodynamic resistances for four extreme surfaces: water-saturated bare soil, dry bare soil, well-watered vegetation, fully-stressed vegetation. Note that the four modeled extreme conditions can be reduced to two—fully dry bare soil and water-stressed vegetation—by setting the cold edge to air temperature for all vegetation covers [15,18]. Whereas the parameterization of bare soil aerodynamic resistance is relatively straightforward, the computation of canopy resistance is more complex, as it involves additional parameters to determine the zero plane displacement, such as vegetation height [15,16,25].

In this context, this study aims to develop a simple theoretical approach to estimate Tends independently from LST images. Soil Tends are simulated by a soil energy balance model, and two formulations of aerodynamic resistance are tested over bare soil. Vegetation Tends are obtained from soil Tends and air temperature. Model-derived soil Tends are first evaluated with the *in situ* measurements made in an irrigated area in Morocco. They are next compared to the soil Tends retrieved from 90-m resolution ASTER data collected over two irrigated areas in Spain and Mexico using an algorithm that combines both $LST - \alpha$ and $LST - f_{vg}$ spaces. Model-derived and image-based Tends are then used as input to a contextual (SEB-1S, [14]) ET model, and results in terms of 90-m resolution and 1-km resolution ET estimates are evaluated over the Mexican site, where ground-based ET measurements are available.

The three study sites in Morocco, Spain and Mexico and their associated experimental data are described in Section 3. In Section 4, the SEB-1S modeling approach is reiterated, and the Tends models forced by either meteorological data or remote sensing data are presented. In Section 5, Tends models are intercompared, and the ET estimates are evaluated at 90-m and 1-km resolutions.

2. Sites and Data Description

The study is based on data collected over three irrigated semi-arid areas: an irrigation perimeter situated in the Haouz plain in central Morocco (hereby referred to as "R3"), the Urgell area in Catalonia, in northeastern Spain, and the Yaqui valley (Sonora State) in northwestern Mexico. A visual representation of the three study sites is provided in Figure 1.

2.1. Site and In Situ Data Description

2.1.1. R3 Perimeter

The Haouz plain in central Morocco covers around 6000 km^2. The climate is semi-arid, typically Mediterranean, with an average annual precipitation of about 250 mm. About 85% of available water is used for agriculture. The R3 irrigated perimeter, situated in the Haouz plain, includes mainly wheat, olive trees, maize and beet. The typical size of crops is 3 up to 4 ha. Wheat is generally sown between mid-November and mid-January, depending on climatic conditions and the start of the rainfall season. Harvest usually occurs about 5 up to 6 months later, in May or June. More details about the description of the R3 site can be found in [26,27].

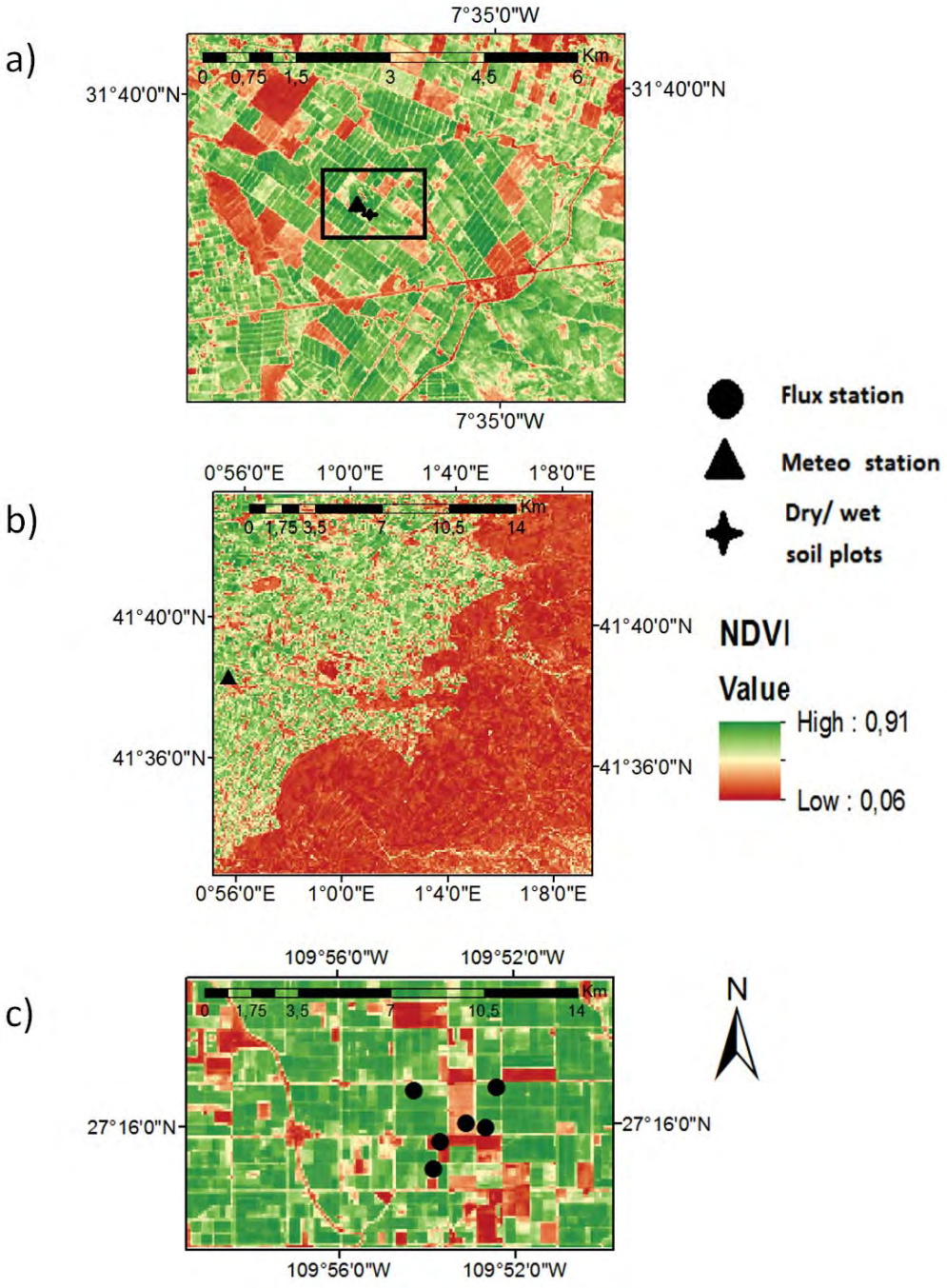


Figure 1. The three study sites included in this paper: (**a**) the R3 perimeter, (**b**) the Urgell area and (**c**) the Yaqui valley.

A field experiment was conducted in a maize parcel of the R3 perimeter on five days: 14 April, 22 April, 30 April, 8 May and 16 May 2014. Maize was seeded in early April, so that the crop field was mostly bare soil during the whole experiment. The experiment focused on measuring the minimum and maximum soil Tends ($T_{s,wet}$ and $T_{s,dry}$) to serve as a reference for the direct validation of Tends modeling approaches using meteorological data. Weather conditions were mostly sunny, with one cloudy

day (16 May) and one post-rainy day (22 April). The different atmospheric conditions met throughout the campaign ensured testing the model in various conditions. The temperature measurements were taken over two small soil plots (2 m by 2 m). One soil plot was kept fully dry, whereas the soil moisture of the other soil plot was maintained equal to the soil moisture at saturation. The dry soil conditions were fulfilled by isolating the respective soil plot with a plastic sheeting, buried at a 15-cm depth to prevent water infiltrations underneath from capillary rising. The dry plot was also covered by plastic sheeting between two consecutive sampling days to protect the soil from possible rainfall events before the experiments took place. On each sampling day, the saturated soil plot was manually and continuously irrigated from about 10–20 min before the start of the temperature measurements and until the time of the last measurement. Much attention was given to maintain the soil plot as wet: irrigation water was poured in sufficient quantity to reach soil saturation, but as slow as possible to avoid flooding and runoff. Figure 2 shows an illustration of the two 2 m by 2 m soil plots where the experiment was undertaken. Note that the soil surface structure was not preserved during the re-filling of the soil pits, but the soil surface was made relatively flat to mimic the surrounding (bare soil) corn field. Continuous measurements of $T_{s,wet}$ and $T_{s,dry}$ were taken between 10:00 a.m. and 11:30 a.m. UTC (Coordinated Universal Time). The observation time is compatible with the overpass times of ASTER, Landsat-7/Landsat-8 and MODIS thermal sensors. The soil temperature for dry and wet soils were monitored using DS1921G-F5 Thermochron iButton sensors. These sensors are rugged, low-cost and self-sufficient systems that record and store time and temperature. They are equipped with a wide (from -40–85 °C) temperature range and fit for a moderately long (from 1–255 min) logging rate. Their accuracy spans from $\pm$ 1 °C in the -30–70 °C range to $\pm$ 1.3 °C outside this range, with a field resolution of 0.5 °C ([28]). In order to increase the confidence in the experimental data and to have a better estimate of the uncertainty in measurements, several Thermochron iButtons were set up at each soil plot. In practice, iButtons were fixed on 7 cm by 7 cm acrylic glass plates, for a set of three iButtons per plate. Because more variability was found in the dry soil measurements, two plates (six iButtons) were used to measure $T_{s,dry}$, whereas one plate was used to estimate $T_{s,wet}$. The plates were buried very close to the soil surface, by making sure that iButtons were not directly exposed to solar radiation. Thermochron data were logged at a 5-min time interval. Entries registered by all Thermochron iButtons were averaged per soil plot, and per 30-min bin, cumulated at 10:30 a.m., 11:00 a.m. and 11:30 a.m. UTC. This resulted in a total of three sets ($T_{s,wet}$ and $T_{s,dry}$) of values per day.

In addition to soil temperature measurements, meteorological data, such as air temperature, relative air humidity, solar radiation and wind speed, were acquired every half hour by a permanent meteorological station located in an alfalfa crop field about 200 m from the dry/wet soil plots. The acquisition height of the meteorological measurements was 2 m.

Figure 3 illustrates the mean (the average of Thermochron iButton entries) soil temperature measurements for wet and dry conditions ($T_{s,wet}$ and $T_{s,dry}$, respectively) obtained on 8 May. The reason why the wet and dry temperatures are very similar before 10:00 a.m. is that the sensors were set up right before and removed right after each field experiment. They were not measuring soil temperatures corresponding to wet and dry conditions, but they were stored in the same media (air) and, thus, logging similar temperatures. Soil Tends measurements perfectly match until 30 min before

the start of the experiment, which reflects excellent confidence in the iButtons' embedded calibration. By computing the average on each soil plot of the standard deviation of instantaneous measurements during the experiment duration, the uncertainty in $T_{s,wet}$ and $T_{s,dry}$ is estimated as 0.90 °C and 1.1 °C for wet and dry conditions, respectively.



Figure 2. Illustration of the dry soil plot in the R3 perimeter. (**a**) The pit was dug at a 15-cm depth and (**b**) the soil plot was kept fully dry by covering it with a plastic sheeting in between experiments.

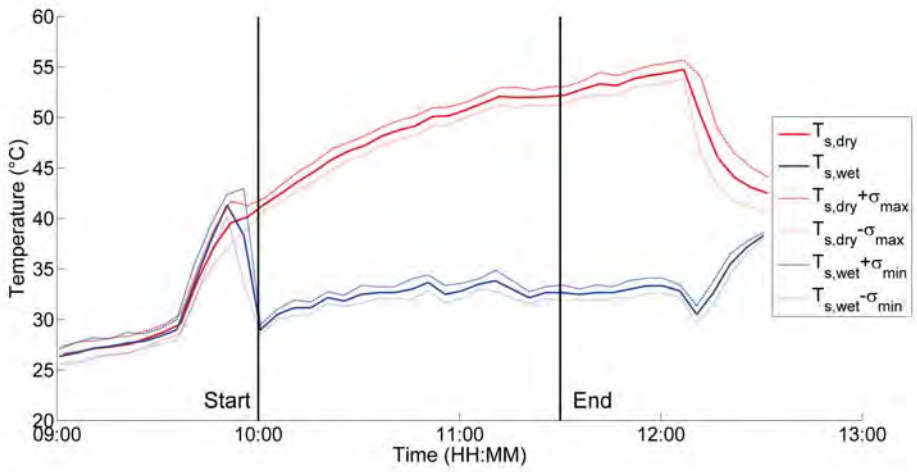


Figure 3. Mean and standard deviation of the *in situ* measurements performed on 8 May on the dry and wet soil plots of the R3 perimeter, Haouz plain. The two vertical lines stand for the beginning and the end of the experiment.

2.1.2. Urgell Site

The Urgell area presents a semi-arid continental Mediterranean climate, with a mean yearly air temperature of 16 °C, precipitation of 400 mm and a number of days with rain of 60. The area of interest includes both irrigated crops consisting of wheat, maize, alfalfa and various fruit (apple and pear) trees, as well as dryland crops, such as barley, olive trees, vineyards and almond trees [29]. The Urgell experiment, conducted in 2011, focused on a 20 km by 20 km area, situated within the Urgell

irrigation region. Meteorological data are continuously acquired over the area. The meteorological data were collected by the automatic weather station of Golmes (WC) as part of the XEMA network operated by the Meteorological Service of Catalonia (SMC). Air temperature, relative air humidity, solar radiation and wind speed were recorded every 30 min at a 2-m height by the Golmes (WC) station.

2.1.3. Yaqui Site

The Yaqui valley in northwestern Mexico presents a semi-arid climate, with an average annual potential evapotranspiration of 2233 mm and an annual precipitation of 290 mm [30]. The main crops include wheat (about 50%), broccoli, beans, chili pepper, potatoes, chickpea, safflower, orange and corn. Irrigation takes up to 90% of the total water expenditure.

The experiment was performed from December 2007–May 2008 on an irrigated area situated in the Yaqui valley. Measurements were focused on a 4 km by 4 km area, while the study area in this paper is defined as a 16 km by 10 km area containing the 4 km by 4 km Yaqui experimental area. Seven micro-meteorological stations equipped with an eddy-covariance flux measurement system were set up in different fields [14]. The six stations, providing data concurrently, with at least four ASTER overpasses, are used in the ET validation analysis. The latent heat flux was measured with KH20 fast response hygrometers at a frequency of 10 Hz and converted to the 30-min average. Meteorological data, including air temperature, relative air humidity, solar radiation and wind speed, were acquired every half hour, from 27 December 2007–17 May 2008 from a station located in the center of the region. The acquisition height was 10 m. Further details on the experiment can be found in [14,30–32].

2.2. Remote Sensing Data

2.2.1. ASTER

ASTER is an imaging instrument on-board NASA's Terra satellite, launched in 1999 and having an equator crossing time at 11 a.m., with a 16-day revisit cycle. The ASTER thermal sensor has five thermal bands, with a spatial resolution of 90 m. ASTER official surface kinetic temperature products (AST_08) were downloaded from the Earth Observing System Data Gateway and extracted over the 16 km by 10 km Yaqui and 20 km by 20 km Urgell areas. Two cloud-free ASTER images were acquired over the Urgell site at around 11:00 a.m. UTC on 16 August and 3 October 2011. Seven cloud-free ASTER images were obtained over the Yaqui area during the agricultural season of 2007–2008, at around 11:30 a.m. local solar time on 30 December, 23 February, 10 March, 11 April, 27 April, 6 May and 13 May.

In order to test the contextual ET models at multiple resolutions, ASTER data are also used at a 1-km (MODIS thermal) resolution by aggregating the original 90-m resolution ASTER temperature using a simple average [31].

2.2.2. Formosat-2

The Formosat-2 satellite is an Earth observation satellite of the National Space Organization of Taiwan, launched in 2004. Its high (8 m) resolution images are provided daily at 9:30 a.m.

equator-crossing time. The reason behind using Formosat-2 derived albedo resides in the unusability of the ASTER shortwave infrared data on its seven overpass days over the Yaqui site [30]. Its high temporal resolution (one day revisit cycle) enabled us to use data collected on the nearest date from each of the seven ASTER overpass dates in order to estimate both NDVI and α from the red and near-infrared reflectances, aggregated at ASTER thermal sensor resolution.

2.2.3. LST, Fractional Green Vegetation Cover, NDVI and Surface Albedo

In order to retrieve the LST, the "temperature and emissivity separation" algorithm [33,34] was applied.

Fractional green (photosynthetically active) vegetation cover and surface albedo are derived over the Urgell and Yaqui study areas from satellite surface reflectances in the red and near-infrared bands. For the Urgell area, reflectance data are obtained from 15-m resolution ASTER red and near-infrared data on its two overpass days. For the Yaqui area, reflectance data are obtained from Formosat-2 data, available on dates closest to the ASTER overpass dates. In both cases, surface reflectances are aggregated from the observation resolution to ASTER thermal resolution (90 m) beforehand.

The Gutman and Ignatov (1998) formula is applied in order to obtain the fractional green vegetation cover (f_{vg}):

$$f_{vg} = \frac{\text{NDVI} - \text{NDVI}_s}{\text{NDVI}_{vg} - \text{NDVI}_s} \quad (1)$$

with NDVI_{vg} corresponding to fully-covering green vegetation and NDVI_s to bare soil or to bare soil partially covered by senescent (non-photosynthetically active) vegetation. In this study, NDVI_{vg} and NDVI_s are set to the maximum and minimum value of the NDVI observed during the agricultural season within the study domain. NDVI endmembers are estimated as 0.93 and 0.018, and 0.93 and 0.18 for the Urgell and Yaqui areas, respectively. The reason behind the small value registered for NDVI_s for Urgell is the presence of open water (reservoir of irrigation water) in the region. NDVI is computed as the ratio of the difference between (re-sampled) near-infrared and red reflectances to their sum.

Surface albedo (α) is estimated as a weighted sum of (re-sampled) red and near-infrared reflectances, using the coefficients provided by [35] and corroborated by [36] and in [30].

3. Models

The main goal of this study is to assess the impact of different modeling approaches of Tends on the ET estimated at multiple resolutions by a contextual model. Herein, the contextual model used for the estimation of ET is the recently-developed SEB-1S model. Like a number of contextual models, SEB-1S computes ET as EF times the available energy, with EF being estimated as the ratio of the distance separating a pixel in the $LST - \alpha$ space from the line defined as the dry edge to the distance separating the dry edge and the line identified as the wet edge. The dry and wet edges are defined from Tends.

Two different approaches are investigated to estimate Tends using (1) available remotely-sensed data (referred to as Tends_RS) and (2) a soil energy balance model forced by available meteorological data

(referred to as Tends_EBsoil). The energy balance model offers the possibility to derive Tends estimates independently of surface conditions and of the spatial resolution of remote sensing data.

3.1. SEB-1S: An EF Model (EF_SEB-1S)

SEB-1S relies on the same $LST - \alpha$ space as the one used in the Simplified Surface Energy Balance Index (S-SEBI, [13]). It computes EF similarly to S-SEBI, as the ratio of the distance separating the point from the dry edge to the distance separating the wet and dry edges [37–42]. However, SEB-1S provides a different interpretation of the wet edge, which has proven to significantly improve ET estimates over the Yaqui area [14]. Figure 4 presents a graphic illustration of the computation methodology applied in SEB-1S.

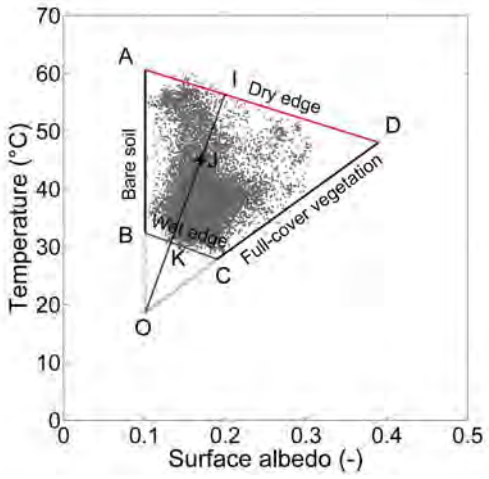


Figure 4. For a given J point, EF is computed as the ratio of IJ to IK. Underlying grey points correspond to 90-m resolution data acquired on 27 April, over the Yaqui area.

ET is defined as EF times the available energy ($Rn - G$), where Rn (W·m^{-2}) is the surface net radiation and G (W·m^{-2}) the ground heat flux. Further details describing the methodology can be found in [14].

3.2. Image-Based Tends Algorithm (Tends_RS)

The Tends_RS approach provides soil and vegetation Tends from a synergy between the $LST - \alpha$ space and the $LST - f_{vg}$ space, as in the original version of the SEB-1S Tends submodule. Note that previous studies [8,15,19,20,43] use only one space in the determination of the extreme temperatures. Figure 5 illustrates side-by-side the SEB-1S interpretation of the edges and vertices of the polygon in the $LST - \alpha$ space and the polygon in the $LST - f_{vg}$ space for data collected on 11 April, over the Yaqui site.

In the $LST - \alpha$ space, SEB-1S identifies four Tends and three α endmembers (α_ends), which correspond to bare soil (α_s), green vegetation (α_{vg}) and senescent vegetation (α_{vs}) albedo. The edges are interpreted as follows: [AB] as "bare soil", [BC] as "wet surface", [CD] as "full-cover vegetation"

and [DA] as "dry surface". In the $LST - f_{vg}$ space, SEB-1S identifies the same four Tends and both f_{vg} endmembers (0 and 1). The edges are interpreted as follows: [AB] as "mixed soil and senescent vegetation", [BC] as "wet surface", [CD] as "full-cover green vegetation" and [DA] as "dry surface". The dry edge [AD] has a negative slope in both S-SEBI [13] and SEB-1S [14] models due to a radiative and convective control in dry conditions, respectively. In SEB-1S, the wet edge BC identified in the $LST - \alpha$ space has a negative slope, also, due to a convective control in wet conditions [14], consistent with the common interpretation of the $LST - f_{vg}$ space.

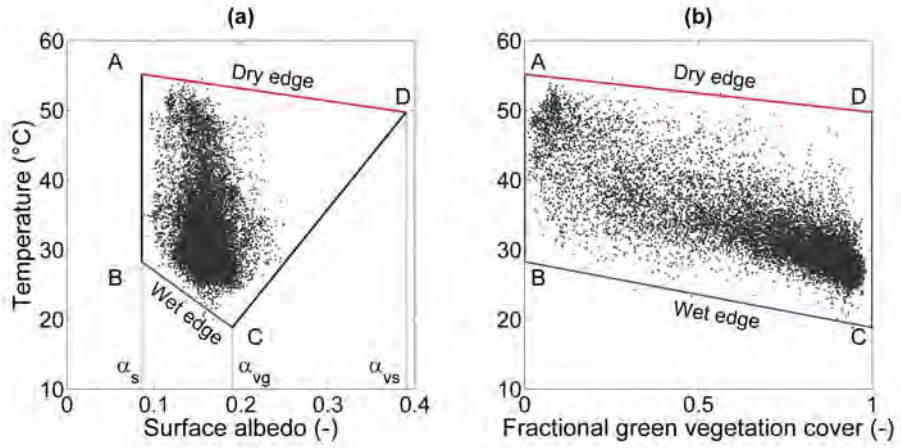


Figure 5. (**a**) Interpretation of the edges and vertices of the $LST - \alpha$ and (**b**) the $LST - f_{vg}$ polygons. Underlying points correspond to 90-m resolution data acquired on 11 April, over the Yaqui site.

As in the original version of the SEB-1S Tends submodule, the minimum vegetation temperature $T_{v,wet}$ is set to the air temperature T_a. The maximum soil temperature $T_{s,dry}$ is the same for both spaces and is set to the maximum surface temperature T_{max}. The other Tends are estimated by combining the Tends retrieved from the $LST - \alpha$ and $LST - f_{vg}$ spaces separately. The minimum soil temperature for each space is computed by extrapolating the wet edge passing through the well-watered full-cover green vegetation index until the bare soil line. The maximum vegetation temperature is computed such that all data points are located in between the dry [AD] and the wet [BC] edges. This process is first undertaken for each space separately and then sets $T_{v,dry}$ to the average of the two retrieved values. Please refer to Appendix A for further details concerning the original version of the algorithm and the adjustments made to it.

3.3. Model-Derived Tends (Tends_EBsoil)

A different methodology is proposed to estimate soil and vegetation Tends independently of remotely-sensed data. A soil energy balance model forced by available meteorological data is used to derive soil Tends. Two simple assumptions are then made for deriving vegetation Tends from model-derived soil Tends. Alternative to the use of a complete surface energy balance model for both soil and vegetation components, the energy balance is modeled herein for bare soil conditions only

(and not for full-cover vegetation). The idea is to develop a meteorological-based approach for Tends determination using a minimum number of *a priori* input parameters.

3.3.1. Soil Energy Balance Model

The maximum and minimum soil temperatures are modeled by setting the near-surface soil moisture to zero and to the soil moisture at saturation respectively, as an input to an energy balance model for bare soils (EBsoil). In practice, EBsoil operates an iterative loop on the soil temperature T_s (initialized to T_a), until thermal equilibrium is established [9]:

$$Rn_s - G = H_s + LE_s \tag{2}$$

with H_s (W·m^{-2}) being the soil sensible heat flux and LE_s (W·m^{-2}) the soil latent heat flux.

The latent heat flux is calculated as:

$$LE_s = \frac{\rho C_p}{\gamma} \frac{e_{sat}(T_s) - e_{sat}(T_a)}{r_{ss} + r_{ah}} \tag{3}$$

with ρ (kg·m^{-3}) being the air density, C_p (J·kg^{-1}·K^{-1}) the specific heat of air at constant pressure, γ (Pa·K^{-1}) the psychrometric constant, $e_{sat}(T_s)$ (Pa) the saturated vapor pressure at soil temperature, $e_{sat}(T_a)$ (Pa) the saturated vapor pressure at air temperature, r_{ss} (s·m^{-1}) the soil evaporation resistance and r_{ah} (s·m^{-1}) the aerodynamic resistance to heat transfer.

r_{ss} is given by the formulation in [44], as a function of soil moisture:

$$r_{ss} = exp(A - B\frac{SM}{SM_{fc}}) \tag{4}$$

where the two best fit parameters A (unitless) and B (unitless) are considered as 8 and 5, respectively [45,46]. SM represents the surface (0–5 cm) soil moisture, whereas SM_{fc} the soil moisture at field capacity.

Given the constraint of Equation (2), EBsoil computes $T_{s,dry}$ and $T_{s,wet}$ by setting SM to 0 and SM_{sat}, respectively.

3.3.2. Aerodynamic Resistance Formulations for Bare Soil

The aerodynamic resistance that intervenes in the computation of the sensible heat flux, as well the latent heat flux (see Appendix B) plays an essential role in the prediction of soil Tends. Different parameterizations have been proposed based on the Monin-Obukhov (MO) similarity theory [47–50]. These parameterizations fall within three different categories: those that follow the MO similarity theory, those that are based on empirical methods and, finally, the so-called semi-empirical ones [51]. Different formulations were developed with the purpose of saving computation time, by introducing various assumptions and simplifications in regards to the stability effects, which perform differently when comparing estimates to measurements of r_{ah} [51].

Under stable conditions, an exact solution of the aerodynamic resistance can be easily obtained due to the linearity of the functions of the stability parameters. However, under non-stable conditions, these

functions are highly non-linear, and an iterative algorithm is used in order to obtain the exact solution. This method is known as the "MO similarity" method.

To assess the impact of r_{ah} on soil Tends, two different formulations are implemented within EBsoil. The first formulation RI we consider in our implementation takes into account the Richardson number R_i (unitless), which represents the importance of natural relative to forced convection and is expressed as in [48]:

$$r_{ah,RI} = \frac{r_{ah,SI}}{(1+R_i)^{\eta}} \tag{5}$$

with η (unitless) being a coefficient equal to 0.75 in unstable conditions ($T_s > T_a$) and to 2 in stable conditions ($T_s < T_a$). We define a simple formulation, based on an "empirical" method, which represents the aerodynamic resistance that neglects natural convection:

$$r_{ah,SI} = \frac{log^2 \frac{Z_r}{Z_{0m}}}{k^2 u_a} \tag{6}$$

with Z_{0m} (m) the roughness length for momentum transfer over bare soil, Z_r (m) the reference height at which the wind speed is measured, k (unitless) the von Kármán constant and u_a (m·s^{-2}) the wind speed.

The bulk Richardson number is a measure of the influence of atmospheric stability on the flux-gradient relationship in the surface layer:

$$R_i = \frac{5gZ_r(T_s - T_a)}{T_a u_a^2} \tag{7}$$

with g (m·s^{-2}) the gravitational constant. This method is referred herein to as the "empirical" method.

The second formulation takes into account the MO length, and is computed as:

$$r_{ah,MO} = \frac{\log \frac{Z_r}{Z_{0m}} - \psi_h}{k u_*} \tag{8}$$

with ψ_h (unitless) the stability correction factor for heat transport and u_* (m·s^{-1}) the friction velocity.

For further details on the derivation of r_{ah} from the MO theory, please refer to Appendix C.

3.3.3. Vegetation Tends

Two assumptions are made for deriving vegetation Tends from model-derived soil Tends.

The minimum vegetation temperature $T_{v,wet}$ corresponding to well-watered unstressed vegetation is set to the air temperature [52–54]. This hypothesis is common to many models, such as SEBAL (Surface Energy Balance Algorithm for Land) [54], S-SEBI [13] and SEB-1S [14]. The work in [55] considers $T_{v,wet}$ as a fraction of the air temperature, using a correction factor determined as 0.993. This correction factor was introduced after observing a systematic bias between T_s and T_a of well-watered vegetation.

The maximum vegetation temperature $T_{v,dry}$ corresponding to fully water-stressed vegetation is estimated as:

$$T_{v,dry} = T_{s,dry} - (T_{s,wet} - T_{v,wet}) \tag{9}$$

by assuming that the dry and wet edges in the $LST - f_{vg}$ space are parallel. This is based on the observation that the lines [AD] and [BC] in the $LST - f_{vg}$ space are practically parallel, especially when using high-resolution remote sensing data in highly heterogeneous conditions [14]. From a physical point of view, it is assumed that the water-saturated bare soil to full-cover unstressed vegetation aerodynamic resistance difference is approximately the same as the bone dry bare soil to full-cover water stressed vegetation aerodynamic resistance difference. In other words, the impact of water stress on vegetation height is neglected, so that the decreasing rate in surface temperature with increasing vegetation cover (and the related increase in roughness length for momentum transfer from bare soil to full-cover vegetation) is practically similar for both dry and wet surface edges.

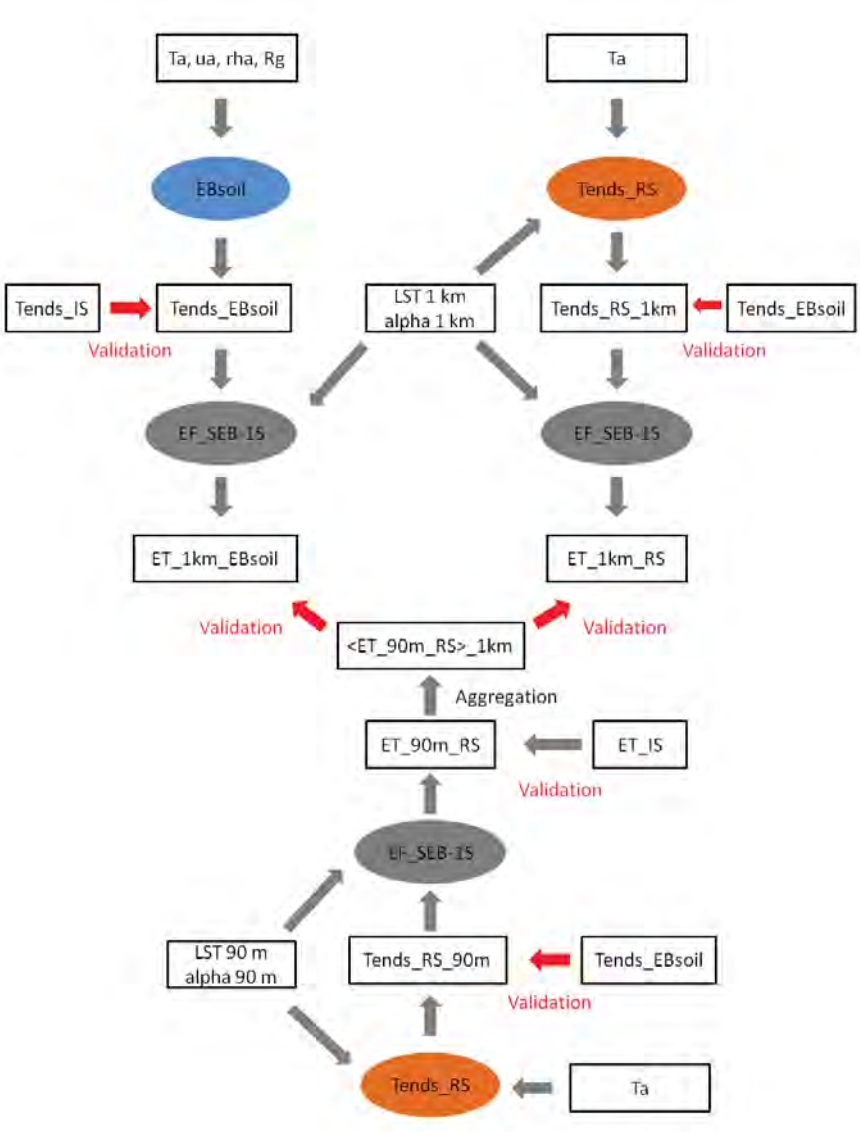


Figure 6. Schematic diagram presenting an overview of the models used (including inputs, outputs and the relationships between them) and the validation strategy of temperature endmembers (Tends) and ET estimates.

Note that by setting the minimum vegetation temperature to T_a, the roughness length for momentum transfer over full-cover unstressed vegetation is implicitly constrained. Equation 9 represents a similar constraint on the roughness length over full-cover water-stressed vegetation (for $T_{v,dry}$ estimation), which is fully consistent with the $T_{v,min} = T_a$ strategy.

Deriving vegetation Tends from soil Tends and T_a data provides the advantage of not requiring the vegetation height information, as well as the zero plane displacement, the roughness length for momentum transfer and the roughness length for heat transfer. The proposed approach only relies on the roughness length for momentum transfer (Z_{0m}) over bare soils, which can be uniformly set to 0.001 m [56]. The range of values of Z_{0m} is typically considered between 0.0006 and 0.035 m [15,22,43,56,57].

3.4. Validation Strategy

Two validation strategies are applied, in terms of both Tends and ET estimates. Figure 6 offers a schematic diagram of the models used, including corresponding inputs and outputs, alongside the performed validations. First, the soil Tends simulated by EBsoil (soil Tends_EBsoil) are compared to the soil Tends measured *in situ* (soil Tends_IS). Second, the soil Tends retrieved from remote sensing data at 90-m resolution (Tends_RS_90m) are compared to the soil Tends_EBsoil. Third, the soil Tends retrieved from remote sensing data at 1-km resolution (Tends_RS_1km) are compared to the soil Tends_EBsoil.

Concerning ET estimates, first of all, a validation at 90-m resolution with *in situ* measurements is presented. SEB-1S is then applied to 1-km resolution data, which were derived from the aggregation (simple average) of 90-m resolution data. This 1-km resolution modeled ET is compared against the ET simulated by SEB-1S at the original 90-m ASTER resolution and subsequently aggregated to 1 km.

4. Results and Discussion

In this section, the soil Tends simulated by the soil energy balance model (Tends_EBsoil) using the RI or the MO r_{ah} formulation are evaluated against: (1) the *in situ* measurements (Tends_IS) in the R3 area; and (2) the soil Tends retrieved from 90-m resolution ASTER data (Tends_RS_90m) in the Yaqui and Urgell areas. The model-derived and image-based Tends are then used separately as input to SEB-1S when applied to 90-m resolution and to 1-km resolution (aggregated) ASTER data. The impact of Tends on contextual ET estimates is finally discussed in terms of surface conditions and observation resolution.

4.1. Modeled Soil Tends

4.1.1. Direct Validation Using *In Situ* Measurements

When comparing the two sets of results, it is clear that the latter gives the best results in terms of root mean square error (RMSE), bias, slope of linear regression and correlation coefficient (see Table 1). The mean RMSE is equal to 2.4 °C for the MO formulation, compared to 3.6 °C for the RI formulation. The lowest mean bias is also obtained for the MO formulation, -1.6 °C, as opposed to -2.2 °C for the RI formulation. The mean correlation is increased from 0.95 (RI) up to 0.97 (MO formulation, respectively).

It is found (not shown) that $T_{s,dry}$ accounts for more bias than $T_{s,wet}$, and its correlation with *in situ* measurements is slightly weaker than for $T_{s,wet}$. This is possibly due to the $T_{s,dry}$ measurements, which present greater variability than the $T_{s,wet}$ measurements, as explained previously.

Table 1. Root mean square error (RMSE), bias, correlation coefficient (R) and slope of the linear regression between soil Tends simulated by the energy balance model for bare soils (EBsoil) with the RI and MO aerodynamic resistance and the soil Tends observed *in situ* at the soil plots of the R3 perimeter.

Time	RMSE (°C)		Bias (°C)		R (-)		Slope (-)	
	RI	MO	RI	MO	RI	MO	RI	MO
10:30 a.m.	3.2	2.3	−1.8	−1.2	0.93	0.96	0.82	0.99
11:00 a.m.	3.3	2.3	−2.2	−1.7	0.97	0.98	0.79	0.94
11:30 a.m.	4.2	2.7	−2.5	−1.9	0.95	0.98	0.74	0.90
Mean	3.6	2.4	−2.2	−1.6	0.95	0.97	0.78	0.94

Figure 7a,b plot the soil Tends modeled using the RI and MO version of r_{ah}, respectively.

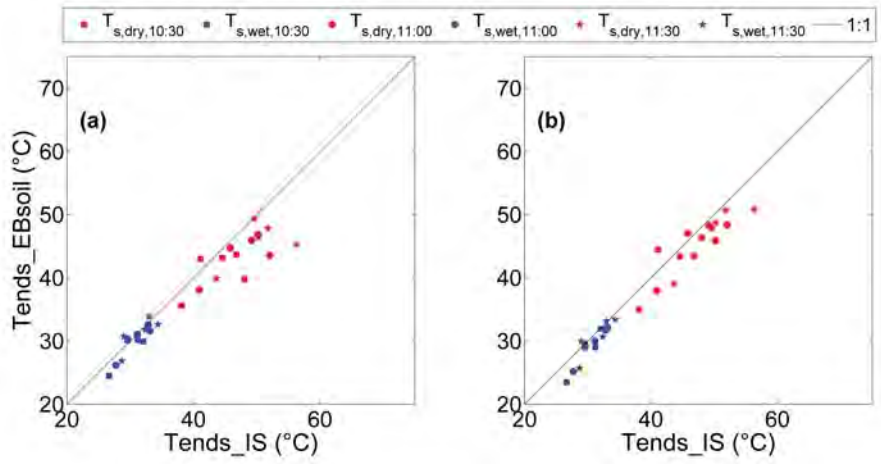


Figure 7. The soil Tends simulated by EBsoil for the (**a**) RI and (**b**) MO aerodynamic resistance are plotted against the soil Tends observed *in situ* at the two soil plots in the R3 perimeter.

As the calculation of Tends depends on r_{ah}, which depends on wind speed, a sensitivity study of simulated r_{ah} with respect to wind speed is also conducted (see Figure 8). It appears that a higher sensitivity is obtained for wind speeds inferior to 2 m·s^{-1} and that $T_{s,dry}$ is impacted slightly more than $T_{s,wet}$. Both RI and MO formulations are similar in terms of $T_{s,wet}$ sensibility to wind speed. However, this is not the case for $T_{s,dry}$, as the RI formulation is more sensitive to wind speeds inferior to 1 m·s^{-1}. As wind speed decreases from 2 m·s^{-1}, the resistance to heat transfer becomes very large and is no longer representative of the turbulent processes occurring.

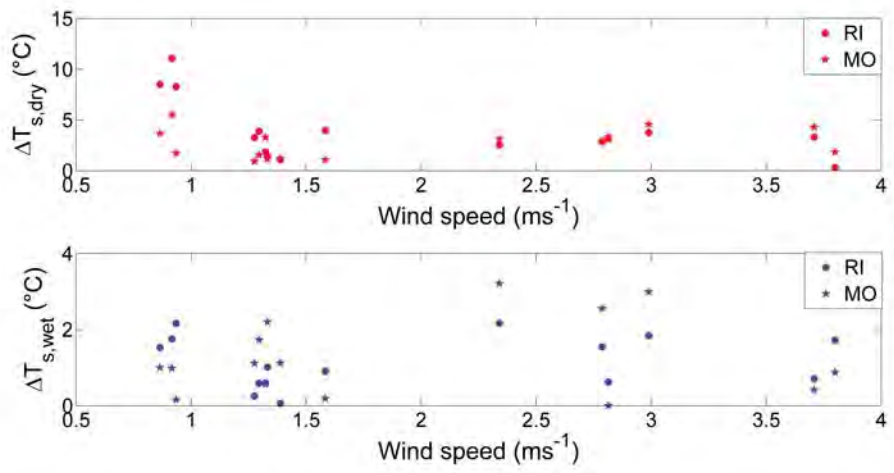


Figure 8. Absolute value of the simulated to observed maximum (**top**) and minimum (**bottom**) soil temperature difference as a function of wind speed for the RI and MO aerodynamic resistance separately.

Results are consistent with those obtained in [57], stating that one of the parameters contributing to important uncertainties in the estimation of aerodynamic resistances is the wind speed.

As a summary, the soil Tends simulated using the MO formulation compare quite well to *in situ* measurements and seem to be relatively stable even in the conditions of low wind speeds. The uncertainty in modeled $T_{s,dry}$ is larger than in modeled $T_{s,wet}$, consistent with a significantly larger range of variation of $T_{s,dry}$.

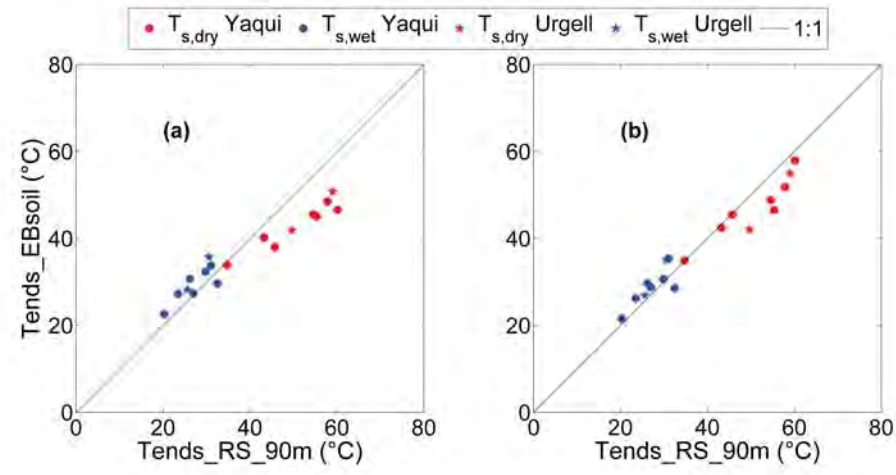


Figure 9. The soil Tends simulated by EBsoil for the (**a**) RI and (**b**) MO aerodynamic resistance are plotted against the image-based Tends for the Yaqui and Urgell regions separately.

4.1.2. Consistency with Remotely-Sensed Soil Tends

The proposed two r_{ah} formulations are now implemented over the Urgell and Yaqui areas. The soil Tends issued from the physical aerodynamic resistance model are compared to the soil Tends derived by the SEB-1S Tends submodel from 90-m resolution ASTER data. Figure 9 plots model-derived (Tends_EBsoil) *versus* image-based (Tends_RS_90m) soil Tends for the Yaqui and Urgell areas separately.

Statistical results in terms of daily bias in $T_{s,wet}$ and $T_{s,dry}$ are provided in Table 2, while values of the slope of the linear regression, correlation coefficient and RMSD between modeled and remotely-sensed Tends can be found in Table 3, for each r_{ah} formulation. When analyzing the results obtained over the Yaqui and Urgell areas combined, the MO r_{ah} formulation gives the best overall statistical results, with a RMSD estimated as 4.1 °C, compared to 6.4 °C for the RI formulation. The consistency obtained between model-derived and image-based soil Tends reflects the robustness and precision of the image-based Tends_RS_90m over highly heterogeneous irrigated areas. Such a consistency is quite satisfying, as it validates both the physical approach based on the MO theory and the remote sensing Tends algorithm based on both $LST - \alpha$ and $LST - f_{vg}$ spaces.

Table 2. Difference between the soil Tends simulated by EBsoil with the RI and MO aerodynamic resistance and the soil Tends retrieved from 90-m resolution data for the minimum and maximum soil Tends, for the Yaqui and Urgell regions separately.

Day	Site	Bias $T_{s,dry}$ (°C)		Bias $T_{s,wet}$ (°C)	
		RI	MO	RI	MO
30 December	Yaqui	−0.84	0.23	2.4	1.3
23 February	Yaqui	−3.0	−0.80	3.8	2.8
10 March	Yaqui	−7.7	−0.21	0.40	1.8
11 April	Yaqui	−9.0	−5.7	4.6	3.6
27 April	Yaqui	−14	−2.2	2.8	4.3
6 May	Yaqui	−10	−8.8	−2.8	− 3.9
13 May	Yaqui	−9.3	−6.0	2.6	0.90
16 August	Urgell	−8.3	−4.0	5.2	4.4
3 October	Urgell	−7.8	−7.6	2.6	1.3
All (mean)	Yaqui	−7.7	−3.4	2.0	1.5
All (σ)	Yaqui	4.4	3.5	2.5	2.7
All (mean)	Urgell	−8.0	−5.8	3.9	2.8

The approach is also implemented using 1-km resolution (aggregated) ASTER data over the two sites. Figure 10 shows side by side the 90-m resolution (grey dots) and the 1-km resolution (black dots) $LST - \alpha$ and $LST - f_{vg}$ spaces plotted for three ASTER overpass dates over the Yaqui site (30 December, 11 April and 6 May, respectively). The polygons obtained using Tends_EBsoil (RI and MO formulations), Tends_RS_90m and Tends_RS_1km are overlaid on each space. The temporal variability

in the boundaries of the two spaces can be explained by the presence or absence of vegetation and by the change in vegetation color during the senescence period, as noted in [14]. Nonetheless, one observes that the contour of 90-m resolution data points is relatively well represented by the Tends_RS_90m polygons. The other three polygons (Tends_RS_1km and Tends_EBsoil for RI and MO formulations) are significantly different, especially in the $LST - \alpha$ space. The albedo endmembers for Tends_EBsoil and Tends_RS_1km are derived from 1-km resolution ASTER data. This is the reason why the three polygons are narrower and closer to the 1-km resolution data points. However, the MO r_{ah} formulation provides dry and wet edges closer to the Tends_RS_90m polygon than the RI r_{ah} formulation (consistent with a better performance in terms of soil Tends). Moreover, the MO r_{ah} significantly outperforms the Tends_RS_1km algorithm, which systematically underestimates the dry edge and overestimates the wet edge. Consequently, one concludes that the major factor that contributes to the estimation of physics-based Tends is the aerodynamic resistance and that the MO formulation appears to be a better candidate than the RI formulation, especially for the dry edge determination. As the spatial resolution of observation decreases, so does the number of data points within both $LST - \alpha$ and $LST - f_{vg}$ spaces. A small number of data points represents a strong constraint on the image-based determination of Tends. In fact, a larger domain size would be needed at 1-km resolution to capture more variability in surface (especially dry and wet) conditions. However, the domain size is another strong constraint on the contextual Tends algorithms, because the application of image-based methods implicitly assumes that meteorological conditions are uniform within the study domain. In particular, Tends are by definition zero-dimensional, that is Tends are not scale-dependent. Therefore, the relevance of image-based methods over a given application domain is a compromise between an extent sufficiently large to meet the heterogeneous requirement (so that dry and wet conditions can really be observed) and an extent sufficiently small to meet the uniform atmospheric forcing requirement (so that Tends are uniform, too). In this paper, image-based algorithms were applied to $\sim$10 km-sized areas to limit the variability of atmospheric conditions within the study domain.

Table 3. Root mean square difference (RMSD), correlation coefficient (R) and slope of the linear regression between the soil Tends simulated by EBsoil for the RI and MO aerodynamic resistance retrieved from 1-m resolution data (RS_1km) and the soil Tends retrieved from 90-m resolution data, for the Yaqui and Urgell regions combined.

RMSD (°C)			**R (-)**			**Slope (-)**		
RI	**MO**	**RS_1km**	**RI**	**MO**	**RS_1km**	**RI**	**MO**	**RS_1km**
6.4	4.1	6.7	0.97	0.97	0.92	0.58	0.76	0.55

4.2. Application to ET Estimation at 90-m and 1-km Resolutions

One key advantage of model-derived Tends lies within the possibility to "de-contextualize" Tends submodels, meaning to extend the applicability of the so-called contextual ET models to remote sensing data at multiple spatial resolutions and to regions less heterogeneous than irrigated areas. In this study, the performance of the Tends algorithms is assessed by applying SEB-1S to both 90-m resolution and

1-km resolution data, using the model-derived or image-based Tends as input. The issue of the scale mismatch between remotely-sensed and ground-based ET estimates [12] is addressed by the following stepwise approach: 90-m resolution SEB-1S ET is evaluated with eddy covariance measurements, and the 1-km resolution SEB-1S ET is compared to the 90-m SEB-1S ET aggregated at 1-km resolution.

Figure 11 plots the ET simulated by SEB-1S at 90-m resolution *versus* the ground-based observations of the six stations of the Yaqui experiment. SEB-1S ET estimates are obtained using the available energy ($Rn - G$) observed by the flux stations and either the model-derived (RI or MO r_{ah} formulation) or the image-based Tends as input. The differences between the three simulation results (see the distribution of points in Figure 11) indicate that ET is greatly influenced by EF, which strongly depends on Tends. Moreover, the r_{ah} modeling approach for computing Tends has a significant impact on ET estimates. The RMSD between modeled and ground-based ET is 65, 82 and 120 W·m^{-2} for Tends_RS_90m, Tends_EBsoil with the MO r_{ah} formulation and Tends_EBsoil with the RI r_{ah} formulation, respectively. Tends_RS_90m-based ET estimates are in good agreement with *in situ* measurements, especially given the wide observation range of ET (0–700 W·m^{-2}) [14]. The comparison between SEB-1S and *in situ* ET is also undertaken using the available energy ($Rn - G$) modeled by SEB-1S. In this case, the RMSD between modeled and ground-based ET is 85, 106 and 130 W·m^{-2} for Tends_RS_90m, Tends_EBsoil with the MO r_{ah} formulation and Tends_EBsoil with the RI r_{ah} formulation, respectively. The RMSD (65–85 W·m^{-2}) of ET obtained when using image-based Tends as input is still in the range reported by the published literature [58–61]. In the case of the Yaqui experiment, RMSD values tend to be high, because most components of the energy budget are often large in semi-arid lands at low latitudes. In addition, a fairly strong bias is present on net radiation, which tends to worsen ET estimations [30]. Note that the significant difference in RMSD when using image-based *versus* model-derived Tends as input to SEB-1S could be attributed to uncertainties in the vegetation Tends (simply derived from soil Tends and air temperature). Nonetheless, the MO r_{ah} still provides acceptable ET estimates at 90-m resolution.

Figure 12 plots the ET simulated by SEB-1S at 1-km resolution (from aggregated ASTER data) *versus* the 1-km resolution ET chosen as a reference over the Yaqui area, on all seven ASTER overpass days. Given the good results obtained with Tends_RS_90m in terms of 90-m resolution ET estimates, the reference dataset for evaluating the 1-km resolution ET models is obtained by aggregating the 90-m resolution Tends_RS_90m-forced SEB-1S ET at 1-km resolution. The Tends used as input to SEB-1S are either modeled using Tends_EBsoil with the RI or MO r_{ah} formulation, or retrieved from (1-km resolution) remote sensing data using Tends_RS_1km. Statistical results, including RMSD, bias, slope of the linear regression and correlation coefficient, between simulated, and reference ET is reported in Table 4, for each Tends algorithm and for each observation day separately. The MO formulation yet again gives the best results, with a mean RMSD of 56 W·m^{-2} compared to 77 W·m^{-2} for the RI formulation. The mean RMSD obtained when using Tends_RS_1km is close to that obtained when using the RI-derived Tends, which is 78 W·m^{-2}. However, the lowest mean bias (15 W·m^{-2}) in ET is obtained for Tends_RS_1km compared to 23 W·m^{-2} and 27 W·m^{-2} for Tends_EBsoil with the MO and RI formulations, respectively. The mean correlation coefficient is increased from 0.95 for Tends_RS_1km up to 0.98 for both Tends_EBsoil algorithms.

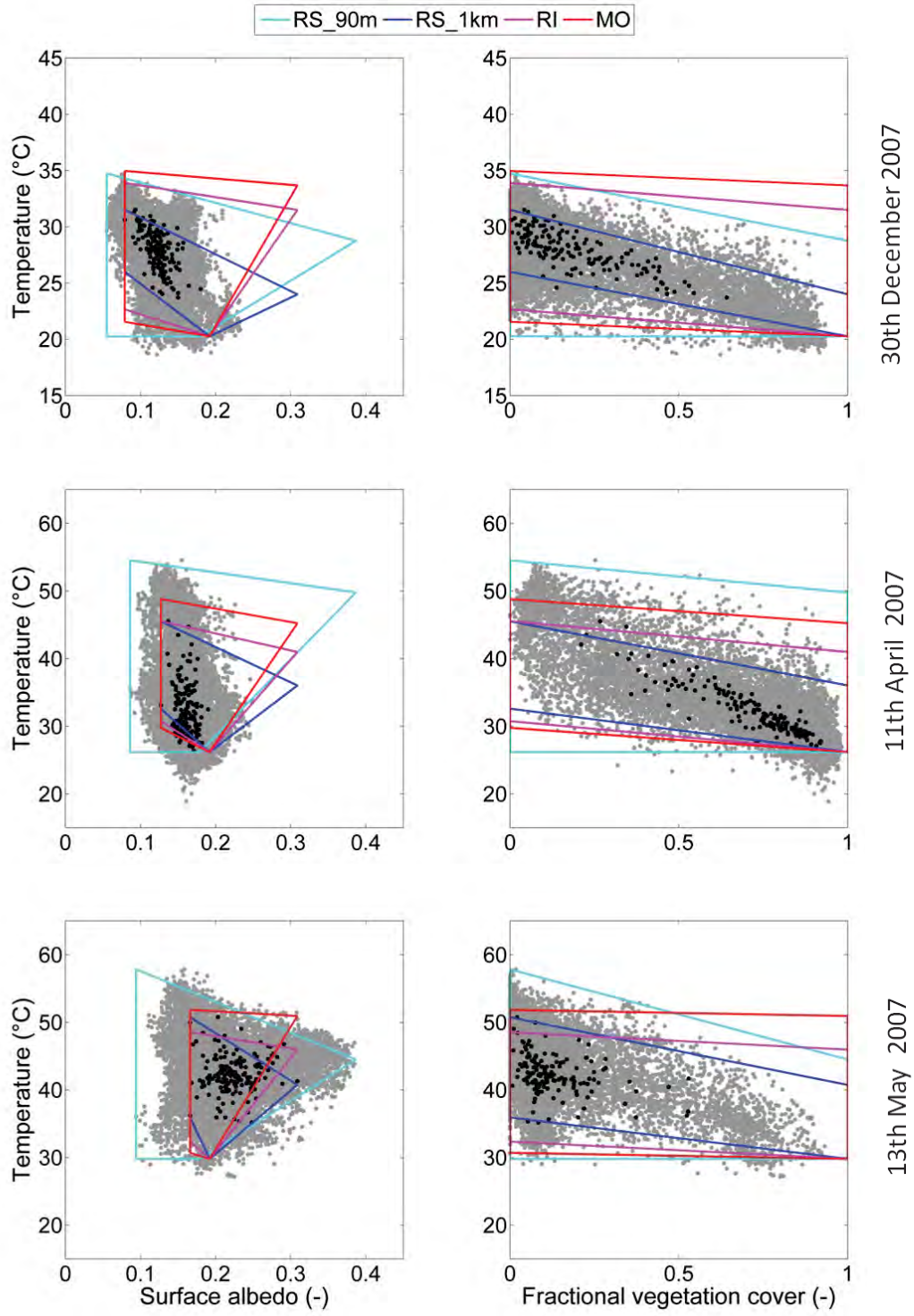


Figure 10. The $LST - \alpha$ and $LST - f_{vg}$ spaces corresponding to 1-km resolution data (black dots) and 90-m resolution data (grey dots) are overlaid with the polygons built using Tends_RS_1km (blue), Tends_RS_90m (light blue) and the Tends_EBsoil for the RI (magenta) and MO (red) aerodynamic resistance. Results are shown over the Yaqui area (30 December, 11 April and 13 May, respectively).

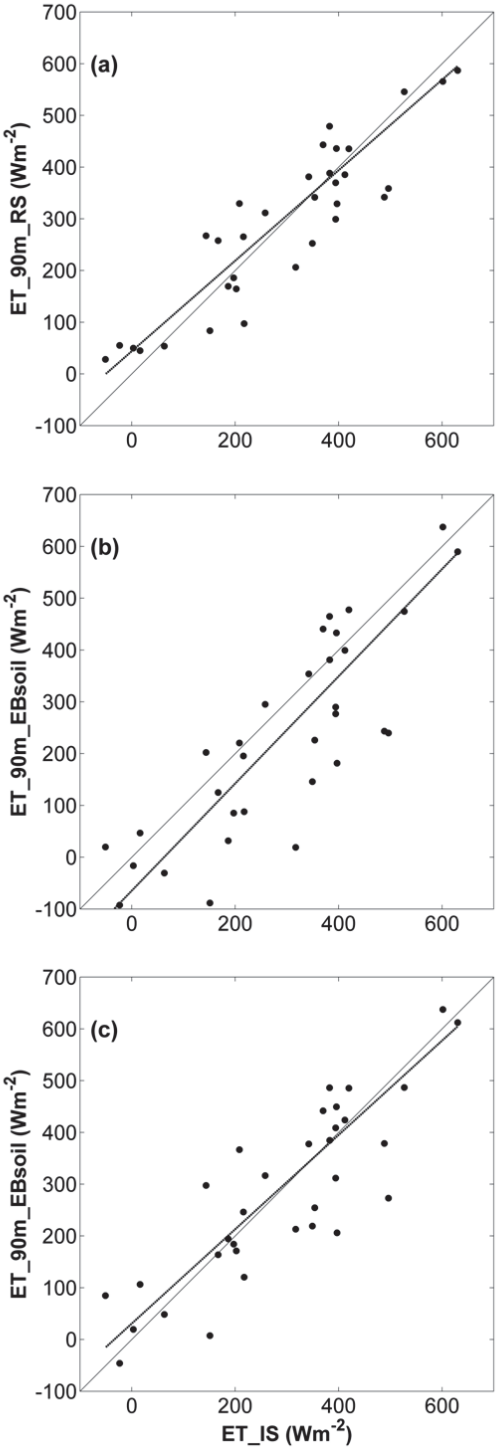


Figure 11. The ET simulated at 90-m resolution is plotted against station measurements for data over the Yaqui region. The Tends used as input to the EF_SEB-1S model consist of (**a**) the Tends retrieved from 90-m resolution data, (**b**) the Tends simulated by EBsoil with the RI and (**c**) MO aerodynamic resistance.

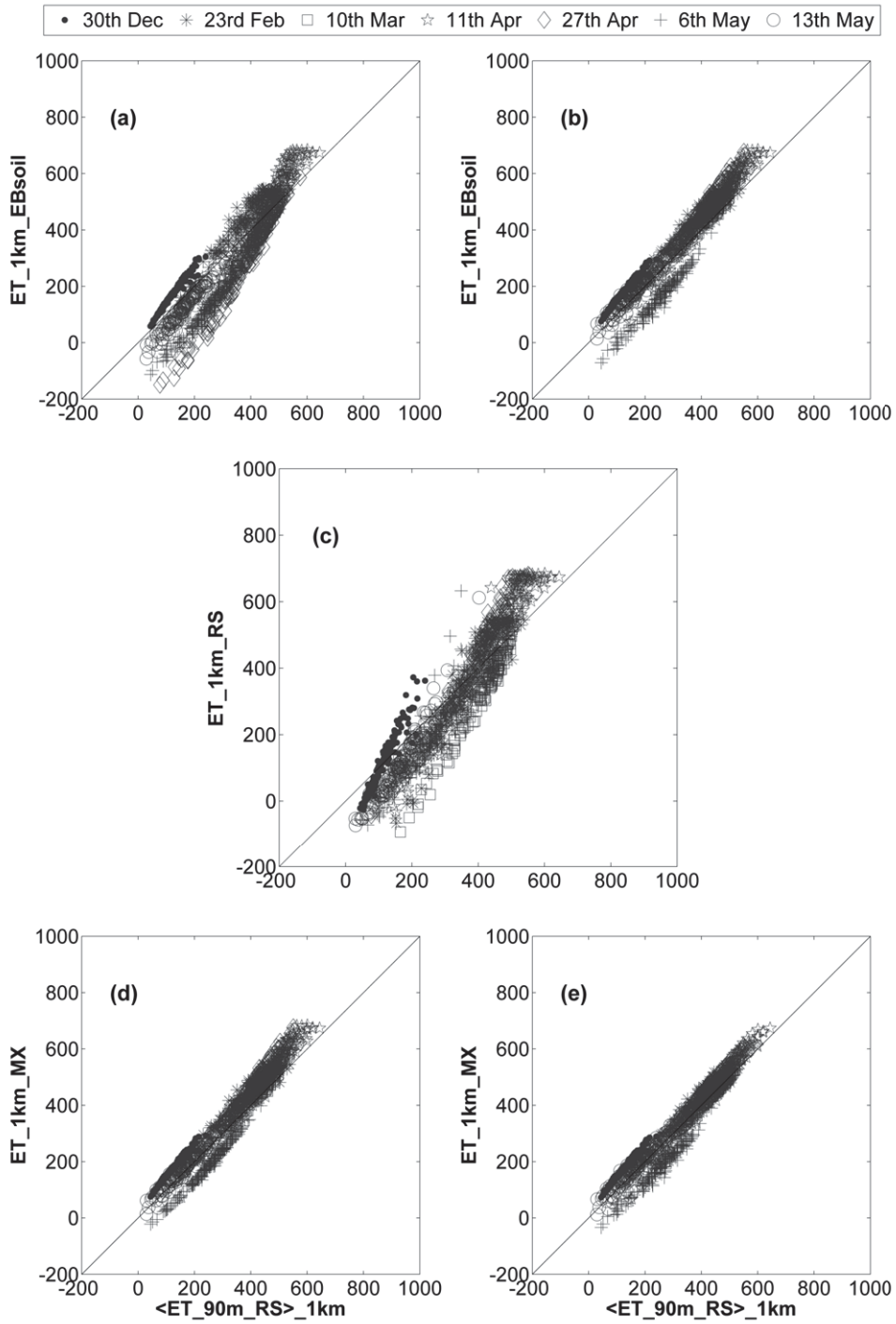


Figure 12. The ET simulated at 1-km resolution is plotted against the 90-m resolution simulated ET, aggregated at 1-km resolution, for data over the Yaqui region, on the seven ASTER overpass dates separately. The Tends used as input to the EF_SEB-1S model consist of (**a**) the Tends simulated by EBsoil with the RI and (**b**) MO aerodynamic resistance, (**c**) the Tends retrieved from 1-km resolution data, (**d**) the Tends simulated with the mixed modeling approach using αends_1km and (**e**) the Tends simulated with the mixed modeling approach using αends_90m.

Table 4. Root mean square difference (RMSD), bias, correlation coefficient (R) and slope of the linear regression between the ET simulated at 1-km resolution and the 90-m resolution simulated ET aggregated at 1-km resolution, for data over the Yaqui region, on the seven ASTER overpass dates separately. The Tends used as input to EF_SEB-1S consist of Tends simulated by EBsoil with the RI and MO aerodynamic resistance and the Tends retrieved from 1-km resolution data (RS).

Date	RMSD ($W \cdot m^{-2}$)			Bias ($W \cdot m^{-2}$)			R (-)			Slope (-)		
	RI	MO	RS	RI	MO	RS	RI	MO	RS	RI	MO	RS
30th Dec	48	55	47	46	54	5.3	0.98	0.98	0.96	1.3	1.2	2.0
23 February	64	58	78	57	55	2.0	0.94	0.96	0.94	1.1	0.98	1.7
10 March	61	36	99	−46	31	−73	0.99	0.97	0.97	1.5	1.1	1.8
11 April	71	57	91	24	37	41	0.97	0.98	0.96	1.6	1.4	1.7
27 April	122	63	79	−102	57	24	0.99	0.99	0.98	1.6	1.2	1.7
6 May	127	95	82	−126	−93	−52	0.98	0.98	0.93	1.2	1.1	1.6
13 May	42	28	67	−38	23	−53	0.98	0.97	0.95	1.2	1.0	1.4
All (mean)	77	56	78	−27	23	−15	0.98	0.98	0.95	1.4	1.1	1.7
All (σ)	34	21	17	72	53	44	0.01	0.01	0.01	0.22	0.12	0.19

At 1-km resolution, the Tends_EBsoil algorithm with MO formulation improves ET predictions. However, the presence of a significant bias can be observed on the ET estimates computed using the MO-derived Tends, on 6 May. In fact, Tends_EBsoil are computed independently of the LST dataset and, thus, can lead to the polygon not being centered on the cluster points in the $LST - \alpha$ and $LST - f_{vg}$ spaces. This can introduce a bias in model-derived Tends and, as a result, in EF/ET. In contrast, the polygon obtained from Tends_RS_1km is centered on the cluster points, because the computation of Tends is made using the observed data points. Overall, the ET estimates derived from Tends_RS_1km have a low bias. However, the mean slope of the linear regression between modeled and reference ET is estimated as 1.7. This is due to the fact that Tends_RS_1km overestimates the wet edge and underestimates the dry edge, which translates into the polygon boundaries being too close to the point cluster. The data points are thus closer to the boundaries at 1-km resolution than at 90-m resolution, influencing the range of variability and leading to slopes significantly larger than one.

SEB-1S, as a contextual model, performs well, provided that the extreme conditions associated with fully dry and wet soil/vegetation components are encountered in the same image. Many studies have documented the sensitivity of contextual models to the wet and dry edges. For instance, Long *et al*. [19] indicated that SEBAL shows great sensitivity to extreme soil Tends, which implies that the extreme conditions play a significant role in the estimation of ET. The dry and wet edges impact the ET estimates in a similar manner: increasing Tends for the wet/dry extremes tend to increase ET based on energy conservation. Yet the distance between the observed dry and wet edges depends the range of EF, which is linked to the growth stage of vegetation and to the spatial representativeness of the study area [8]. It also depends on the domain size. The impact of the domain size on ET estimates can reach

50–80 W·m^{-2} [19]. When applied to moderate or low spatial resolution images, contextual models could fail to discriminate extreme soil wetness conditions and, hence, to detect EF, especially for relatively small study sites [20]. Conversely, deriving the physical boundaries independently of remote sensing data could help improve ET estimates, especially by reducing the underestimation and overestimation of the dry and wet edge, respectively. The theoretical approach developed herein, based on a simple soil energy balance model and the MO r_{rah} formulation, is well adapted for large-scale applications using moderate to low resolution data.

4.3. A Mixed Modeling Remote Sensing Approach for Improved Tends and ET estimates

As the Tends_EBsoil with the MO formulation and the Tends_RS_1km provide complementary results in terms of bias and slope of the linear regression between simulated and reference ET, a mixed modeling approach is proposed for $T_{s,dry}$. The new formulation considers $T_{s,dry}$ as the maximum between the two estimates:

$$T_{s,dry,MX} = max(Tends_EBsoil, Tends_RS) \quad (10)$$

Furthermore, a new constraint concerning the albedo endmembers is applied. It consists of using either the 1-km resolution α endmembers (αends_1km) or the 90-m resolution α endmembers (αends_90m) in the estimation of Tends_RS_1km. A comparison between the 1-km resolution mixed-modeled and reference ET is also presented in Figure 12, for the Yaqui area, on all seven ASTER overpass days, and using αends_1km (e) and αends_90m (f). Results in terms of RMSD, bias, correlation coefficient and slope of the linear regression between simulated and reference ET are reported in Table 5, for each of the two mixed Tends formulations (MX_αends_1km and MX_αends_90m, respectively) and for the two previous configurations (MO-derived Tends_EBsoil and Tends_RS_1km, respectively). The statistical results are presented for all seven ASTER overpass dates combined. Both mixed modeling formulations significantly improve the results, with RMSD values of 43 W·m^{-2} and 52 W·m^{-2} (for MX_αends_90m and MX_αends_1km, respectively), as compared to 59 W·m^{-2} and 79 W·m^{-2} (for the MO_1km and RS_1km formulation, respectively).

Table 5. Root mean square difference (RMSD), bias, correlation coefficient (R) and slope of the linear regression between the ET simulated at 1-km resolution and the 90-m resolution simulated ET aggregated at 1-km resolution, for data over the Yaqui region, on all of the seven ASTER overpass dates combined. The Tends used as input to EF_SEB-1S consist of Tends simulated by EBsoil with the MO aerodynamic resistance, retrieved from 1-km resolution data (RS), simulated with the mixed modeling approach using αends_1km (MX_αends_1km) and with αends_90m (MX_αends_90m).

Configuration	RMSD (W·m^{-2})	Bias (W·m^{-2})	R (-)	Slope (-)
MO	59	23	0.95	1.1
RS	79	−15	0.93	1.3
MX_αends_1km	52	25	0.96	1.1
MX_αends_90m	43	9.3	0.96	1.0

As expected, when using the αends_90m as input to the Tends_RS_1km, the computed boundaries are much closer to the boundaries estimated at 90-m resolution. This results in the lowest bias obtained between the mixed-modeled ET at 1-km resolution (using MX_αends_90m derived Tends) and the 1-km resolution reference ET. Nevertheless, the improvement provided by the αends_90m configuration is relatively small, so that the MX_αends_1km strategy seems to be a good compromise in terms of ET accuracy and data availability. Especially the MX_αends_1km does not require high (90 m) resolution remote sensing data for calibrating αends.

5. Conclusions

This paper develops a simple theoretical approach to estimate Tends independently from LST images, for use as input to contextual ET models. The soil Tends simulated by a soil energy balance model using the RI and MO r_{ah} formulations are evaluated with *in situ* measurements made over an irrigated area in Morocco. Results indicate that the MO formulation provides the best soil Tends estimates, with a mean RMSE of 2.4 °C, compared to 3.6 °C for the RI formulation. Model-derived soil Tends are also compared to the image-based soil Tends derived from the SEB-1S Tends submodel, which combines both $LST - \alpha$ and $LST - f_{vg}$ spaces as additional constraints on wet and dry edges. The approach is tested using 90-m resolution ASTER data over two semi-arid irrigated areas in Mexico and Spain. The MO formulation still provides the best results, with an RMSD between simulated and image-based soil Tends estimated as 4.1 °C and 6.4 °C for the MO and RI formulations, respectively.

The performance of Tends algorithms in terms of ET estimation is assessed by applying the SEB-1S EF submodel to both 90-m resolution and 1-km resolution (aggregated) ASTER data using the model-derived or image-based Tends as input. In the case of the theoretical Tends algorithm, vegetation Tends are estimated from model-derived soil Tends and air temperature. The RMSD between 90-m resolution SEB-1S and *in situ* ET is estimated as 65 and 82 W·m^{-2}, and the RMSD between 1-km resolution SEB-1S and aggregated 90-m resolution SEB-1S ET is estimated as 78 and 56 W·m^{-2} for the image-based and model-derived Tends (MO formulation), respectively. In light of the above results, one concludes that Tends should be estimated *a priori* when contextual models are applied to low resolution images. Nonetheless, image-based Tends provide accurate ET estimates when high-resolution images are used in highly heterogeneous (irrigated) landscapes. In these conditions, the consistency between model-derived and high-resolution image-based Tends provides a meaningful basis for developing mixed modeling observational approaches.

Regarding the estimation of ET using contextual models with image- and physics-based edge determination, this study differs from previous studies in four main aspects: (1) Tends are derived from a r_{ah} formulation for bare soil only, although most theoretical Tends algorithms are based on r_{ah} formulations for both soil and vegetation [10,19,43]; (2) to the authors' knowledge, this is the first time that soil Tends have been evaluated using *in situ* measurements; (3) the image-based edge determination relies on both the $LST - \alpha$ and $LST - f_{vg}$ spaces, while previous studies have been based on one space only [8,15,19,20,43]; and (4) the sensitivity of ET simulations to observation resolution is assessed in terms of image-based and physics-based edge determination.

The successful implementation of a soil energy balance model in contextual ET models increases the accuracy in ET estimates at multiple resolutions and has many other potential applications. First, the consistency between model-derived and image-based Tends could lead to improving the parameterization of aerodynamic resistance by remote sensing means. Secondly, the minimum and maximum soil temperatures are the boundary conditions for LST-based soil evaporative efficiency (SEE, being defined as the ratio of actual to potential evaporation) models [62,63]. Improving the accuracy in soil Tends opens the path for monitoring SEE from optical/thermal remote sensing. Thirdly, LST-derived SEE is tightly coupled to the near-surface soil moisture retrieved from microwave radiometers, such as SMOS (Soil Moisture and Ocean Salinity), especially under semi-arid non-energy-limited conditions [64]. The DISPATCH (DISaggregation based on a Physical And Theoretical scale CHange; [29]) algorithm improves the spatial resolution of SMOS-like data by converting LST-derived SEE into high resolution soil moisture fields. Any improvement in LST-based SEE models would contribute to further developments of DISPATCH and to associated refinements of soil moisture products. Last, but not least, remotely-sensed estimates of LST-derived SEE and microwave-derived soil moisture data would be key for better constraining the partitioning of LST-derived ET into soil evaporation and plant transpiration [65]. Such information is crucial for better quantifying crop water needs and managing water resources over semi-arid areas.

Acknowledgments

This study was supported by TREMA (Télédétection et Ressources en Eau en Méditerranée semi-Aride) international laboratory, IRD (Institut de Recherche pour le Développement), CNRS (Centre National de la Recherche Scientifique) and ANR (Agence Nationale de la Recherche) MIXMOD-E ANR-13-JS06-0003-01 funding.

Author Contributions

Vivien Stefan collected the *in situ* measurements of extreme soil temperatures at the Haouz site, developed the EBsoil algorithm, integrated it within SEB-1S and performed all of the simulations. Olivier Merlin developed the former version of the SEB-1S model and supervised the present research. Salah Er-Raki and Said Khabba assisted with the implementation of the field campaign at the Haouz site and helped interpret the results. Maria-José Escorihuela provided the meteorological data used for the Urgell site. All co-authors have given insightful information concerning the final version of this paper.

Conflicts of Interest

The authors declare no conflict of interest.

Nomenclature

α	Surface albedo
α_s	Bare soil albedo
α_{vg}	Green vegetation albedo

α_{vs} Senescent vegetation albedo
$\alpha_{th,1}$ Threshold albedo, computed as the average between α_{vg} and α_s
α_{avg} Average of all albedo values
αends_90m Albedo endmembers at 90-m resolution
αends_1km Albedo endmembers at 1-km resolution
ASTER Advanced Spaceborne Thermal Emission and Reflection Radiometer
C_p Specific heat of air at constant pressure ($J \cdot kg^{-1} \cdot K^{-1}$)
$e_{sat}(T)$ Saturated vapor pressure at temperature T (Pa)
ϵ_s Soil emissivity
η Stability coefficient
EF Evaporative fraction
ET Evapotranspiration ($W \cdot m^{-2}$)
ET_IS Evapotranspiration given by *in situ* measurements ($W \cdot m^{-2}$)
ET_90m_RS Evapotranspiration estimated at 90-m resolution, by using image-based temperature endmembers as input ($W \cdot m^{-2}$)
ET_90m_EBsoil Evapotranspiration estimated at 90-m resolution, by using model-derived temperature endmembers as input ($W \cdot m^{-2}$)
ET_1km_RS Evapotranspiration estimated at 1-km resolution, by using image-based temperature endmembers as input ($W \cdot m^{-2}$)
ET_1km_EBsoil Evapotranspiration estimated at 1-km resolution, by using model-derived temperature endmembers as input ($W \cdot m^{-2}$)
ET_1km_MX Evapotranspiration estimated at 1-km resolution, by using as input the temperature endmembers derived from mixed-modeling ($W \cdot m^{-2}$)
f_{vg} Fractional green vegetation cover
$f_{vg,avg}$ Average of all fractional green vegetation cover values
g Gravitational constant ($m \cdot s^{-2}$)
G Ground heat flux ($W \cdot m^{-2}$)
γ Psychrometric constant ($Pa \cdot K^{-1}$)
H_s Soil sensible heat flux ($W \cdot m^{-2}$)
k Von Kármán constant
LST Land surface temperature (°C)
LE_s Soil latent heat flux ($W \cdot m^{-2}$)
L_{mo} Monin–Obukhov length (m)
MO Monin–Obukhov
NDVI Normalized Difference Vegetation Index
ψ_h Stability correction factor for heat transport
ψ_m Stability correction factor for momentum transport
r_{ah} Aerodynamic resistance to heat transfer ($s \cdot m^{-1}$)
$r_{ah,RI}$ Aerodynamic resistance to heat transfer based on the Richardson number ($s \cdot m^{-1}$)
$r_{ah,MO}$ Aerodynamic resistance to heat transfer based on the Monin–Obukhov length ($s \cdot m^{-1}$)

r_{ss}	Soil evaporation resistance	($s \cdot m^{-1}$)
ρ	Air density	($kg \cdot m^{-3}$)
RI	Richardson	
R_i	Richardson number	
RMSD	Root mean square difference	
RMSE	Root mean square error	
Ra	Incident atmospheric radiation at large wavelengths	($W \cdot m^{-2}$)
Rg	Incident solar radiation at short wavelengths	($W \cdot m^{-2}$)
Rn	Surface net radiation	($W \cdot m^{-2}$)
SEB-1S	Surface energy balance-mono-source	
σ	Stefan–Boltzmann constant	($W \cdot m^{-2} \cdot K^{-4}$)
SM	Surface (0–5 cm) soil moisture	($m \cdot^3 m^{-3}$)
SM_{fc}	Soil moisture at field capacity	($m \cdot^3 m^{-3}$)
SMC	Meteorological Service of Catalonia	
T_a	Air temperature	(°C)
T_{max}	Maximum surface temperature	(°C)
Tends	Temperature endmembers	(°C)
Tends_IS	Soil temperature endmembers measured *in situ*	(°C)
Tends_RS_90m	Soil temperature endmembers derived from 90-m resolution images	(°C)
Tends_RS_1Km	Soil temperature endmembers derived from 1-km resolution images	(°C)
Tends_EBsoil	Model-derived soil temperature endmembers	(°C)
$T_{s,dry}$	Temperature of a fully dry bare soil	(°C)
$T_{s,dry,MX}$	Maximum soil temperature, derived from the mixed modeling approach	(°C)
$T_{s,wet}$	Temperature of a fully wet bare soil	(°C)
$T_{v,dry}$	Temperature of fully water-stressed vegetation	(°C)
$T_{v,wet}$	Temperature of well-watered vegetation	(°C)
u_a	Wind speed	(ms^{-2})
u_*	Friction velocity	(ms^{-2})
UTC	Coordinated Universal Time	
WC	Automatic weather station of Golmes	
XEMA	Automatic weather station network	
Z_{0m}	Roughness length for momentum transfer over bare soil	(m)
Z_r	Reference height at which the wind speed is measured	(m)

Appendix

A. Image-Based Temperature Endmembers Algorithm

Adjustments to the estimation of image-based Tends, as implemented in the original version of SEB-1S, are made in this study. Especially, an effort is made to improve the robustness of the SEB-1S Tends submodule when applied to remote sensing data collected at a spatial resolution significantly

lower than 90 m. Modifications mainly consist of defining new image-based thresholds to estimate the slopes of the wet and dry edges more accurately using data that are not necessarily representative of extreme (wet/dry) soil and vegetation conditions. Below, we provide a summary of the methodology for the estimation of minimum vegetation temperatures ($T_{v,wet,1}$ and $T_{v,wet,2}$), minimum soil temperatures ($T_{s,wet,1}$ and $T_{s,wet,2}$), maximum soil temperatures ($T_{s,dry,1}$ and $T_{s,dry,2}$) and maximum vegetation temperatures ($T_{v,dry,1}$ and $T_{v,dry,2}$) for the $LST - \alpha$ and $LST - f_{vg}$ spaces, respectively.

With respect to the $LST - \alpha$ space, no modification is made to the computation of the minimum vegetation temperature $T_{v,wet,1}$ and the maximum soil temperature $T_{s,dry,1}$, respectively:

- $T_{v,wet,1}$ (if $\alpha = \alpha_{vg}$) is set to the air temperature T_a;
- $T_{s,dry,1}$ (if $\alpha = \alpha_s$) is set to the maximum temperature (T_{max}) observed within the study area.

Several modifications are made in the estimation of $T_{s,wet,1}$ and $T_{v,dry,1}$:

- The minimum soil temperature $T_{s,wet,1}$ is defined as the intercept at $\alpha = \alpha_s$ of the line passing through the point (α_{vg}, T_a) and the point with $\alpha < \alpha_{th,1}$, such that the slope of the line is maximum (meaning that all of the other data points with $\alpha < \alpha_{th,1}$ are located above the wet surface edge), with $\alpha_{th,1}$ being the average between α_{vg} and α_s. In the original version of SEB-1S, this threshold was set to α_{vg}.
- The maximum vegetation temperature $T_{v,dry,1}$ is defined as the intercept at $\alpha = \alpha_{vs}$ of the line passing through (α_s, $T_{s,dry,1}$) and the point with $\alpha > \alpha_{avg}$, such that the slope of the line is maximum (meaning that all of the other data points with $\alpha > \alpha_{avg}$ are located below the dry surface edge) with α_{avg} being the average of all α values within the study area.

In a similar manner, regarding the $LST - f_{vg}$ space, $T_{v,wet,2}$ and $T_{v,dry,2}$ remain unchanged:

- $T_{v,wet,2}$ (if $f_{vg} = 1$) is set to the air temperature T_a;
- $T_{s,dry,2}$ (if $f_{vg} = 0$) is set to T_{max}.

Several modifications are made in the estimation of $T_{s,wet,2}$ and $T_{v,dry,2}$:

- The minimum soil temperature $T_{s,wet,2}$ is computed as the intercept (at $f_{vg} = 0$) of the line passing through the point (1, T_a) and the point with $f_{vg} < f_{vg,avg}$, such that the slope of the line is maximum (meaning that all of the other data points with $f_{vg} < f_{vg,avg}$ are located above the wet surface edge) with $f_{vg,avg}$ being the average of all f_{vg} values within the study area. In the original version of SEB-1S, the threshold value ($f_{vg,avg}$) was set to 0.5. It is now computed for each day separately.
- The maximum vegetation temperature $T_{v,dry,2}$ is defined as the intercept (at $f_{vg} = 1$) of the line passing through the point (0, $T_{s,dry,2}$) and the point with $f_{vg} > f_{vg,avg}$, such that the slope of the line is maximum (meaning that all of the other data points with $f_{vg} > f_{vg,avg}$ are located below the dry surface edge).

Finally, as in the original version of SEB-1S, an estimation of the four Tends is given by averaging the two Tends sets obtained for each space separately:

$$T_{s,dry} = T_{s,dry,1} = T_{s,dry,2} = T_{max} \tag{11}$$

$$T_{s,wet} = (T_{s,wet,1} + T_{s,wet,2})/2 \tag{12}$$

$$T_{v,wet} = T_{v,wet,1} = T_{v,wet,2} = T_a \tag{13}$$

$$T_{v,dry} = (T_{v,dry,1} + T_{v,dry,2})/2 \tag{14}$$

B. Energy Balance Model for Bare Soil

The net radiation at the soil surface is given by:

$$Rn_s = (1 - \alpha_s)Rg + \epsilon_s(Ra - \sigma T_s^4) \tag{15}$$

Where α_s (unitless) represents the soil albedo, Rg (W·m^{-2}) the incident solar radiation at short wavelengths, ϵ_s (unitless) the soil emissivity, set to 0.96, Ra (W·m^{-2}) the incident thermal radiation at large wavelengths, σ (W·m^{-2}·K^{-4}) the Stefan–Boltzmann constant and T_s ($^\circ C$) the soil temperature.

The soil heat conduction flux can be approximated as [10,66]:

$$G = 0.32Rn_s \tag{16}$$

The soil sensible heat flux is computed as:

$$H_s = \frac{\rho C_p(T_s - T_a)}{r_{ah}} \tag{17}$$

with ρ (k·g·m^{-3}) being the air density and C_p (J·kg^{-1}·K^{-1}) the specific heat of air at constant pressure.

C. Aerodynamic Resistance Modeling (MO Formulation)

The MO length L_{mo} is computed as follows:

$$L_{mo} = -\frac{\rho C_p T_a u_*^3}{Kg(H + 0.61 C_p T_a L E_s)} \tag{18}$$

u_* (ms^{-1}) represents the friction velocity and is expressed as:

$$u_* = \frac{u_a k}{\log \frac{Z_m}{Z_{0m}} - \psi_m} \tag{19}$$

ψ_m represents the stability correction factor for momentum transport and is given by:

$$\psi_m = \frac{\psi_h}{2} + 2\log\frac{1+x}{2} - 2\arctan x + 0.5\pi \tag{20}$$

The stability correction factor for heat transport is expressed as:

$$\psi_h = 2\log\frac{1+x^2}{2} \tag{21}$$

where x is a function of the MO length L_{mo} and of the reference height for wind speed observations:

$$x = (1 - 16\frac{Z_r}{L_{mo}})^{0.25} \quad (22)$$

References

1. Garrido, A.; Llamas, M.; Varela-Ortega, C.; Novo, P.; Rodgriguez-Casado, R.; Aldaya, M. Water footprint and virtual water trade in Spain: Policy implications. *Nat. Resource Manag. Policy* **2010**, *35*, doi:10.1007/978-1-4419-5741-2.
2. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration—Guidelines for Computing Crop Water Requirements—FAO Irrigation and Drainage Paper 56*; FAO: Rome, Italy, 1998.
3. Allen, R.G. Using the FAO-56 dual crop coefficient method over an irrigated region as part of an evapotranspiration intercomparison study. *J. Hydrol.* **2000**, *229*, 27–41.
4. Allen, R.G.; Tasumi, M.; Morse, A.; R.Trezza. A Landsat-based energy balance and evapotranspiration model in Western US water rights regulation and planning. *Irrig. Drain. Syst.* **2005**, *19*, 251–268.
5. Bausch, W.C.; Neale, C. Spectral inputs improve corn crop coefficients and irrigation scheduling. *Trans. ASAE* **1989**, *32*, 1901–1908.
6. Long, D.; Singh, V.P. Integration of the GG model with SEBAL to produce time series of evapotranspiration of high spatial resolution at watershed scales. *J. Geophys. Res. Atmos.* **2010**, *115*, doi:10.1029/2010JD014092.
7. Stisen, S.; Sandholt, I.; Nørgaard, A.; Fensholt, R.; Jensen, K.H. Combining the triangle method with thermal inertia to estimate regional evapotranspiration—Applied to MSG-SEVERI data in the Senegal River basin. *Remote Sens. Environ.* **2008**, *112*, 1242–1255.
8. Tang, R.; Li, Z.L.; Tang, B. An application of the Ts-VI method with enhanced edges determination for evapotranspiration estimation from MODIS data in arid and semi-arid regions: Implementation and validation. *Remote Sens. Environ.* **2010**, *114*, 540–551.
9. Norman, J.M.; Kustas, W.P.; Humes, K.S. A two-source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. *Agric. For. Meteorol.* **1995**, *77*, 263–293.
10. Su, Z. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrol. Earth Syst. Sci.* **2002**, *6*, 85–99.
11. Kalma, J.D.; McVicar, T.R.; McCabe, M.F. Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data. *Surv. Geophys.* **2008**, *29*, 421–469.
12. Li, Z.L.; Tang, R.; Wan, Z.; Bi, Y.; Zhou, C.; Tang, B.; Yan, G.; Zhang, X. A review of current methodologies for regional evapotranspiration estimation from remotely sensed data. *Sensors* **2009**, *9*, 3801–3853.
13. Roerink, G.J.; Su, Z.; Menenti, M. S-SEBI: A simple remote sensing algorithm to estimate the surface energy balance. *Phys. Chem. Earth* **2000**, *25*, 147–157.
14. Merlin, O. An original interpretation of the surface temperature-albedo space to estimate crop evapotranspiration (SEB-1S). *Hydrol. Earth Syst. Sci.* **2013**, *17*, 3623–3637.

15. Long, D.; Singh, V.P. A two-source trapezoid model for evapotranspiration (TTME) from satellite imagery. *Remote Sens. Environ.* **2012**, *121*, 370–388.
16. Moran, M.S.; Clarke, T.R.; Inoue, Y.; Vidal, A. Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. *Remote Sens. Environ.* **1994**, *49*, 246–263.
17. Carlson, T. An overview of the "Triangle Method" for estimating surface evapotranspiration and soil moisture from satellite imagery. *Sensors* **2007**, *7*, 1612–1629.
18. Jiang, L.; Islam, S. A methodology for estimation of surface evapotranspiration over large areas using remote sensing observations. *Geophys. Res. Lett.* **1999**, *26*, 2773–2776.
19. Long, D.; Singh, V.P.; Li, Z.L. How sensitive is SEBAL to changes in input variables, domain sizes and satellite sensor? *J. Geophys. Res.* **2011**, *116*, doi:10.1029/2011JD016542.
20. Long, D.; Singh, V.P.; Scalon, B.R. Deriving theoretical boundaries to address scale dependencies of triangle models for evapotranspiration estimation. *J. Geophys. Res.* **2012**, *117*, doi:10.1029/2011JD017079.
21. Long, D.; Singh, V.P. Assessing the impact of end-member selection on the accuracy of satellite-based spatial variability models for actual evapotranspiration estimation. *Water Resour. Res.* **2013**, *49*, 2601–2618.
22. Timmermans, W.J.; Kustas, W.P.; Anderson, M.C.; French, A.N. An intercomparison of the Surface Energy Balance Algorithm for Land (SEBAL) and the Two-Source Energy Balance (TSEB) modeling schemes. *Remote Sens. Environ.* **2007**, *108*, 369–384.
23. Yang, Y.; Shang, S. A hybrid dual-source scheme and trapezoid framework-based evapotranspiration model (HTEM) using satellite images: algorithm and model test. *J. Geophys. Res. Atm.* **2013**, *118*, 2284–2300.
24. de Tomás, A.; Nieto, H.; Guzinski, R.; Salas, J.; Sandholt, I.; Berliner, P. Validation and scale dependencies of the triangle method for the evaporative fraction estimation over heterogeneous areas. *Remote Sens. Environ.* **2014**, *152*, 493–511.
25. Venturini, V.; Bisht, G.; Islam, S.; Jiang, L. Comparison of evaporative fractions estimated from AVHRR and MODIS sensors over South Florida. *Remote Sens. Environ.* **2004**, *93*, 77–86.
26. Duchemin, B.; Hadria, R.; Er-Raki, S.; Boulet, G.; Maisongrande, P.; Chehbouni, A.; Escadafal, R.; Hoedjes, J.; Karrou, H.; Mougenot, B.; *et al*. Monitoring wheat phenology and irrigation in Central Morocco: on the use of relationship between evapotranspiration, crops coefficients, leaf area index and remotely-sensed vegetation indices. *Agric. Water Manag.* **2006**, *79*, 1–27.
27. Er-Raki, S.; Chehbouni, A.; Guemouria, N.; Duchemin, B.; Ezzahar, J.; Hadria, R. Combining FAO-56 model and ground-based remote sensing to estimate water consumptions of wheat crops in a semi-arid region. *Agric. Water Manag.* **2007**, *87*, 41–54.
28. iButtonLink Technology. Available online: http://www.ibuttonlink.com/products/ds1921g (accessed on 3 February 2014).
29. Merlin, O.; Escorihuela, M.J.; Mayoral, M.A.; Hagolle, O.; Al Bitar, A.; Kerr, Y. Self-calibrated evaporation-based disaggregation of SMOS soil moisture: an evaluation study at 3 km and 100 m resolution in Catalunya, Spain. *Remote Sens. Environ.* **2013**, *130*, 25–38.

30. Chirouze, J.; Boulet, G.; Jarlan, L.; Fieuzal, R.; Rodriguez, J.C.; Ezzahar, J.; Er-raki, S.; Bigeard, G.; Merlin, O.; Garatuza-Payan, J.; *et al*. Inter-comparison of four remote sensing based surface energy balance methods to retrieve surface evapotranspiration and water stress of irrigated fields in semi-arid climate. *Hydrol. Earth Syst. Sci. Discuss.* **2013**, *10*, 895–963.
31. Merlin, O.; Duchemin, B.; Hagolle, O.; Jacob, F.; Coudert, B.; Chehbouni, G.; Dedieu, G.; Garatuza, J.; Kerr, Y. Disaggregation of MODIS surface temperature over an agricultural area using a time series of Formosat-2 images. *Remote Sens. Environ.* **2010**, *114*, 2500–2512.
32. Fieuzal, R.; Duchemin, B.; Jarlan, L.; Zribi, M.; Baup, F.; Merlin, O.; Hagolle, O.; Garatuza-Payan, J. Combined use of optical and radar satellite data for the monitoring of irrigation and soil moisture of wheat crops. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 1117–1129.
33. Gillespie, A.; Rokugawa, S.; Matsunaga, T.; Cothern, S.; Hook, S.; Khale, A. A temperature and emissivity separation algorithm for Advanced Spaceborne Thermal Emission and Reflection (ASTER) images. *IEEE Trans. Geosci. Remote Sens.* **1998**, *36*, 1113–1126.
34. Schmugge, T.J. Applications of passive microwave observations of surface soil moisture. *J. Hydrol.* **1998**, *212–213*, 188–197.
35. Weiss, M.; Baret, F.; Leroy, M.; Begué, A.; Hautecoeur, O.; Santer, R. Hemispherical reflectance and albedo estimates from the accumulation of across track sun synchroneous satellite data. *J. Geophys. Res.* **1999**, *104*, 221–232.
36. Bsaibes, A.; Courault, D.; Baret, F.; Weiss, M.; Olioso, A.; Jacob, F.; Hagolle, O.; Marloie, O.; Bertrand, N.; Desfond, V.; Kzemipour, F. Albedo and LAI estimates from FORMOSAT-2 data for crop monitoring. *Remote Sens. Environ.* **2009**, *113*, 716–729.
37. Gómez, M.; Olioso, A.; Sobrino, J.A.; Jacob, F. Retrieval of evapotranspiration over the Alpilles/ReSeDA experimental site using airborne POLDER sensor and a thermal camera. *Remote Sens. Environ.* **2005**, *96*, 399–408.
38. Sobrino, J.A.; Gómez, M.; Jiménez-Munoz, J.C.; Olioso, A.; Chehbouni, G. A simple algorithm to estimate evapotranspiration from DAIS data: Application to the DAISEX campaigns. *J. Hydrol.* **2005**, *315*, 117–125.
39. Sobrino, J.A.; Gómez, M.; Jiménez-Munoz, J.C.; Olioso, A. Application of a simple algorithm to estimate daily evapotranspiration from NOAA-AVHRR images for the Iberian Peninsula. *Remote Sens. Environ.* **2007**, *110*, 139–148.
40. Fan, L.; Liu, S.; Bernhofer, C.; Liu, H.; Berger, F.H. Regional land surface energy fluxes by satellite remote sensing in the Upper Xilin River watershed (Inner Mongolia, China). *Theor. Appl. Climatol.* **2007**, *88*, 231–245.
41. Galleguillos, M.; Jacob, F.; Prévot, L.; Liang, S. Mapping daily evapotranspiration over a Mediterranean vineyard watershed. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 168–172.
42. Galleguillos, M.; Jacob, F.; Prévot, L.; French, A.; Lagacherie, P. Comparison of two temperature differencing methods to estimate daily evapotranspiration over a Mediterranean vineyard watershed from ASTER data. *Remote Sens. Environ.* **2011**, *115*, 1326–1340.
43. Allen, R.G.; Tasumi, M.; Trezza, R. Satellite-based energy balance for mapping evapotranspiration with internalized calibration (METRIC) - Model. *J. Irrig. Drain. Eng. ASCE* **2007**, *133*, 380–394.

44. Sellers, P.J.; Heiser, M.D.; Hall, F.G. Relations between surface conductance and spectral vegetation indices at intermediate (100 m^2 to 15 km^2) length scales. *J. Geophys. Res.* **1992**, *97*, 19033–19059.
45. Kustas, W.P.; Schmugge, T.J.; Humes, K.S.; Jackson, T.J.; Parry, R.; Weltz, M.A.; Moran, M.S. Relationships between evaporative fraction and remotely sensed vegetation index and microwave brightness temperature for semiarid rangelands. *J. Appl. Meteor.* **1993**, *32*, 1781–1790.
46. Crow, W.T.; Kustas, W.P.; Prueger, J.H. Monitoring root-zone soil moisture through the assimilation of a thermal remote sensing-based soil moisture proxy into a water balance model. *Remote Sens. Environ.* **2008**, *112*, 1268–1281.
47. Monteith, J.L. *Principles of Environmental Physics*; Edward Arnold Press: London, UK, 1973; p. 241.
48. Choudhury, B.; Reginato, R.J.; Idso, S.B. An analysis of infrared temperature observations over wheat and calculation of latent heat flux. *Agric. For. Meteor.* **1986**, *37*, 75–88.
49. Lee, H. Notes and correspondence: improvement of surface flux calculation in the atmospheric surface layer. *J. Appl. Meteor.* **1997**, *36*, 1416–1423.
50. Yang, K.; Tamai, N.; Koike, T. Analytical solution of surface layer similarity equations. *J. Appl. Meteor.* **2001**, *40*, 1647–1653.
51. Liu, Y.; Hiyama, T.; Yamaguchi, Y. Scaling of land surface temperature using satellite data: A case examination on ASTER and MODIS products over a heterogeneous terrain area. *Remote Sens. Environ.* **2006**, *105*, 115–128.
52. Carlson, T.N.; Capehart, W.J.; Gillies, R.R. A new look at the simplified method for remote-sensing of daily evapotranspiration. *Remote Sens. Environ.* **1995**, *54*, 161–167.
53. Prihodko, L.; Goward, S.N. Estimation of air temperature from remotely sensed surface observation. *Remote Sens. Environ.* **1997**, *60*, 335–346.
54. Bastiaanssen, W.G.M.; Menenti, M.; Feddes, R.A.; Holtslag, A.A.M. A remote sensing surface energy balance algorithm for land (SEBAL) 1. Formulation. *J. Hydrol.* **1998**, *212–213*, 198–212.
55. Senay, G.; Bohms, S.; Singh, R.; Gowda, P.; Velpuri, N. Operational evapotranspiration mapping using remote sensing and weather datasets: A New parametrization for the SSEB approach. *J. Am. Water Assoc.* **2013**, *49*, 577–591.
56. Yang, K.; Koike, T.; Ishikawa, H. Turbulent flux transfer over bare soil surfaces: Characteristics and parameterization. *J. Appl. Meteor.* **2008**, *40*, 276–290.
57. Liu, S.; Mao, D.; Jia, L. Evaluating parameterizations of aerodynamic resistance to heat transfer using field measurements. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 769–783.
58. Gonzales-Dugo, M.P.; Neale, C.M.U.; Mateos, L.; Kustas, W.P.; Prueger, J.H.; Anderson, M.C.; Li, F. A comparison of operational remote sensing-based models for estimating crop evapotranspiration. *Agric. For. Meteor.* **2009**, *149*, 1843–1853.
59. McCabe, M.F.; Wood, E.F. Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors. *Remote Sens. Environ.* **2006**, *105*, 271–285.

60. Minacapilli, M.; Agnese, C.; Blanda, F.; Cammalleri, C.; Ciraolo, G.; D'Urso, G.; Iovino, M.; Pumo, D.; Provenzano, G.; Rallo, G. Estimation of actual evapotranspiration of Mediterranean perennial crops by means of remote-sensing based surface energy balance models. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 1061–1074.
61. Su, H.; Wood, E.; McCabe, M.; Su, Z. Evaluation of remotely sensed evapotranspiration over the CEOP EOP-1 reference sites. *J. Meteorol. Soc. Jpn* **2007**, *85A*, 439–459.
62. Nishida, K.; Nemani, R.R.; Glassy, J.M.; Running, S.W. Development of an evapotranspiration index from Aqua/MODIS for monitoring surface moisture status. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 493–501.
63. Merlin, O.; Walker, J.P.; Chehbouni, A.; Kerr, Y. Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency. *Remote Sens. Environ.* **2008**, *112*, 3935–3946.
64. Merlin, O.; Chehbouni, G.; Walker, J.P.; Panciera, R.; Kerr, Y. A simple method to disaggregate passive microwave based soil moisture. *IEEE Trans. Geosc. Remote Sens.* **2008**, *46*, 786–796.
65. Merlin, O.; Chirouze, J.; Olioso, A.; Jarlan, L.; Chehbouni, G.; Boulet, G. An image-based four-source surface energy balance model to estimate crop evapotranspiration from solar reflectance/thermal emission data (SEB-4S). *Agric. For. Meteor.* **2014**, *184*, 188–203.
66. Kustas, W.P.; Daughtry, C.S.T. Estimation of the soil heat flux/net radiation ratio from spectral data. *Agric. For. Meteor.* **1989**, *49*, 205–223.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

Chapter 4

Global modeling of the temporal behavior of soil evaporative efficiency

Contents

4.1 Introduction 96

4.2 Global Modeling 97

4.2.1 Concepts 97

4.2.2 Method formulation 98

4.2.3 Algorithm 99

4.3 Data description 100

4.3.1 Transfert Eau Chaleur (TEC) model 100

4.3.2 TEC Outputs 100

4.3.3 Data pre-processing 102

4.4 Application and Results 103

4.4.1 SEE Global Model 103

4.4.2 Model validation 109

4.5 Summary and Conclusions 110

4.1 Introduction

Soil evaporative efficiency (SEE), defined as the ratio of actual (LE_s) to potential evaporation (LE_p), is a key variable in retrieving soil evaporation (E) by remote sensing means. Since SEE is much more directly linked to remote sensing data than E is, various SEE models have been developed in conjunction with two observable variables - land surface temperature (LST) and surface soil moisture (SM). While the first part of this PhD focused on mapping evapotranspiration at mulitple resolutions from an instantaneous LST–based model, this second part is aimed at characterizing the temporal dynamics of SEE.

Characterizing SEE for various soil types is important, since soil texture plays an important role in the dynamics of the daily SEE cycle (*Komatsu*, 2003; *Chanzy and Bruckler*, 1993). However, SEE models have been rarely tested for different soil types. In mechanistic models, the link between SEE and soil texture is ensured through pedotransfer functions, which connect soil hydraulic properties with sand and clay percentages. *Merlin et al.* (2016) developed a meta–analysis approach to represent SEE over a large range of soil and atmospheric conditions, by using data collected at over 30 sites worldwide (article found in Annex). A significant relationship has been observed between a semi–empirical E model parameter and soil texture.

In this part, attempts have been made to characterize the daily cycle of SEE for various soil types, starting from synthetic SEE time series. One possibility of doing that is by using the global modeling technique, which relies on the theory of nonlinear dynamical systems. This technique is a fairly recent one, having been developed in the 1990s (*Gouesbet and Letellier*, 1994; *Letellier et al.*, 1995). Its application to the environmental dynamics is even more recent (*Maquet et al.*, 2007; *Mangiarotti et al.*, 2014). It uses the phase space in order to represent, analyze and understand the temporal evolution of the dynamical behavior. The representation in the phase space is useful because it offers a geometrical representation of the dynamical behavior. One important point of interest of the approach comes from the fact that, in principle, such a phase space can be reconstructed from observational time series (*Takens*, 1981). In its work (*Lorenz*, 1963), Edward Lorenz has shown that equations of a simplified atmosphere may lead to unpredictable behaviors in the long term. One reason to use the global modeling approach when studying SEE is that its dynamics is directly coupled to the atmosphere. Similar properties of low predictability for the behavior of SEE as the ones obtained by Lorenz may thus be expected from it. To study the temporal behavior of the SEE, it appears natural to use the theory of nonlinear dynamical systems which is well adapted to such type of dynamics. There is a novelty aspect related to this approach, since this technique has been used for a very small number of environmental contexts, with even fewer models obtained that are able to reproduce the original dynamics.

First of all, a description of the global modeling technique and an overview of the method used, along with its limitations, are presented in Section 4.2. A description of the synthetic SEE data and the model used to generate it is offered in Section 4.3. Results in terms of the global SEE model and its validation are proposed in Section 4.4, with a summary and general conclusions given in Section 4.5. The article based on this analysis can be found in Section 4.6, which is a summary of the results obtained in terms of the SEE global model and its predictability in time.

4.2 Global Modeling

The work of Henri Poincaré (1854 - 1912) is at the basis for the development of the theory of nonlinear dynamical systems, making way for the emergence of the chaos theory (*Letellier*, 2006). Poincaré used the phase space in order to develop new concepts such as recurrence maps (nowadays known as Poincaré map), limit cycles and fixed points. These concepts and the whole work of Poncaré have brought the fundamental concepts to understand, from a qualitative point of view, how in a deterministic context, small differences in the initial conditions can lead to the exponential divergence of the time evolution. With the development of digital computers in the 1960s, the first illustrations to Poincaré's theory have increased. Amongst all, the most notable one belongs to Edward Lorenz, who in 1963 discovered the first tridimensional autonomous system, which is both deterministic and sensitive to initial conditions, making it unpredictable in the long–term. As for more practical "applications", the most notable onse are the first global models obtained from experimental data (*Letellier et al.*, 1995, 2009). Below, a description of the concepts used in this approach, along with a presentation of the method and algorithm behind the global modeling technique.

4.2.1 Concepts

Some concepts are required when using the theory of nonlinear dynamical systems, which are presented below. A more in–depth description behind the notions and concepts can be found in various works such as *Abarbanel* (1996), *Bergé et al.* (1998) and *Letellier* (2006).

Determinism

In order to understand and foresee the observed dynamical behaviors of systems, it is necessary to identify the cause and effect links between observed phenomena. Determinism assumes both the causality condition (every state is causally linked to the prior state) and the necessary condition (the transition from one state to the next state is entirely determined and is given by a unique relationship). By knowing the current state and the evolution laws, one can predict all future states and also predict backward all past states.

Phase space

The phase space is a space that allows the representation of all possible states of a system. In the theory of dynamical systems, it provides a geometrical representation of a deterministic behavior. One important result of the theory of nonlinear dynamical system is that such a space can be reconstructed for one given system from one single observable. This is guaranteed by a theorem by F. Takens (*Takens*, 1981). In practice, the phase space can be reconstructed either using differential coordinates (one variable and its successive derivatives), or delayed coordinates (the same variable represented as a function of itself for various time delays).

Global modeling technique

The aim of most modeling techniques is to reproduce the observations. A valid model is obtained when it is able to simulate the same temporal evolution as the observed system. They are often referred to as "local" techniques because they are applied on a specific time window. Global approaches however focus on finding the underlying dynamics of the sytem rather than the temporal evolution itself (since observed temporal evolution is the result of specific initial conditions). In practice, global modeling consists in projecting the experimental data, embedded in a reconstructed space spanned by derivative or delayed coordinates, onto a basis of functions (*Gouesbet and Letellier*, 1994; *Aguirre and Billings*, 1995; *Mangiarotti et al.*, 2012a).

Phase portrait

Phase portraits are used in global modeling because they offer a geometrical representation of

the dynamical behavior. Basically, a phase space portrait is a representation of the trajectories of a dynamical system in the phase plane. There are two types of phase portraits: differential portrait (based on the successive derivatives of the original system) and delayed portrait (based on the delayed coordinates of the original system). These are, in principle, equivalent. In this study, the first type of phase portrait was used, as it does not require any choice of parameters.

Chaos

Two conditions are required for chaos: 1) determinism and 2) high sensitivity to initial conditions (*Kaplan and Glasss*, 1992). Most of the available methodologies are not able to verify the underlying determinism of a system, even though they are able to extract geometrical and dynamical nonlinear invariants. On the other hand, topological invariants are a more powerful way of detecting the type of chaos associated to the system. However, their practical derivation supposes being able to derive the deterministic component, in order to further apply the topological analysis. Global modeling is, in this context, a robust method of detecting the presence of a deterministic component from real world observations (*Freitas et al.*, 2009; *Mangiarotti et al.*, 2012b). Therefore, it can allow to get, all together in a single synthetic object, the necessary conditions for chaos together with all the other properties of chaos (*Mangiarotti et al.*, 2014).

(Chaotic/Strange) Attractor

When integrating the equations of a dynamical system, its trajectory (if not divergent) will, after a given time corresponding to a transitory, tend to evolve towards a set of numerical values, which is called an attractor. Attractors can be either fixed points, or limit cycles (of a given period), or they can have a more complex aperiodic structure (known as chaotic attractors). Chaotic attractors are represented by exponential diverging trajectories: two very close initial trajectories will diverge exponentially and will present a fractal structure. Strange attractors also exhibit a fractal structure but do not require the properties of a divergent trajectory.

Nonlinear invariants

A major difficulty in characterizing chaotic dynamics is the high sensitivity to initial conditions. One must be able to find some properties that are not sensitive to initial conditions, invariants that can be quantitatively estimated. Three types of nonlinear invariants can be distinguished:

- dynamical invariants, which describe the stability of the dynamics (*Chlouverakis and Sprott*, 2004; *F. Grond*, 2003, 2005; *Wolf et al.*, 1985)
- geometrical invariants, linked to the self–similarity of attractors when changing scale within the phase domain, (*Grassberger and Procaccia*, 1983; *Kaplan and Yorke*, 1979)
- topological invariants, which describe the structure of the attractor (*Gilmore*, 1998)

Embedding dimensions

The embedding dimension corresponds to the number of dimensions required to reconstruct the dynamical behavior. It corresponds to the number of variables of the original dynamics.

4.2.2 Method formulation

Instead of solving the system analytically, the global modeling technique uses the phase space to identify and study the associated dynamical behavior. In theory, it is possible to reconstruct the phase space of a given behavior starting from a unique time series (*Takens*, 1981). In its differential form, given a unique variable X_1, the approach aims to obtain models whose canonical form is:

$$\begin{cases} \dot{X}_i = X_{i+1}, i = 1, ..., n-1 \\ \dot{X}_n = F(X_1, X_2, ..., X_n) \end{cases} \tag{4.1}$$

where n is the model dimension and X_i the successive derivatives of the observed variable X_1. In practice, modeling the dynamical behavior comes back to obtaining an approximation of the function F that allows to reproduce the observed dynamics in a satisfying way. Identifying this function without any *a priori* knowledge is a difficult task. Therefore, strong tools are needed. The tool used in this study is GloMo (*Mangiarotti et al.*, 2012a, 2016). This tool has been successfully tested on theoretical and experimental systems (*Mangiarotti et al.*, 2012b). It was also used to model a number of different environmental variables, such as cereal crop cycles (*Mangiarotti et al.*, 2014), karstic spring (*Zhang et al.*, 2016) and plague epidemics (*Mangiarotti*, 2015; *Huc and Mangiarotti*, 2016) to name a few.

4.2.3 Algorithm

The algorithm used to apply the global modeling to the individual series of SEE is GloMo (*Mangiarotti et al.*, 2012a). Its aim is to find a parameterization of the function F in Equation 4.1, using a Gram–Schmidt procedure, as in (*Gouesbet and Letellier*, 1994). The algorithm allows any algorithm dimension with any polynomial degree.

GloMo provides an accurate estimate of the parameterization that will then be used to test the integrability of the model. An interactive interface allows the user to add or remove polynomial terms from the generic polynomial. After each change in terms, a new identification can be performed and the reliability of the model with respect to numerical integrability tested. Additional information about the suitability of the monomials is also provided after each run. This is attained by removing monomials one by one and then proceed to compare the fittings to the one obtained with all monomials. Such a suitability can reflect the contribution of a term only in reference to the corresponding structure. Another structure can lead to different results.

The main difficulty of global modeling is that no information regarding the model structure is available. In other words, what polynomial terms should be included in the model and which ones should be neglected. If the chosen structure is not well adapted (that is if the chosen structure does not correspond to the function F) then the model will diverge. Selecting a proper polynomial structure is in itself a difficult task because of all the possible combinations of terms. For example, let us consider a system of dimension 3 - that is consisting of three variables X_1, $X_2 = \dot{X}_1$ and $X_3 = \dot{X}_2$.

If one decides to limit the models only to quadratic polynomials (i.e. a polynomial degree equal to 2 at most), then there are ten possible terms for the global model: 1, X_1, X_2, X_3, X_1X_2, X_1X_3, X_2X_3, X_1^2, X_2^2 and X_3^2). That mounts to 2^{10} possible models among which only one may be the exact model (in some rare cases), and only a few ones may be a good approximation for it. In degree 5, this number of possible models will be 2^{56}.

Another problem is observability. Depending on the original dynamical system and on the chosen (or available) observed variable, one may provide a good observability of the dynamics of the system. As a consequence, the function F may be more or less difficult to retrieve or to approximate by a polynomial (*Letellier et al.*, 1995). In some cases, it may be impossible to obtain it. This point is of particular interest since the variable of interest, SEE, is defined in such a way that observability problems may arrise. In particular, the potential evaporation LE_p is ~ 0 during night–time, which might induce singularities in the model.

4.3 Data description

Synthetic SEE series are generated using the TEC (Transfert Eau Chaleur, *Chanzy and Bruckler* (1993)) model, for different soil textures. A brief description of the model and its outputs is offered below.

4.3.1 Transfert Eau Chaleur (TEC) model

The Transfert Eau Chaleur (TEC) model (*Chanzy and Bruckler*, 1993) has been used in order to derive our variable of interest, the soil evaporative efficiency. The TEC model is based on the theory of soil heat and water flows in partially saturated media. It considers the soil as a multilayer system and solves the coupled differential equations of heat and mass in order to derive surface fluxes such as evapotranspiration, latent heat, radiance, soil moisture profiles and soil temperature profiles. TEC is based on soil properties that are estimated using pedotransfer functions in order to characterize the soil hydraulic properties, and uses a climatic forcing. These pedotransfer functions are highly sensitive to initial soil water conditions and lower boundary conditions. The four pedotransfer functions used by TEC in order to estimate soil hydraulic properties are: *Wösten* (1997), *Cosby et al.* (1984), *Rawls and Brackensiek* (1985) and *Vereecken et al.* (1989) and they are herein referred to as "wos", "cos", "bra" and "ver", respectively.

4.3.2 TEC Outputs

The TEC outputs in terms of soil evaporation are different, depending on the formulation of the pedotransfer function used. Simulations are generated at an hourly time step, using a climatic forcing typical of a Mediterranean climate with heavy rain and long dryness periods (*Chanzy et al.*, 2008). Figure 4.1 is a visual representation of the soil evaporation (LE_s) and potential evaporation (LE_p), obtained using the four different pedotransfer functions.

When using the *Cosby et al.* (1984) and *Vereecken et al.* (1989) pedotransfer functions, the soil evaporation LE_s seems to be overestimated as compared to simulations given by the *Wösten* (1997) and *Rawls and Brackensiek* (1985) formulations. In addition, the formulation given by *Rawls and Brackensiek* (1985) offers negative values of LE_s. Therefore, simulations obtained using the *Wösten* (1997) pedotransfer function have been chosen for further analysis.

Both LE_s and LE_p follow a daily cycle, with LE_p reaching a peak around noon, with high incoming solar radiation. LE_s reaches its peak in the morning, due to capillary rise that happens during the night, allowing a high moisture profile in the morning. Therefore, it is maximum in the morning, and it decreases throughout the day with soil moisture. SEE follows the same cycle as LE_s, being maximum in the morning; it is not determined during night–time, when the evaporative demand is 0. A visual representation of the daily cycles of LE_s and LE_p, corresponding to TEC simulations using the *Wösten* (1997) pedotransfer function can be seen in Figure 4.2. Very small negative values obtained sometimes for LE_s and LE_p can be observed in the figure. This could be explained by dew and/or infiltration. In order to avoid potential problems on SEE, these negative values have been removed when deriving the SEE series.

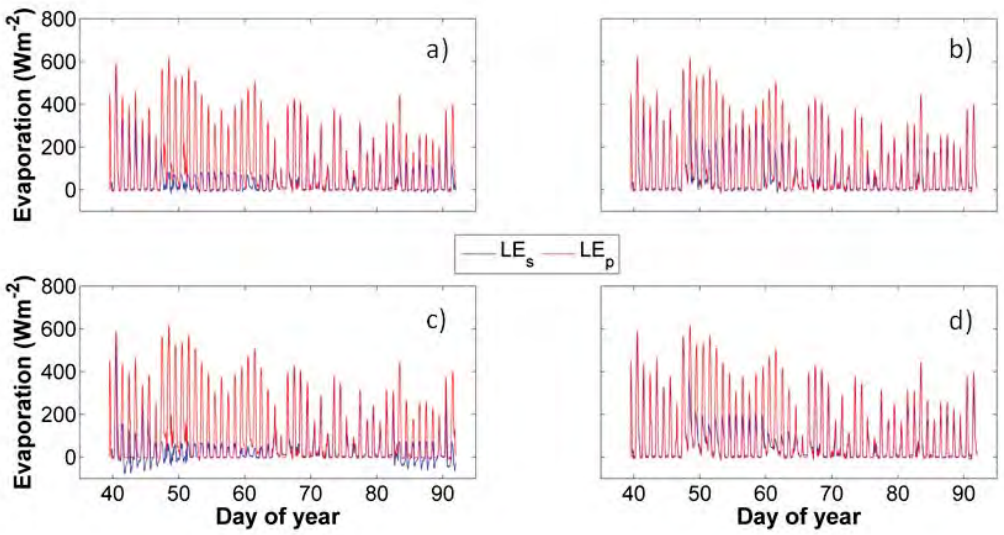


Figure 4.1: Soil evaporation (LE_s) and potential soil evaporation (LE_p) as derived from TEC, using four formulations of the pedotransfer functions: *Wösten* (1997) (a), *Cosby et al.* (1984) (b), *Rawls and Brackensiek* (1985) (c) and *Vereecken et al.* (1989) (d)

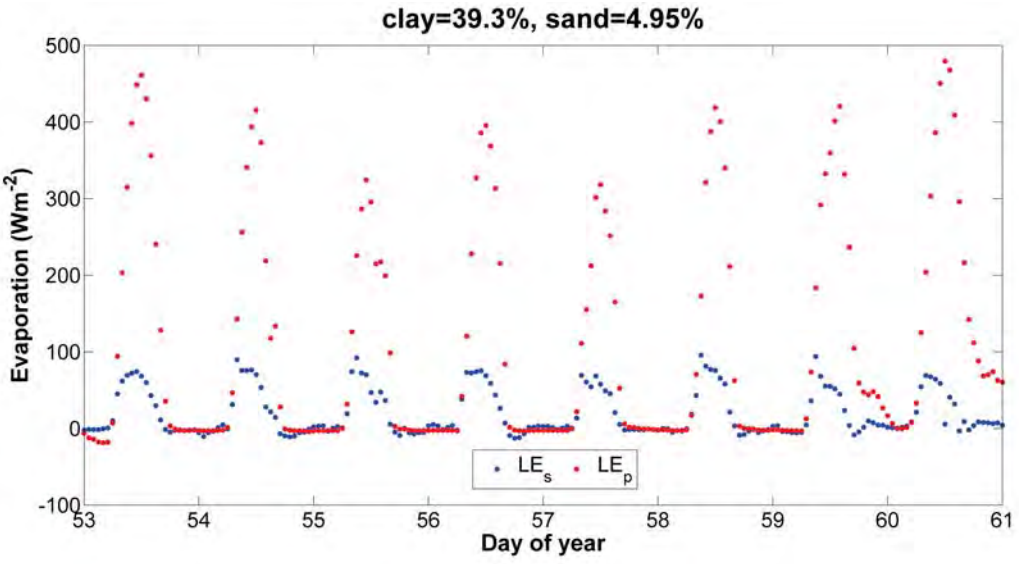


Figure 4.2: Daily cycle of soil evaporation (LE_s) and potential soil evaporation (LE_p). Data was generated using TEC, for the *Wösten* (1997) pedotransfer function.

Since soil texture is an important aspect in our modeling, four different soil types have been considered in generating the outputs. Soil texture has a low influence on the daily dynamics of LE_p, but a higher influence on LE_s estimates. A visual representation of LE_s estimates associated to the "wos" formulation is presented in Figure 4.3, for four different soil types.

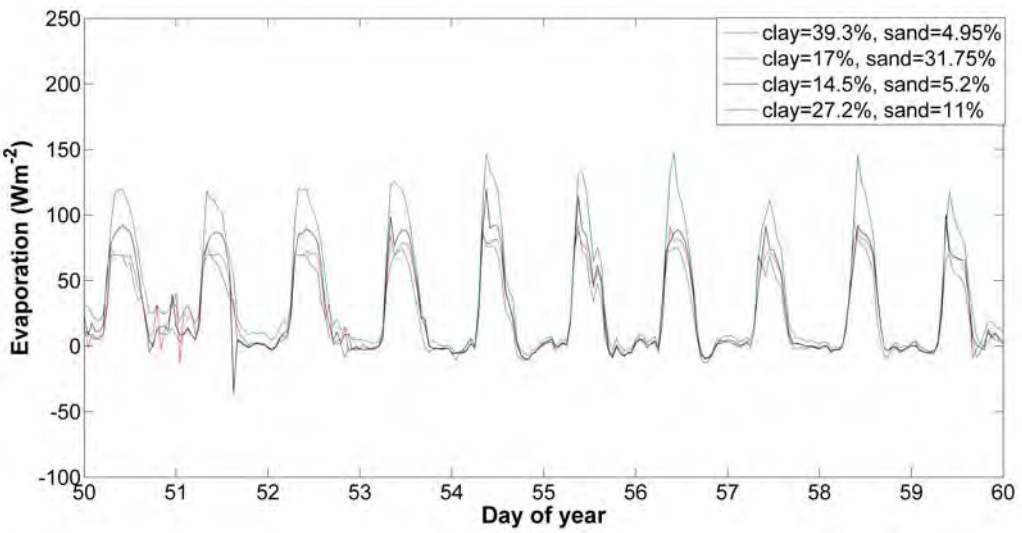


Figure 4.3: Evaporation (LE_s) generated using TEC with the "wos" formulation of the pedo-transfer function, using four different soil textures.

Soil texture clearly influences the magnitude of the LE_s estimates, with higher estimates registered for a texture configuration consisting of a clay percentage of 27.2 % and a sand percentage of 11 %. One can also observe that the daily dynamics obtained at night-time between days of year 50 and 51 is irregular, presenting several oscillations, regardless of the texture used.

4.3.3 Data pre-processing

Our variable of interest, SEE is obtained from individual LE_s and LE_p series derived from TEC. LE_p is influenced by the incoming radiation. Therefore, by dividing LE_s by LE_p, one neglects this meteorological component, which is supposed to be known. Already available simulations of LE_s and LE_p, given by the TEC model (*Chanzy et al.*, 2008), were used to derive the SEE, which was then used as input to GloMo.

The goal is to describe the temporal variability of soil evaporative efficiency for different soils. Several steps of data–preprocessing have been used in order to derive time series which can be further used in the global modeling approach. First of all, abnormal values of the SEE (inferior to 0 or superior to 1) have been filtered out of the derived SEE series. Then, a cubic spline has been applied to re–sample the original signal at an hourly time–step. The successive derivatives necessary for the global modeling approach have been calculated using a Savitzky-Golay filter. A visual representation of a SEE series after post–processing can be found in Figure 4.4.

The SEE series is characteristic of a clayey soil type, obtained over Avignon in southern France. The climatic sequence used in TEC and corresponding to this location is typically Mediterranean, with heavy rain and long periods of dryness (*Chanzy et al.*, 2008). We can observe that the SEE generally presents a daily cycle, peaking in the morning. However, important irregularities can be observed. For example, for day of year 53, the SEE presents two cycles. One possible explanation could be the clouds passing by, which would induce a change in the evaporative demand. Day of year 56 also presents an additional small peak around noon. Days of year 50 and 52, even though similar to each other, are very different from all other days; and even though they are similar to each other, day of year 52 has a faster increase, with a shorter

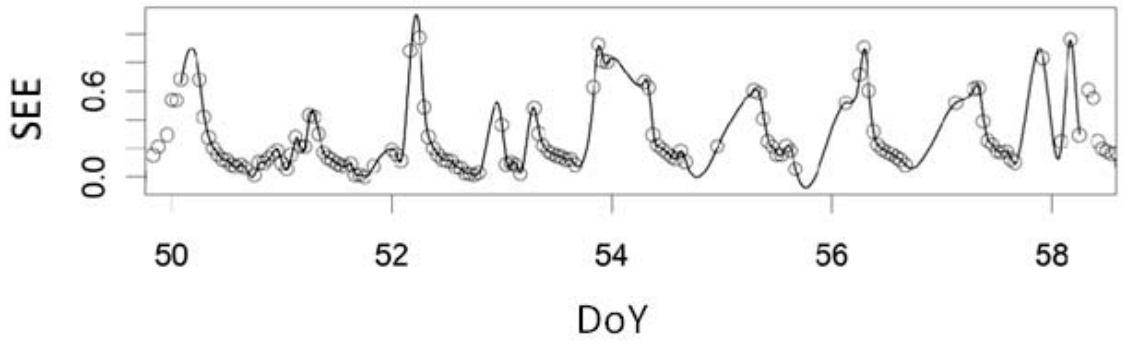


Figure 4.4: Original (circle) and processed (lines) SEE time series. The processed SEE time series is used as input to the global modeling.

time for the maximum value. The passages between days of year 53 and 54 and between 57 and 58 are also very singular. One can say that only days of year 55 and 57 are similar to each other.

4.4 Application and Results

The global modeling technique was applied to the four different time series presented in Fig. 4.3 (corresponding to four different soil types). However, a model could be obtained for only one series. This SEE time series is characteristic of a clayey soil (clay = 39.3%, sand = 4.95 %) and it spans a total of nine days. The obtained model is presented below.

4.4.1 SEE Global Model

The global modeling approach was first applied in dimension 3, but no satisfying model was obtained. A model of dimension 4 was obtained, which exhibits a complex behavior. The canonical form of the model is:

$$\begin{cases} \dot{SEE}_1 = SEE_2 \\ \dot{SEE}_2 = SEE_3 \\ \dot{SEE}_3 = SEE_4 \\ \dot{SEE}_4 = F(SEE_1, SEE_2, SEE_3, SEE_4) \end{cases} \tag{4.2}$$

The F function that corresponds to the obtained model can be written as:

$$\mathrm{F} = \mathrm{a}_1 SEE_4^2 + \alpha a_2 SEE_4^3 + a_3 SEE_3 + a_4 SEE_3 SEE_4 + a_5 SEE_3 SEE_4^2 + a_6 SEE_3^2 + a_7 SEE_3^3 + a_8 SEE_2 + a_9 SEE_2 SEE_4 + a_{10} SEE_2 SEE_4^2 + a_{11} SEE_2 SEE_3 + a_{12} SEE_2 SEE_3 SEE_4 + a_{13} SEE_2^2$$

$+ a_{14}SEE_2^2SEE_4 + a_{15}SEE_2^2SEE_3 + a_{16}SEE_2^2 + a_{17}SEE_1 + a_{18}SEE_1SEE_4^2 + a_{19}SEE_1SEE_3$
$+a_{20}SEE_1SEE_3SEE_4+a_{21}SEE_1SEE_3^2+a_{22}SEE_1SEE_2+a_{23}SEE_1SEE_2SEE_4+a_{24}SEE_1SEE_2$
$SEE_3 + a_{25}SEE_1SEE_2^2 + a_{26}SEE_1^2 + a_{27}SEE_1^2SEE_3 + a_{28}SEE_1^2SEE_2 + a_{29}SEE_1^3$
(4.3)

The values obtained for parameters a_i are reported in Table 4.1.

Table 4.1: Parameters obtained for the 4 dimensional model presented in Equation 4.3.

Parameter											
a_1	9.328630	a_6	2.3951097	a_{11}	4.7741240	a_{16}	−11.8738967	a_{21}	−3.6953403	a_{26}	0.2203366
a_2	−5.5659542	a_7	−2.3068889	a_{12}	−52.0660917	a_{17}	−0.1187456	a_{22}	2.0236262	a_{27}	−1.6147321
a_3	−2.2207074	a_8	−0.4506487	a_{13}	14.8157017	a_{18}	−16.7687883	a_{23}	−37.5016044	a_{28}	−2.2573969
a_4	3.9735013	a_9	23.2924234	a_{14}	−33.1414233	a_{19}	2.1924822	a_{24}	−7.7941180	a_{29}	−0.2418988
a_5	−24.0569615	a_{10}	26.4741226	a_{15}	−31.6852736	a_{20}	−6.0891262	a_{25}	−23.1251478		

The fitting parameter α (associated to parameter a_2) in Equation 4.3 is equal to 1 by default. The 4 dimensional chaotic model obtained exhibits a transient dynamics that however, for long time periods, converges towards a periodical solution. Modifying the fitting parameter to α = 0.99258 allows to maintain the chaotic behavior for longer time periods. The model is a reasonably good approximation of the original dynamics, as we can see from the original and the model phase portraits for various projections of the phase space, presented in Figure 4.5.

One can see that the dynamics of all variables, with the exception of SEE_4, can be reproduced with a good approximation by the model. In general, their maximum amplitudes are slightly larger than the amplitudes of the original data, but the general form is reproduced. The structure is well reproduced by the global model, even though slightly more dense than that of the original model. Variable SEE_4 is the only one that presents much larger amplitudes and a rather spiky behavior as opposed to the original series. This is an indication that this variable is less constraint than the other variables. As stated before, one of the difficulties in obtaining a model is the observability of the system (*Letellier et al.*, 2005), that depends on the variable used to reconstruct the system's dynamics. One possible explanation behind such a behavior for the variable SEE_4 can be explained, at least in part, by the definition of the SEE itself ($SEE = LE_s/LE_p$). Because SEE is defined as the ratio of LE_s over LE_p, this divison could pose a problem in the global modeling, since the variable of interest is the quotient of two complex behaviors. This can reduce or even compromise the observability of the system. It can also be one explanation as to why no global model was found for the three other SEE series, characteristic of different soil textures.

An illustration of the model in the time domain is shown in Figure 4.6. In the time domain, the obtained time series does not reproduce the original time series. This difference of time evolution is completely expected and directly arises from the chaotic regime of the dynamics that leads to the extremly rapid divergence between the observed and model time evolutions. However, even taking this element into account, the model cannot be considered as satisfying. In particular, the reproduced series presents faster oscillations than the daily signal. One would expect to obtain a model whose behavior oscillates around a daily cycle (because the LE_s signal is directly related to the day–night cycle). The model does not reproduce the daily cycle, but only faster variations. The reason is that the daily signal is not found (or almost not) in the observed dataset itself. The observed signal sometimes exhibits a full daily cycle. But it also shows very rapid, very irregular and complex changes (that are weakly predictable). And it is

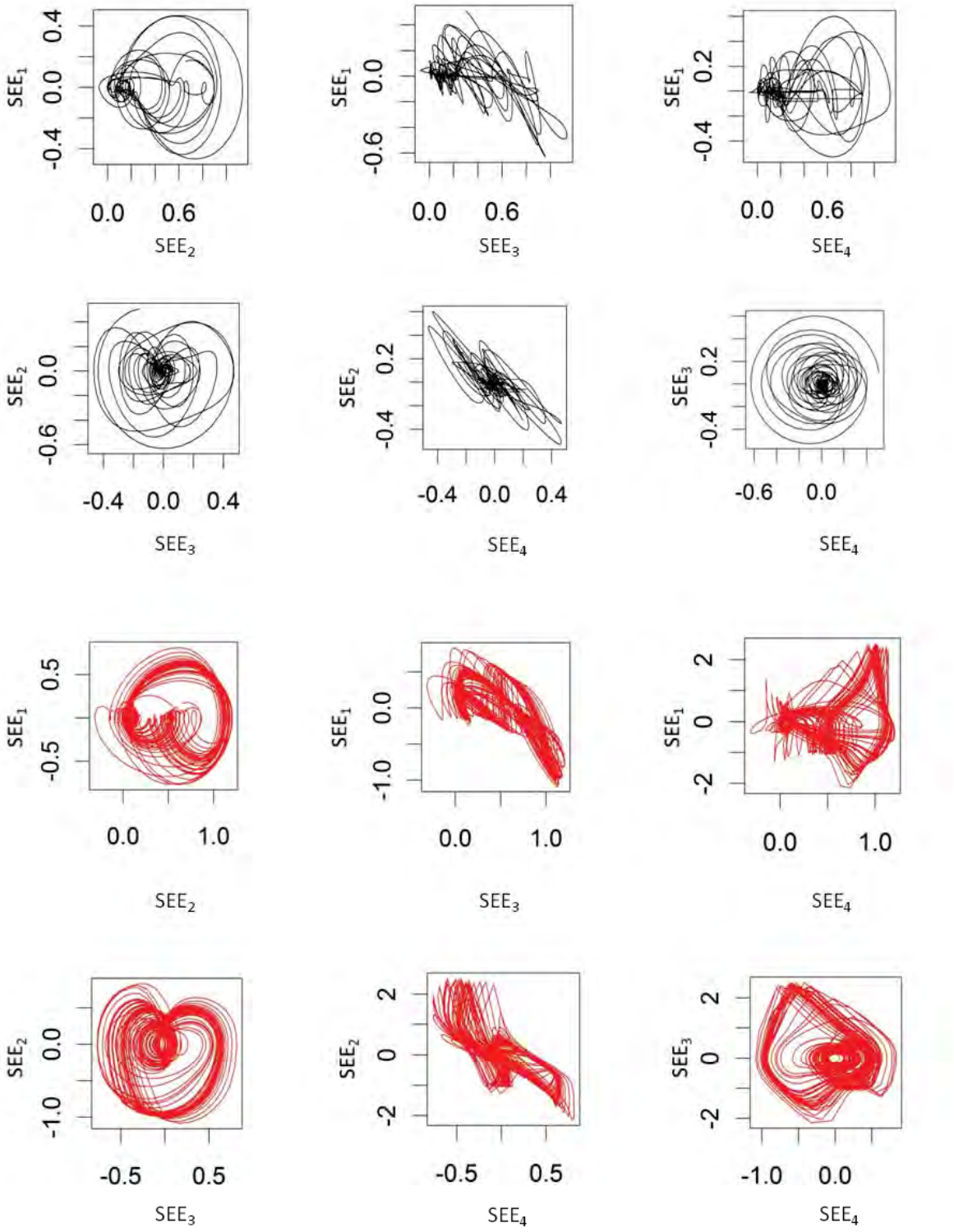


Figure 4.5: Original (black) and model (red) phase portraits for various projections of the phase space.

precisely these properties of the behavior that are found in the model.

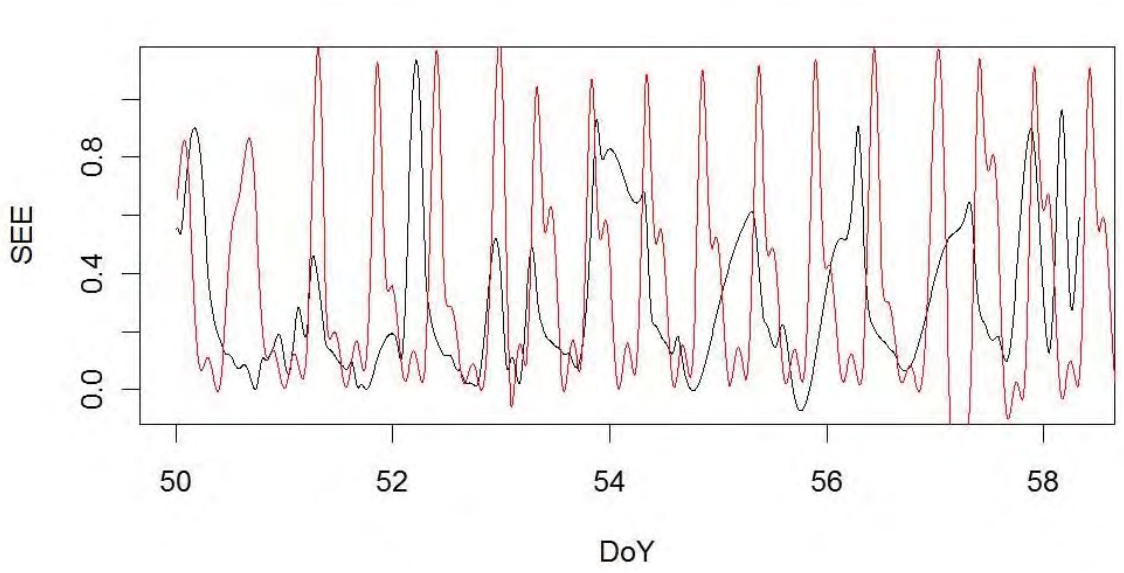


Figure 4.6: Original (black) and reconstructed (red) SEE time series.

The daily cycle is expected to be obtained because we know it, but it is not obvious in the original data either, which also presents several cycles per day. In particular, multimodal daily cycles can be observed for days of year 53 and 51. An important element to keep in mind in understanding the model's dynamical behavior is that the model obtained is autonomous: no forcing is used. In other words, there is nothing to synchronize the model on the daily cycle, a fact which is contrary to other existing models that are usually used to model such dynamics. The model is able to reproduce the complex dynamics of the behavior (see phases portraits) but is also able to perform satisfying forecasts only on a short–time scale. Although not synchronized to the observed signal, numerous properties of the complex behavior found in the original signal can be retrieved in the model simulation: very quick increases sometimes, more progressive ones other times, decreases in single or double stages, small oscillations during the night. Despite this prediction ability at short time scales, the forecasts become obviously impossible at time horizons larger than 3 hours. In order to have a comparison between the data and the model over time, several succesive simulations were launched starting from different initial conditions, that were then compared to the data. The SEE model obtained (Equation 4.3) is used to perform the simulations. The model thus simulates the variables SEE_1, SEE_2, SEE_3 and SEE_4 and then starting from the chosen initial conditions, one can look at the evolution of the predictions either in the time domain (Figure 4.7) or in the phase domain (Figure 4.8).

Figure 4.7 is the visual representation of the predictability analysis, in the time domain. It is the same as Figure 4.6, but predictions are made starting from different initial conditions. The error between the prediction and the data is also shown. It shows that the model can qualitatively anticipate complex incoming variations, even though it's not in a completely systematic way. Also, it shows that some behaviors are more difficult to predict than others. All in all, the model appears to be quite effective when looking at both the time and the phase domain representations, for sufficiently short prediction horizons (three hours in this case).

Figure 4.8 provides a visual representation in the phase space of the dynamics constructed from the successive dynamical states experienced by the system, as well as a representation of

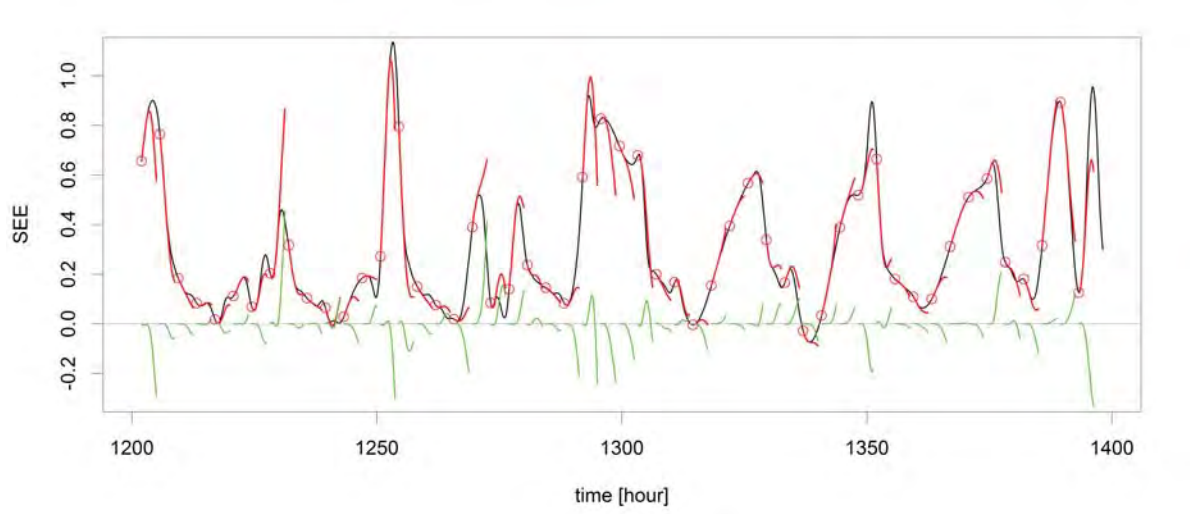


Figure 4.7: Original (black) and reconstructed (red) SEE time series. The reconstructed SEE corresponds to simulations starting from different initial conditions (red circles). The green curves represent the error between the model and the original data.

the phase portraits of the error between the prediction and the data.

The low forecasting skills of the model at large time–scales can also be interpreted in a physical way. An important observation is that the daily SEE series is different from one day to another and also presents strong intraday variations, with the daily expected cycle being sometimes hard to be observed. Intraday variability is very probably due to the presence of clouds that will drastically change the atmospheric and the evaporative conditions (via the evaporative demand LE_p) and will "blur" the daily signal. This quick and unpredictable variability of the dynamics seems to be well represented by the model. This is an important achievement of the global model, since this type of behavior is particularly erratic and difficult to formulate in equations. The model is able to retrieve this behavior (in an approximate way) but since this type of behavior is weakly predictable (at least from a single variable), the resulting model is also very sensitive to initial conditions. The global model thus has a high sensitivity to initial conditions, sensitivity that is directly linked to cloud cover.

The predictability of such a cloud cover is very short–term, and therefore one cannot expect to get better results with the obtained model. A prediction over a period of several days can have only a statistical meaning.

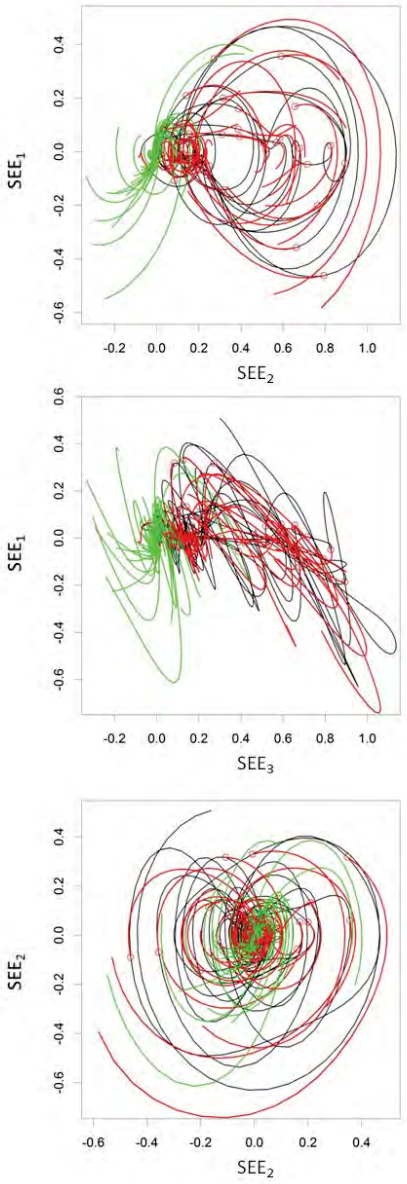


Figure 4.8: Phase portraits of the original (black) and reconstructed (red) SEE series for various projections of the phase space. The reconstructed series corresponds to simulations starting from different initial conditions (red circles). The green curves represent the phase portraits of the error between the model and the original data.

4.4.2 Model validation

From a predictability point of view, validating a model is a difficult task since most of the environmental variables have a limited predictability in terms of dynamics. In order to validate the model obtained, the forecasting error growth of the model can be estimated (*Mangiarotti et al.*, 2012b). Since a chaotic system is characterized by the high sensitivity to the initial conditions, the error growth depends on the stability of the initial conditions as well as on the model skills. Starting from different initial conditions, a large number of forecasts are launched, for horizons spanning 9 hours. The forecasting error $e_t(h)$ is defined as the error corresponding to time $t + h$ when predicting from time t for the horizon h. It is computed as the difference between the prediction made at time $t+h$ and the data at time $t+h$. So this error depends both on the initial state of the system at time t, and on the prediction horizon h. This error growth is represented by a two dimensional diagram in Figure 4.9.

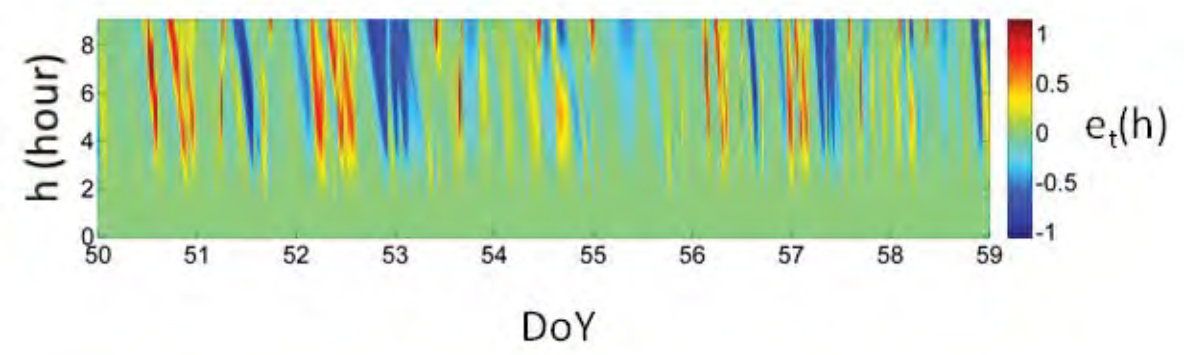


Figure 4.9: Error growth $e_t(h)$ (in color) as a function of t (expressed as the Day of Year - DoY - in the abscissa) and the horizon of prediction h (expressed in hours, as the ordinate).

One can notice that as the prediction horizon increases, so does the error. Two different patterns can be observed: 1) vertical traits, that correspond to errors caused by poor or unstable initial conditions from which the prediction is made and 2) diagonal traits, that correspond to poorly predictable approaching events.

The standard deviation of the error growth is plotted (in corresponding SEE units and in percentage) as a function of the prediction horizon in Figure 4.10. The error level is an increasing function of the forecasting horizon, with low values obtained at short horizons. Results show that, for example, for a maximum error level of 40 % of the signal variance (corresponding to an error of 0.1 in terms of SEE), the horizon of predictability is close to 3 hours, approximately one third of the diurnal part of the day. If higher precision is expected, for example with an error level of 20 % (corresponding to an error of 0.07 in terms of SEE), then the horizon of predictibility is of $\sim$ 2 hours. If predictions are sought after 6 hours, then the error might be larger than 100 % (corresponding to an error larger than 0.3 in terms of SEE) of the signal variance.

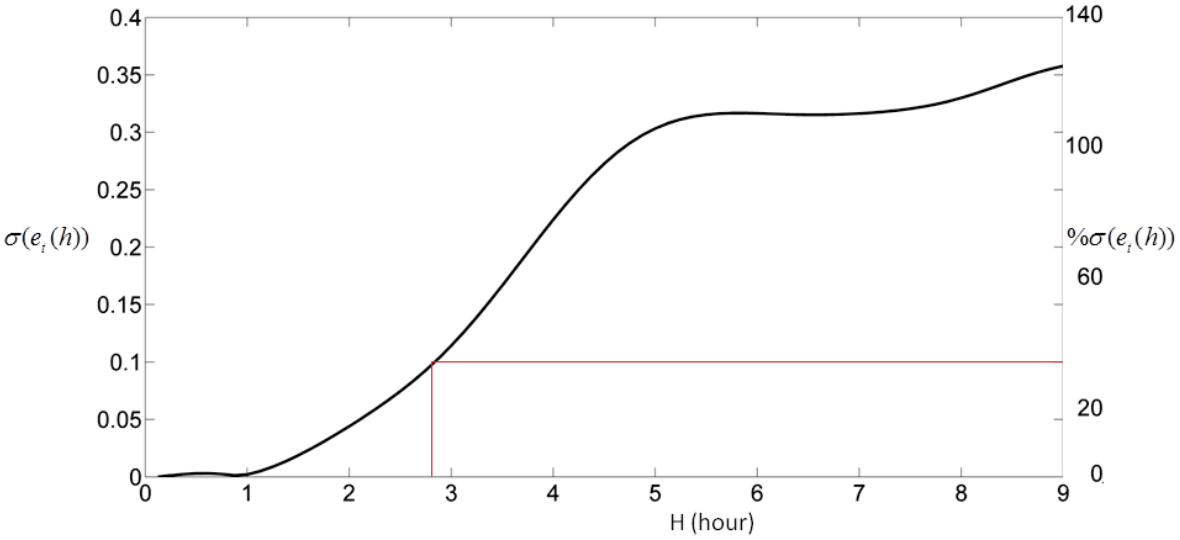


Figure 4.10: The standard deviation of the error growth is plotted (in corresponding SEE units and in percentage) as a function of the prediction horizon.

4.5 Summary and Conclusions

A four dimensional chaotic model was obtained for the daily cycle of soil evaporative efficiency. This model is able to reproduce the dynamics of the variable with a good approximation in the phase domain. The forecasting error growth study shows that the prediction horizon is of 2 hours for an error threshold equal to 20% of the signal variance, which is poor in terms of model predictability. However, these results are interesting for various reasons. To the best of our knowledge, it is the very first time that a chaotic model is obtained for the SEE. It also shows that the SEE dynamics can be approximated by a low–dimensional autonomous model. From a theoretical point of view, it is also interesting to note that only very few low-dimensional models could be directly obtained for environmental dynamics, and that four–dimensional models are even rarer. One of the difficulties in obtaining such a model is linked to the variable itself, as the SEE is obtained as the quotient of two dynamical variables, which could lead to singularities in the model. Moreover, the evaporative demand is not defined during night–time, which can be another source of potential singularities in the model. Another difficulty is given by the non–clear sky conditions and the influence on soil evaporation and evaporative demand. More specifically, presence of clouds could be the reason behind the faster oscillations of the SEE series observed some days, which is not representative of the daily SEE cycle. All this can affect the observability of the variable and lead to difficulties in the model reconstruction. The ability to obtain an autonomous model for soil–atmosphere transfers remains however encouraging as it shows the ability to model this dynamics from a limited number of observed variables (here one single variable). From the perspective of a multi–scale modelization of soil evaporation from remote sensing data, the model's skills, at this stage, remain however limited. Moreover, it seems difficult, without any improvements, to model such dynamics from a spatially distributed point of view. Data assimilation could be a potential lead into further investigating the model's skills,

by taking into account a forcing in terms of meteorological and soil moisture data. Nevertheless, obtaining such an autonomous model is a feat in itself, considering all the restrictions mentioned above.

4.6 Article

L'Efficacité Évaporative du Sol modélisée par approche globale

V. Stefan[1], S. Mangiarotti[1], O. Merlin[1,2] & A. Chanzy[3]

[1] CESBIO, UPS-CNRS-CNES-IRD, 18 av. Ed. Belin, 31401 Toulouse cedex 9, France
[2] Faculté des Sciences Semlalia Marrakech (FSSM), Marrakech, Maroc
[3] Institut National de la Recherche Agronomique (INRA), Avignon, France
sylvain.mangiarotti@ird.fr

Résumé. La modélisation des transferts sol-atmosphère s'appuie généralement sur des formulations considérant la dynamique climatique comme un forçage. Dans cette étude, la modélisation par technique globale est appliquée aux cycles journaliers d'Efficacité Évaporative du Sol (SEE) pour obtenir un modèle autonome, sans faire d'hypothèse sur la structure algébrique du modèle. Un modèle en équations aux dérivées ordinaires de dimension quatre est ainsi obtenu. Bien que le modèle converge vers un cycle limite, il présente néanmoins une dynamique transitoire chaotique longue et en bonne cohérence avec les données originales.

Abstract. In Soil-Atmosphere transfer models, the climatic component is generally considered as an external forcing. In the present study, the global modeling technique is applied to daily cycles of Soil Evaporative Efficiency in order to obtain an autonomous model for this variable. A four-dimensional model of Ordinary Differential Equations is obtained. Although converging to a limit cycle, the model exhibits a long chaotic transient which is consistent with the original data.

1 Introduction

La compréhension de la dynamique du contenu en eau des sols est très importante. L'eau évaporée par le sol joue en effet un rôle important dans le cycle de l'eau mais cette dynamique est difficile à quantifier de manière spatialisée. D'un point de vue appliqué, la quantité d'eau évaporée dans les régions semi-arides peut représenter ∼80% des ressources disponibles. Pour modéliser la dynamique de l'eau du sol, on fait généralement appel à des modèles nécessitant de disposer d'un forçage climatique. Si l'on ne dispose pas d'un tel forçage, ce qui est souvent le cas lorsqu'on s'intéresse à des dynamiques spatialement distribuées, il peut alors être utile de disposer de modèles alternatifs ne nécessitant pas un tel forçage.

La technique de modélisation globale s'appuie sur la théorie des systèmes dynamiques non linéaires. Plutôt que de chercher à résoudre les équations analytiquement, cette théorie privilégie le recours à l'espace des phases pour étudier les comportements dynamiques. L'une des raisons de recourir à cette théorie vient du fait qu'il est – en principe – possible de reconstruire un tel espace pour un comportement dynamique donné à partir d'une série temporelle unique [1]. En pratique, ce résultat n'est toutefois pas toujours garanti et cette possibilité tient en particulier à un problème d'observabilité [2] : en raison des relations de nonlinéarités qui existent entre les variables d'un système étudié, les observables n'offrent pas la même observabilité. La possibilité de reconstruire un comportement dynamique à partir d'une séries de mesures observationnelles plutôt qu'une autre va donc dépendre de l'observable choisie.

La technique de modélisation globale a été développée au cours des années 1990 [3,4,5]. Son application à des dynamiques environnementales est beaucoup plus récente [6]. Au cours des dernières années, des développements méthodologiques ont permis de multiplier le nombre des modèles chaotiques pouvant être obtenus directement à partir de séries observationnelles. Ces travaux ont notamment permis d'obtenir le premier modèle faiblement dissipatif directement issu de mesures observationnelles [7] ; d'autres modèles ont été obtenus pour les cycles de surfaces enneigées [8] ou pour la dynamique des épidémies de peste [9].

L'objectif de cette étude est de mettre en évidence la possibilité de recourir à la technique globale pour modéliser la dynamique de l'eau dans le sol. Dans cette étude nous nous intéresserons plus spécifiquement à l'Efficacité Évaporative du Sol car cette variable présente une dynamique plus directement liée aux

© **Non Linéaire Publications**, Avenue de l'Université, BP 12, 76801 Saint-Étienne du Rouvray cedex

propriétés du sol [10]. Dans le paragraphe suivant sont présentées les données synthétiques utilisées dans cette étude ; le modèle TEC utilisé pour générer ces données y sera donc présenté auparavant. La méthodologie sera présentée au paragraphe 3. Les résultats seront ensuite présentés au paragraphe 4.

2 Données

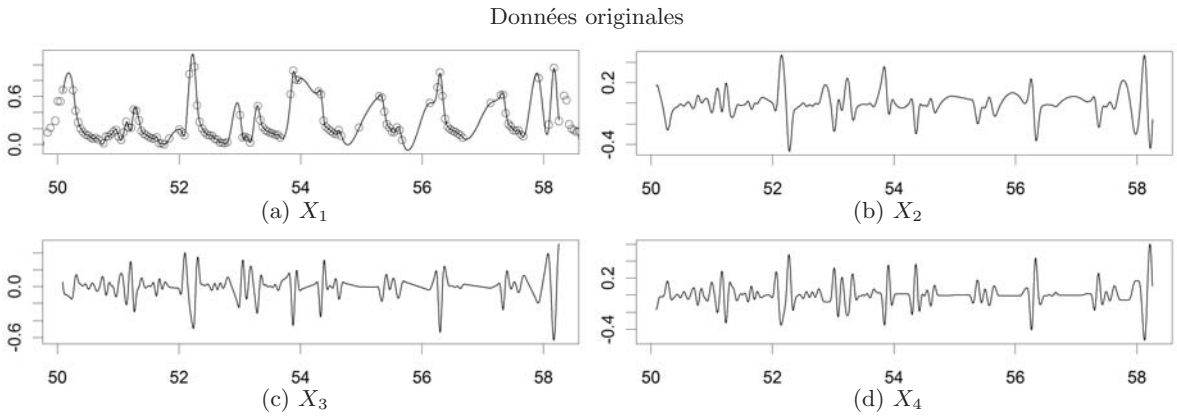
2.1 Modèle TEC

Le modèle Transfert Eau Chaleur (TEC) [11] a été utilisé pour générer la série temporelle de SEE utilisée dans cette étude. Le modèle TEC est un modèle mécaniste s'appuyant sur un paramétrage physique et hydrologique pour décrire les processus impliqués dans l'évaporation du sol. Le sol est modélisé par un système multi-couche dans lequel les flux de chaleur et de masse sont gouvernés par des équations différentielles et d'où sont calculées l'évaporation potentielle du sol E_0 (correspondant à une situation où il serait suffisamment alimenté en eau), et l'évaporation effective E, et d'où la SEE peut directement être déduite : $SEE = E/E_0$.

2.2 Prétraitements

La variable SEE a été formulée pour tenter de s'affranchir de la composante météorologique (c'est l'objet de la division par E_0). Cette normalisation pose toutefois une double difficulté. D'un point de vue algébrique, cette formulation est intéressante à première vue. Elle revient toutefois à considérer le quotient de deux comportements dynamiques complexes qui peut réduire, voire compromettre l'observabilité du système. D'un point de vue pratique, les variables E et E_0 tendent vers zéro pendant la nuit. Pour cette raison, seules les valeurs diurnes sont généralement prises en compte.

Pour la présente étude, nous nous sommes donc concentrés sur des séries temporelles pour lesquelles les valeurs d'évaporation potentielles sont telles que $E_0 \gg 0$. Ce critère nous a conduit à sélectionner une série temporelle de SEE d'un longueur de neuf jours (Fig. 1) issue d'une simulation caractéristique d'un sol argileux par le modèle TEC. Une spline cubique a d'abord été appliquée au signal pour le ré-échantillonner à un pas de temps horaire. Les dérivées nécessaires à la technique de modélisation globale ont été calculées en utilisant un filtre de type Savitzky-Golay [12].



(a) X_1 (b) X_2 (c) X_3 (d) X_4

Figure 1. Séries temporelles originales d'Efficacité Évaporative du sol $X_1 = SEE$ issue du modèle TEC (cercles) [11] ou filtrées (traits pleins) (a) ; et ses dérivées première X_2 (b), seconde X_3 (c) et troisième X_4 (d).

3 Méthodologie

La modélisation globale s'appuie sur l'espace des phases pour obtenir des équations de comportements dynamiques directement à partir de séries temporelles de mesures [5]. Dans sa forme différentielle et partant d'une variable observationnelle unique X_1, cette technique vise à obtenir des modèles de forme canonique

$$\begin{cases} \dot{X}_i = X_{i+1} & \text{pour } i = 1..(n-1) \\ \dot{X}_n = F(X_1, X_2, ..., X_n), \end{cases} \tag{1}$$

où n est la dimension du modèle et les X_i les dérivées successives de la variable observée X_1. En pratique, modéliser un comportement dynamique par approche globale revient à obtenir une approximation de la fonction F qui permette de reproduire la dynamique observée de façon satisfaisante.

Identifier cette fonction sans disposer de structure *a priori* est très difficile et nécessite de disposer d'outils puissants. Les outils PoMoS et GloMo et la suite de leurs développements ont pu être testés avec succès, notamment à la modélisation du système de Rössler à partir de sa variable z présentant un très faible degré d'observabilité (voir [13] pour discussion). L'approche a également été appliquée à plusieurs variables environnementales et les cas d'application apparaissent de plus en plus nombreux (voir par exemple dans le présent ouvrage [14,15]).

4 Analyses

Les portraits de phases différentiels reconstruits à partir des séries temporelles tirées du modèle TEC (voir section 2.1) sont présentées en Figure 2 suivant différentes projections. Visuellement, ces projections exhibent des formes très distinctes les unes des autres.

La technique de modélisation globale a d'abord été appliquée en dimension trois puis en dimension quatre. Aucun modèle de dimension $n = 3$ n'a pu être obtenu et un seul modèle de dimension $n = 4$ a été obtenu. La fonction F de ce modèle prend la forme

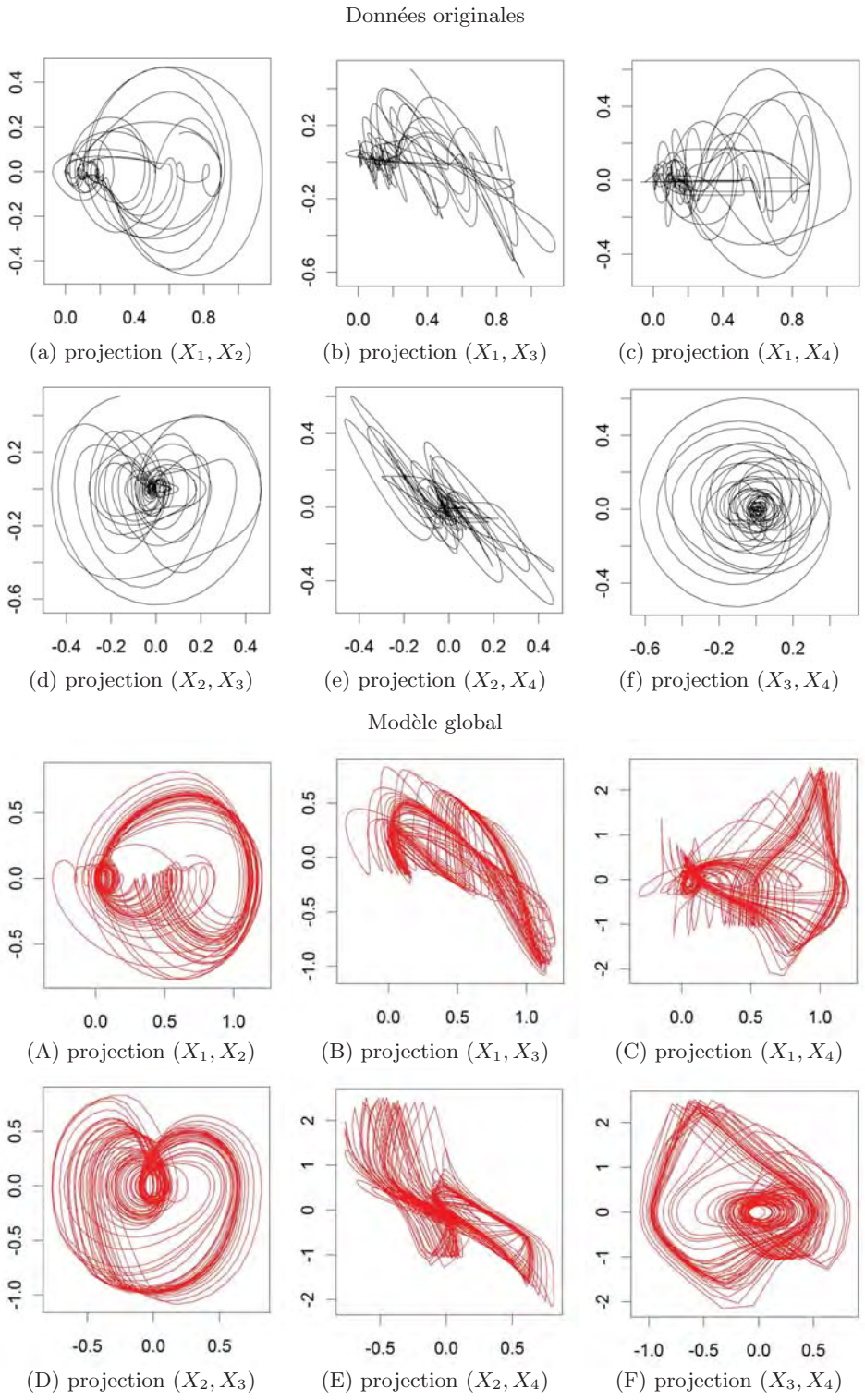
$$\begin{aligned} F(X_1, X_2, X_3, X_4) = {} & a_1 X_4^2 + \alpha\ a_2 X_4^3 + a_3 X_3 + a_4 X_3 X_4 + a_5 X_3 X_4^2 + a_6 X_3^2 + a_7 X_3^3 + a_8 X_2 \\ & + a_9 X_2 X_4 + a_{10} X_2 X_4^2 + a_{11} X_2 X_3 + a_{12} X_2 X_3 X_4 + a_{13} X_2^2 + a_{14} X_2^2 X_4 + a_{15} X_2^2 X_3 \\ & + a_{16} X_2^3 + a_{17} X_1 + a_{18} X_1 X_4^2 + a_{19} X_1 X_3 + a_{20} X_1 X_3 X_4 + a_{21} X_1 X_3^2 + a_{22} X_1 X_2 \\ & + a_{23} X_1 X_2 X_4 + a_{24} X_1 X_2 X_3 + a_{25} X_1 X_2^2 + a_{26} X_1^2 + a_{27} X_1^2 X_3 + a_{28} X_1^2 X_2 + a_{29} X_1^3, \end{aligned} \tag{2}$$

où les paramètres a_i sont donnés en Table 1, et où le paramètre d'ajustement α (associé au coefficient a_2) est pris égal à un. Le modèle ainsi obtenu présente une dynamique transitoire chaotique. Néanmoins, pour des temps longs, le modèle converge vers une solution périodique. Une légère modification du paramètre d'ajustement tel que $\alpha = 0.99258$ permet de préserver une dynamique chaotique sur des temps beaucoup plus long. Il s'agit toutefois d'un chaos peu développé.

La dynamique du modèle global montre une bonne cohérence avec les données comme le montrent la comparaison des portraits de phases originaux et modélisés (Fig. 2). La modélisation de la variable X_4 présente toutefois des amplitudes beaucoup plus fortes que les observations, ce qui semble indiquer que cette variable est moins bien contrainte que les autres variables.

Paramètres															
	a_1	9.3283630	a_6	2.3951097	a_{11}	4.7741240	a_{16}	-11.8738967	a_{21}	-3.6953403	a_{26}	0.2203366
	a_2	-5.5659542	a_7	-2.3068889	a_{12}	-52.0660917	a_{17}	-0.1187456	a_{22}	2.0236262	a_{27}	-1.6147321
	a_3	-2.2207074	a_8	-0.4506487	a_{13}	14.8157017	a_{18}	-16.7687883	a_{23}	-37.5016044	a_{28}	-2.2573969
	a_4	3.9735013	a_9	23.2924234	a_{14}	-33.1414233	a_{19}	2.1924822	a_{24}	-7.7941180	a_{29}	-0.2418988
	a_5	-24.0569615	a_{10}	-26.4741226	a_{15}	-31.6852736	a_{20}	-6.0891262	a_{25}	-23.1251478		

Table 1. Paramètres a_i du modèle global présenté Eq. (1).



(a) projection (X_1, X_2) (b) projection (X_1, X_3) (c) projection (X_1, X_4)

(d) projection (X_2, X_3) (e) projection (X_2, X_4) (f) projection (X_3, X_4)

(A) projection (X_1, X_2) (B) projection (X_1, X_3) (C) projection (X_1, X_4)

(D) projection (X_2, X_3) (E) projection (X_2, X_4) (F) projection (X_3, X_4)

Figure 2. Portraits de phase du jeu de données original (a-f) et du modèle (A-F) suivant les projections correspondant aux différentes combinaisons des axes (X_1,X_2,X_3,X_4).

4.1 Validation

Dans les contextes de prévisibilité limitée comme c'est le cas pour un très grand nombre de dynamiques environnementales, valider un modèle est toujours une tâche difficile. Pour valider le modèle obtenu ici, on se propose de se baser sur les capacités prédictives du modèle en s'appuyant sur une statistique de la croissance de l'erreur de prévision du modèle [16].

Pour cela, un grand nombre de prévisions est lancé, pour des horizons τ compris entre 0 et 9 heures, à partir d'un ensemble de conditions initiales offrant une bonne couverture du jeu de données disponible. L'erreur de prévision $e(t, \tau)$ dépend donc à la fois de l'état initial du système au temps t et de l'horizon de prévision au temps $t+\tau$. Cette erreur peut être représentée par le diagramme à deux dimensions présenté en Figure 3. Plusieurs éléments peuvent y être observés à commencer par la croissance progressive de l'erreur avec la croissance de l'horizon de prévision τ (présenté en ordonnée). On peut y noter deux types de structures au sein de ce diagramme : (1) les structures verticales qui résultent de la difficulté d'effectuer une prévision à partir d'une condition initiale donnée ; (2) les structures diagonales qui proviennent de la difficulté de prévoir un événement particulier. (Attention, suivant la convention de représentation choisie, les structures verticales et diagonales peuvent être inversées, voir [16] pour une interprétation détaillée de ce type de figure). La croissance d'erreur peut être obtenue en faisant la statistique des erreurs en

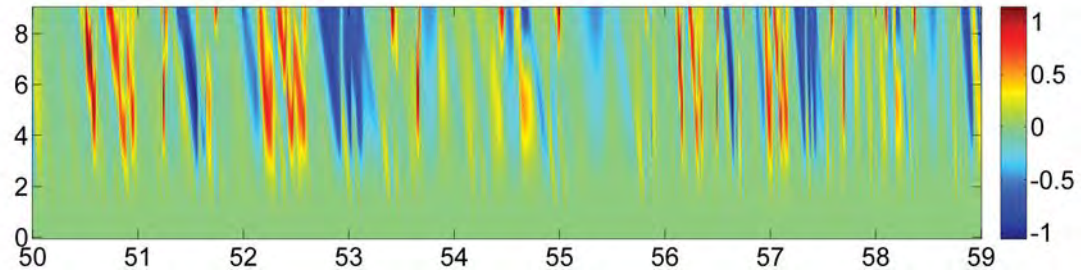


Figure 3. Erreur de prévision $e(t, \tau)$ (en couleur) en fonction de la date t (tracée en ordonnée, exprimée en jour) et de l'horizon de prédiction τ (tracé en abscisse, exprimé en heure).

fonction de l'horizon τ. La croissance d'erreur ainsi obtenue est présentée en Figure 4 en pourcentage de la variance du signal. Ce graphe montre que la croissance d'erreur est très rapide puisqu'au bout de 3 heures, l'erreur atteinte est proche de 40% de la variance du signal. Si cette analyse de prévisibilité permet d'apporter un premier élément de validation, d'un point de vue prévisionnel, le modèle ainsi obtenu apparaît néanmoins d'un intérêt très relatif.

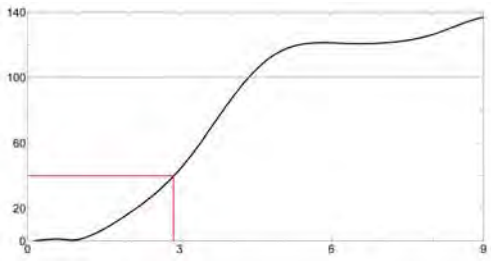


Figure 4. Erreur de prédiction (rmse) exprimée en pourcentage de la variance du signal, en fonction de l'horizon de prédiction τ (en heure). Pour un niveau d'erreur de 40% rmse, l'horizon de prévisibilité est de $\sim$3 heures.

5 Conclusions

Un modèle autonome est obtenu pour les cycles journaliers de l'Efficacité Évaporative du Sol. Il s'agit d'un modèle chaotique de dimension quatre. Ce modèle permet de reproduire la dynamique observée de façon très cohérente. Ses capacité prédictives restent toutefois limitées à un horizon de prévisibilité de 3 heures (pour un seuil d'erreur de 40% de la variance du signal).

Les difficultés rencontrées pour obtenir ce modèle provient très vraisemblablement du choix de la variable utilisée qui, construite comme le quotient de deux variables dynamiques, peut conduire à l'existence de singularités. L'utilisation de cette variable se traduit vraisemblablement par une diminution du degré d'observabilité du système modélisé qui peut gêner la reconstruction du modèle.

Le fait de pouvoir obtenir un modèle autonome pour les transferts Sol-Atmosphère reste toutefois encourageant puisqu'il montre la possibilité de modéliser cette dynamique à partir d'un nombre restreint de variables observées et laisse entrevoir la possibilité de modéliser cette dynamique de manière spatialement distribuée, en s'appuyant par exemple sur des séries observationnelles issues de la télédétection spatiale.

Références

1. F. Takens, Detecting strange attractors in turbulence, *Lecture Notes in Mathematics*, **898**, 366-381, 1981.
2. C. Letellier, L. A. Aguirre, J. Maquet, Relation between observability and differential embeddings for nonlinear dynamics, *Physical Review E*, **71** (6), 066213, 2005.
3. G. Gouesbet & C. Letellier, Global vector field reconstruction by using a multivariate polynomial L_2-approximation on nets, *Physical Review E*, **49** (6), 4955-4972, 1994.
4. C. Letellier, L. Le Sceller, P. Dutertre, G. Gouesbet, Z. Fei & J. Hudson, Topological characterization and global vector field reconstruction of an experimental electrochemical system, *The Journal of Physical Chemistry*, **99** (18), 7016–7027, 1995.
5. C. Letellier, L. A. Aguirre & U. S. Freitas, Frequently asked questions about global modeling, *Chaos*, **19**, 023103, 2009.
6. J. Maquet, C. Letellier & L. A. Aguirre, Global models from the Canadian lynx cycles as a direct evidence for chaos in real ecosystems, *Journal of mathematical biology*, **55** (1), 21–39, 2007.
7. S. Mangiarotti, L. Drapeau, M. Huc & C. Letellier, Two chaotic global models for cereal crops cycles observed from satellite in Northern Morocco, *Chaos*, **24**, 023130, 2014.
8. S. Mangiarotti, Modélisation globale et caractérisation topologique de dynamiques environnementales : de l'analyse des enveloppes fluides et du couvert de surface de la Terre à la caractérisation topolodynamique du chaos, Habilitation to direct Researches, Université de Toulouse 3, 2014.
9. S. Mangiarotti, Low dimensional chaotic models for the plague epidemic in Bombay (1896–1911), *Chaos, Solitons & Fractals*, **81**(A), 184-196, 2015.
10. O. Merlin, V. G. Stefan, A. Amazirh, A. Chanzy, E. Ceschia, S. Er-Raki, P. Gentine, T. Tallec, J. Ezzahar, S. Bircher, J. Beringer & S. Khabba, Modeling soil evaporation efficiency in a range of soil and atmospheric conditions using a meta-analysis approach, *Water Resources Research*, **soumis**, 2016.
11. A. Chanzy, M. Mumen & G. Richard, Accuracy of the top soil moisture simulation using a mechanistic model with limited soil characterization, *Water Resources Research*, **44**, W03432, 2008.
12. A. Savitzky & M. J. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Analytical chemistry*, **36** (8), 1627–1639, 1964.
13. S. Mangiarotti, R. Coudret, L. Drapeau & L. Jarlan, Polynomial search and global modeling : two algorithms for modeling chaos, *Physical Review E*, **86**(4), 046205, 2012.
14. S. Mangiarotti, A. K. Sharma, S. Muddu, S. Corgnes & L. Ruiz, La modélisation du chaos appliquée à la classification des couverts agricoles (bassin versant de Berambadi, Inde du Sud), *Rencontres du Non Linéaire*, **19**, 2016.
15. M. Huc & S. Mangiarotti, Un modèle chaotique quadridimensionnel pour l'épidémie de fièvre hémorragique à virus Ebola en Afrique de l'Ouest, *Rencontres du Non Linéaire*, **19**, 2016.
16. S. Mangiarotti, P. Mazzega, P. Hiernaux & E. Mougin, Predictability of vegetation cycles over the semi-arid region of Gourma (Mali) from forecasts of AVHRR-NDVI signals, *Remote Sensing of Environment*, **123**, 246-257, 2012.

Chapter 5

Evaporation–based disaggregation of soil moisture data

Contents

5.1 Introduction 120

5.2 Methodology 121

5.2.1 DISPATCH 121

5.2.2 Temperature–based SEE model 121

5.2.3 Soil moisture–based SEE model 124

5.2.4 DISPATCH–E 125

5.3 Results 127

5.3.1 DISPATCH/DISPATCH–E sensitivity analysis 127

5.3.2 Temperature–based *versus* soil moisture–based SEE estimates 137

5.3.3 Validation of disaggregation methodology 139

5.4 Summary and Conclusions 141

5.5 Article 142

5.1 Introduction

The objective of this PhD is to improve the representation of soil evaporation (E) from a spatio–temporal point of view, in a range of different soil and atmospheric conditions, by making use of available remote sensing data. By taking into account the monitoring capabilities of E–related quantities through remote sensing means, several studies have shown the possibility to combine shortwave–derived vegetation cover, thermal–derived land surface temperature and microwave–derived soil moisture in order to partition evapotranspiration into soil evaporation and plant transpiration and therefore successfully retrieve E. Under non–energy limited conditions, soil moisture (SM) is amongst the main factors controlling soil evaporation, and therefore most of the climatological, meteorological, hydrological and agricultural applications require SM data at high resolution (HR). Since the resolution of the SM products retrieved from microwave sensors ranges from 40 to 60 km (*Njoku et al.*, 2003; *Kerr et al.*, 2012), this resolution is too coarse for most hydrological and agricultural applications. In this context, disaggregation methodologies based on the synergy between microwave and optical data have been proposed in order to derive HR SM products. Disaggregation methodologies distinguish themselves in two categories: purely empirical (polynomial–based) and semi–empirical (based on the triangle/trapezoidal approach).

DISPATCH (*Merlin et al.*, 2008b) is a semi–empirical–based algorithm that converts HR MODIS–derived SEE fields into HR SM fields by expanding a first order Taylor series of a SEE model around the low resolution (LR) SMOS SM. It relies on two different SEE models: a temperature–based model, used to derive HR SEE from MODIS data, and a SM–based model, used to link the temperature–based SEE to SM. More specifically, HR LST and NDVI data are used to derive in a first step the low resolution extreme temperatures (T_{ext}), which are then used to estimated HR SEE. A self–calibrated SM–based SEE model is then used to build a downscaling relationship linking the disaggregated SM, SMOS SM and MODIS-derived SEE.

Since the current version of DISPATCH uses image–based T_{ext}, it can be considered a contextual method. In an effort to remove the limitations that come with the contextual nature of the algorithm, T_{ext} are now estimated by using a soil energy balance model (EBsoil, see Chapter 3) forced by meteorological data in both dry and wet conditions. By introducing EBsoil forced by meteorological data at low resolution in the estimation of T_{ext}, a new version of DISPATCH is developed, hereby addressed as DISPATCH–E. The goals behind the development of DISPATCH–E are to improve both the physics as well as the robustness of the temperature–based SEE model and hence the calibration strategy of the SM-based SEE model. SAFRAN meteorological data are used as input to EBsoil.

In a first step, the description of the DISPATCH downscaling methodology is offered in Section 5.2. A sensitivity analysis of DISPATCH to various factors is first presented in Section 5.3. The same section regroups summarized results that were obtained when applying DISPATCH and DISPATCH–E in Spain. More specifically, the behavior of the temperature–based SEE model as a function of SMOS SM is analyzed. A non–linear SM–based SEE model is then fitted to the temperature–based SEE model and DISPATCH and DISPATCH–E are again applied. Both DISPATCH and DISPATCH–E temperature–based SEE models are compared in time with SMOS SM. As a final step, results in terms of disaggregated SM are compared to *in situ* measurements over the Urgell area, to serve as validation of the methodology. The non–linear SM–based SEE model is thus evaluated by validating the outputs of DISPATCH–E. A summary and general conclusions can be found in Section 5.4. The article (*in preparation*) corresponding to this analysis can be found in Section 5.5, providing additional results in terms of extreme soil temperatures, DISPATCH's sensitivity with respect to atmospheric conditions, a comparison between the behaviors of the non–linear SM–based and the temperature–based SEE models, as well as detailed statistics concerning the SM estimates.

5.2 Methodology

5.2.1 DISPATCH

The approach used for disaggregating SMOS SM is based on the latest operational version (*Molero et al.*, 2016) of the improved DISPATCH algorithm, presented in (*Merlin et al.*, 2013b). DISPATCH relies on two different SEE models and a first order Taylor series expansion around the coarse scale SM observations in order to convert 1 km resolution MODIS–derived soil temperature fields into 1 km resolution SM fields. The input dataset is generated by taking into account different combinations of SMOS SM (ascending 6 AM and descending 6 PM) and MODIS (Terra overpass 10:30 AM and Aqua 1:30 PM from one day before until one day after the SMOS overpass). That being said, an implicit assumption is made, that no rainfall occurs between the SMOS and MODIS overpasses and that the MODIS spatial variability is similar to the actual SM variability at the SMOS overpass time. The DISPATCH product is the average at 1 km resolution of the disaggregated output ensemble.

The downscaling relationship is expressed as:

$$SM_{HR} = SM_{LR} + \frac{\partial SM_{mod}}{\partial SEE}(SEE_{LR}) \cdot (SEE_{HR} - SEE_{LR}) \tag{5.1}$$

Where SM_{HR} represents the 1 km disaggregated SM, SM_{LR} the SMOS observation, $\frac{\partial SM_{mod}}{\partial SEE}(SEE_{LR})$ the partial derivative of SM with respect to SEE, evaluated at SMOS scale, SEE_{HR} the MODIS–derived SEE, and SEE_{LR} its average within the SMOS pixel.

The temperature–based SEE is therefore used to derive SM_{HR} from MODIS data, while the SM–based model (partial derivative of SM with respect to SEE in Equation 5.1) is used to link the temperature-based SEE to SM.

The main hyphothesis behind Equation 5.1 is that the relationship linking the SEE to SM is valid across all scales. This is true if an implicit assumption of uniform meteorological conditions at the SMOS pixel scale is made.

An overview of these models is offered below.

5.2.2 Temperature–based SEE model

The HR SEE is assumed to be constant during daytime, provided that clear sky conditions are met (*Shuttleworth et al.*, 1989; *Nichols and Cuenca*, 1993). It is estimated as:

$$SEE_{HR} = \frac{T_{s,dry} - T_s}{T_{s,dry} - T_{s,wet}} \tag{5.2}$$

With T_s the MODIS-derived soil temperature. A linear relationship between soil temperature and SEE is thus assumed, relationship which was recently verified by (*Merlin et al.*, 2016) by using a physically–based soil energy balance model for different soil types. Note that Equation 5.2 is consistent with the evaporative fraction models of contextual evapotranspiration approaches (see Chapter 3).

The MODIS LST dataset is corrected beforehand for topographic effects (decrease of air temperature with altitude) as follows:

$$T_{MODIS} = T_{MODIS,init} + \gamma(H_{HR} - H_{LR}) \tag{5.3}$$

Where T_{MODIS} is the corrected MODIS LST, γ ($^{\circ}$C$\cdot$m^{-1}) the mean lapse rate (set to 0.006 $^{\circ}$C$\cdot$m^{-1}), $T_{MODIS,init}$ the original MODIS LST, H_{HR} the elevation of the MODIS pixel and H_{LR} the mean elevation within the LR pixel.

The downscaling methodology decouples soil evaporation from the 0 - 5 cm soil layer and the vegetation transpiration from the root zone soil layer by separating MODIS surface temperature into soil and vegetation components. This partitioning approach is based on the "hourglass" method presented in (*Moran et al.*, 1994). Soil temperature is obtained through a linear decomposition of the surface temperature into its two basic components: soil and vegetation. The MODIS–derived soil temperature is given by the following relationship:

$$T_s = \frac{T_{MODIS} - f_{vg}T_v}{1 - f_{vg}} \tag{5.4}$$

Where f_{vg} is the MODIS–derived fractional vegetation cover and T_v the vegetation temperature. With respect to the vegetation temperature, the "hourglass" approach is used to estimate it. By defining the $LST - f_{vg}$ space, up to four different zones can be distinguished by plotting (see Figure 5.1) the diagonals of the polygon (*Moran et al.*, 1994; *Merlin et al.*, 2012a, 2014). Therefore, up to four different formulations of the vegetation temperature can be expressed:

1. Evaporation-controlled zone

$$T_v = \frac{T_{v,wet} + T_{v,dry}}{2} \tag{5.5}$$

   Where $T_{v,wet}$ and $T_{v,wet}$ are the vegetation temperatures corresponding to well watered full cover vegetation and water–stressed senescent vegetation. The points associated to this zone correspond to LSTs controlled essentially by soil evaporation, thus leading to an optimal sensitivity to SM. This zone will be hereby referred to as Zone A.

2. Mixed surface (unstressed) zone

$$T_v = \frac{T_{v,wet} + T_{v,dry,MODIS}}{2} \tag{5.6}$$

   Where $T_{v,dry,MODIS}$ is the vegetation temperature complementary to SEE = 1 ($T_s = T_{s,wet}$). This zone will be hereby referred to as Zone B.

3. Mixed surface (stressed) zone

$$T_v = \frac{T_{v,wet,MODIS} + T_{v,dry}}{2} \tag{5.7}$$

   Where $T_{v,wet,MODIS}$ is the vegetation temperature complementary to SEE = 0 ($T_s = T_{s,dry}$). This zone will be hereby referred to as Zone C.

4. Transpiration controlled zone

$$T_v = \frac{T_{v,wet,MODIS} + T_{v,dry,MODIS}}{2} \tag{5.8}$$

The points associated to this zone correspond to LSTs controlled primarily by vegetation transpiration, leading to no sensitivity to SM. This zone will be hereby referred to as Zone D.

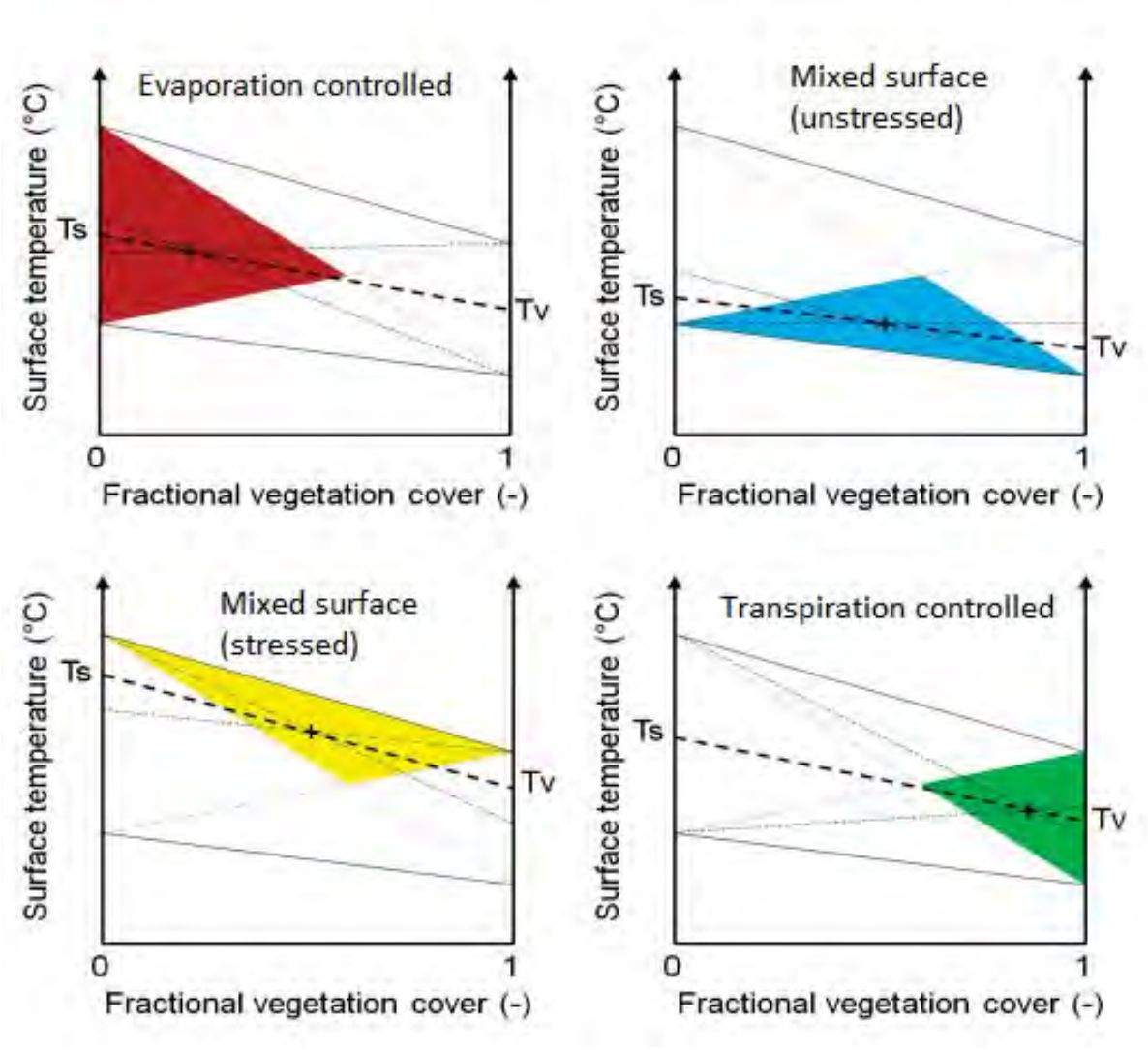


Figure 5.1: The four different zones that can be distinguished in the $LST - f_{vg}$ space (*Merlin et al.*, 2012a).

The extreme temperatures are calculated either directly from remote sensing data (image–based: DISPATCH), or by using a soil energy balance model (model–based: DISPATCH-E). The image–based T_{ext} are estimated from the $LST - f_{vg}$ space, as in the original version of the SEB–1S module, presented in Chapter 3. In an effort to improve the algorithm's robustness, the soil energy balance (EBsoil) model introducted in Chapter 3 was used in order to derive T_{ext} independently of the surface conditions within the study area. The methodology behind the derivation of model–based T_{ext} was presented in Section 3.3.

5.2.3 Soil moisture–based SEE model

Besides the temperature–based SEE model, DISPATCH also uses a SM–based linear model. A non–linear SM–based model is now introduced both in DISPATCH and DISPATCH–E, an its impact on disaggregated soil moisture products is investigated. A description of both the linear and non–linear SM–based SEE models is offered below. Several studies have reported a non–linear behavior of the SEE with respect to SM (*Komatsu*, 2003; *Lee and Pielke*, 1992; *Noilhan and Planton*, 1989). This non–linear behavior is a limitation of the relationship between SM and its disaggregating parameters in the higher ranges that the linear model does not take into account.

5.2.3.1 Linear model

SM is linked to SEE through a semi–empirical linear model (*Budyko*, 1956; *Manabe*, 1969):

$$SEE_{HR} = \frac{SM_{HR}}{SM_p} \tag{5.9}$$

where SM_p is a self–estimated parameter at LR:

$$SM_p = \frac{SM_{LR}}{SEE_{LR}} \tag{5.10}$$

Merlin et al. (2013b) have succesfully implemented this linear model, and its performance has been thoroughly evaluated, by showing that the linearity hypothesis is adequate at kilometric scales, while a potential robustness over non–linear models with errouneous behaviors can be expected. In addition, the SM_p parameter is calibrated from daily observed SEE and SM at low resolution. Because of the linearity in Equation 5.9, there are direct compensations between SEE and SM_p, which are helpful in describing the real behavior of the SEE starting from a range of different SM_p values. Nevertheless, as several studies (*Lee and Pielke*, 1992; *Chanzy and Bruckler*, 1993; *Merlin et al.*, 2011) have proved, SEE has a strongly non–linear behavior over the full SM range. This is a fundamental constraint of the SEE model in Equation 5.1, as the derivative of SEE with respect to SM is influenced by the SM range, and generally decreases with SM (*Komatsu*, 2003).

5.2.3.2 Non–linear model

Multiple non–linear models (*Komatsu*, 2003; *Lee and Pielke*, 1992; *Noilhan and Planton*, 1989) have been implemented in earlier versions of DISPATCH. The model described in (*Komatsu*, 2003) has been proven to have a greater stability (already present in the linear mode), with slope values more stable around 1, than the models in (*Lee and Pielke*, 1992; *Noilhan and Planton*, 1989). The effects that the nonlinearity of the SEE–SM relationship could have on the disaggregated SM products has also been thouroughly addressed, by concluding that a non–linear model is a better choice when using HR ($\sim$ 100 m resolution) data, since it increases the slope of the linear regression between downscaled products and *in situ* measurements. This translates into a better spatial representativeness of SM.

Amongst the three non–linear models, the one presented in (*Komatsu*, 2003) is of special interest. The arguments behind this are linked to the the derivative $\frac{\partial SM_{mod}}{\partial SEE}(SEE_{LR})$:

- it is lowered with low SM_{LR} values

- it is increased with high SM_{LR} values

A small derivative at $SM_{LR} \sim 0$ means that the uniform 1 km SM data is approximately equal to SM_{LR}, which the models in (*Lee and Pielke*, 1992; *Noilhan and Planton*, 1989) do not allow. In particular, when using the latter models, a large derivative is obtained, leading to erroneous large disaggregated values.

This behavior can be observed when looking at Figure 5.2, which is a theoretical representation of the three non–linear models mentioned above (*Komatsu*, 2003; *Noilhan and Planton*, 1989; *Lee and Pielke*, 1992).

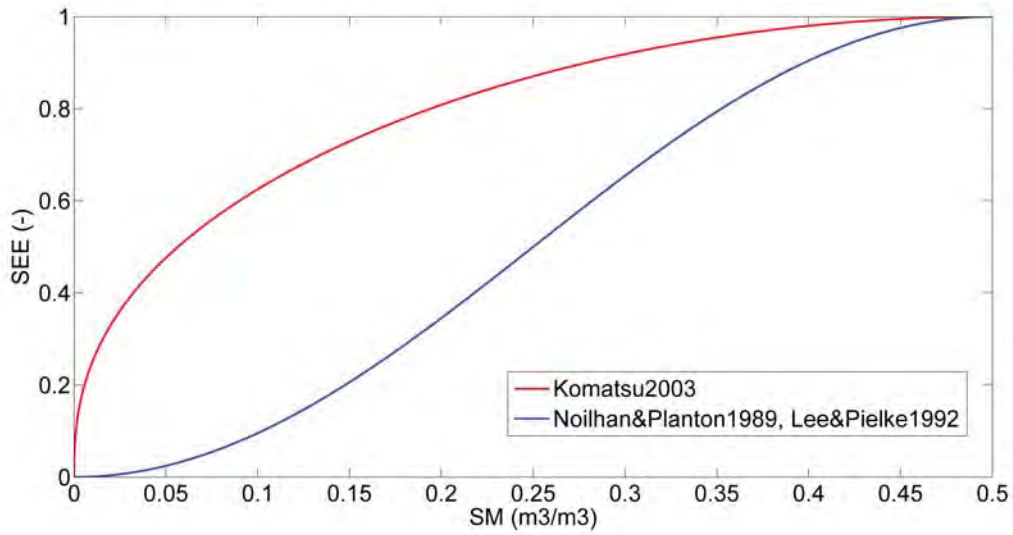


Figure 5.2: Visual representation of three non–linear SM–based SEE models.

Taking everything into account, the exponential form of (*Komatsu*, 2003) (hereby mentioned as K03) is chosen as the non–linear SEE model to be calibrated:

$$SEE_{sim} = 1 - exp(\frac{SM}{SM_c}) \tag{5.11}$$

with SM_c a semi-empirical parameter. In this analysis, it is supposed to be constant in space (within each LR pixel) and it is calibrated at LR for each SMOS pixel from the times series of SMOS, MODIS and SAFRAN data. An iterative loop is run on SM_c (initialized at 0) until the simulated SEE approaches the observed SEE, i.e. until the cost function $||SEE_{sim} - SEE_{obs}||^2$ is minimum. SEE_{obs} is considered as SEE_{LR} (computed as the spatial linear average of SEE_{HR}).

5.2.4 DISPATCH–E

In order to remove some limitations that arise when estimating extreme temperatures from medium and low resolution remote sensing data (and subsequently the limitations it implies on disaggragated SM products), the soil energy balance model (EBsoil) introduced in Chapter 3 is integrated within DISPATCH and a new version of DISPATCH is thus developed: DISPATCH–E. Soil T_{ext} are now estimated independently of MODIS data, by using EBsoil, forced by SAFRAN

meteorological data. SAFRAN data was aggregated from a 5 km original resolution to the SMOS resolution, by using a linear average. An overview of both models (DISPATCH and DISPATCH–E) is presented in Figure 5.3.

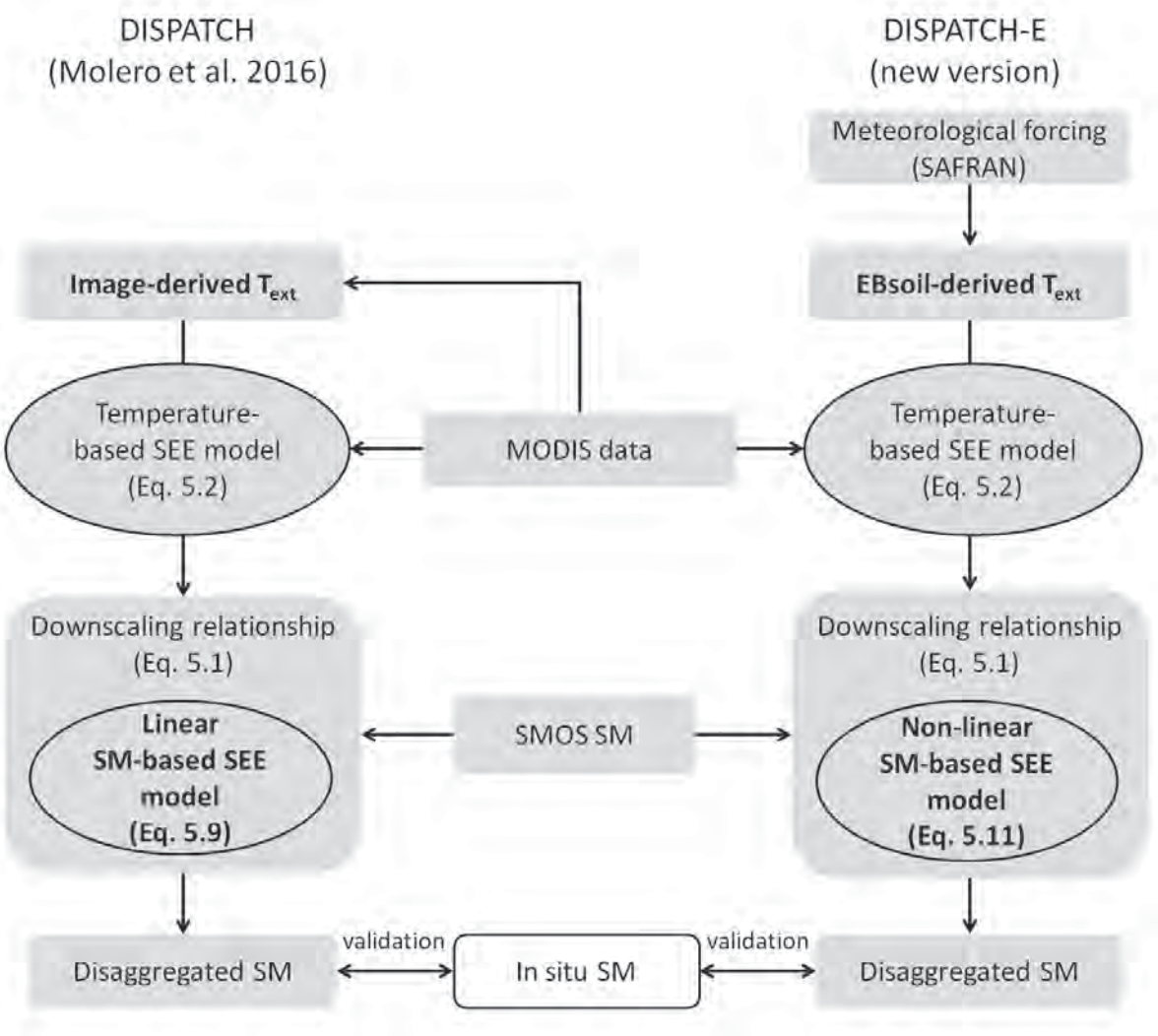


Figure 5.3: Schematic overview of DISPATCH (left) and DISPATCH–E (right).

5.3 Results

A sensitivity analysis of DISPATCH is first presented. Then, the performance of DISPATCH–E (DISPATCH coupled with EBsoil) is evaluated over the study area in terms of: SEE behavior with respect to SMOS SM, and disaggregated SM values. The results of the calibrated non-linear model are first presented: the temperature–based SEE estimates obtained from either the MODIS– or EBsoil–derived extreme temperatures are compared to the SM–based SEE estimates. Then, the SMOS SM disaggregated by DISPATCH and DISPATCH–E in both linear and non–linear modes is evaluated against the *in situ* measurements collected in 2011 and 2012.

5.3.1 DISPATCH/DISPATCH–E sensitivity analysis

An analysis of the sensitivity of DISPATCH and DISPATCH–E with respect to several parameters is shown in this section. In particular, their sensitivity with respect to two input parameters is studied:

- the number of MODIS days chosen for the input LST dataset: 1 day (just two datatsets corresponding to Aqua and Terra, on the same day as the SMOS overpass) or three days (six datasets corresponding to Aqua and Terra from one day before until one day after the SMOS overpass). Taking into account MODIS data from three consecutive days provides better temporal coverage. However, the uncertainty in the methodology is lower when using MODIS data corresponding to 1 day. This can be explained because the 3 day derived product has an underlying assumption, that SM fields are temporally stable for periods of at least one day around the SMOS overpass, which can sometimes be violated if precipitation or irrigation take place.

- the CNT datatset (see Chapter 2, Section 2.3.5 for more details) that is taken into account when generating the aggregated SM product. Different values are taken into consideration: 4, 8, 12. It represents the number of SM–LST combinations used by DISPATCH to produce one output. Low values are associated with missing input data (due to SMOS RFI - Radio Frequency Interference - contamination, cloud presence etc.). Using a low CNT value means that the downscaled SM fields do not profit from the reduction in independent random errors as result of averaging (*Molero et al.*, 2016). However, the total number of days taken into consideration when performing a temporal analysis drops with a high CNT value.

Moreover, the influence of the partitioning method of Equation 5.4 is also evaluated, by selecting LST points that correspond to different areas within the $LST - f_{vg}$ space (see Figure 5.1). In practice, the behavior of the temperature–based SEE estimates obtained using either MODIS– or EBsoil–derived extreme temperatures is analyzed as a function of the SMOS SM. Results are analyzed per individual SMOS pixel. As the behavior of SEE is heavily influenced by the evaporative demand, which varies throughout the day (with a maximum value registered at noon associated with higher incoming solar radiation), results are separated per MODIS overpass.

Figure 5.4 is a visual representation of the study area on which the downscaling algorithm is applied. It engulfs 30 re–sampled 40 km by 40 km SMOS pixels. Six SMOS pixels in particular (hereby numbered as 11, 12, 13, 21, 22 and 23) were chosen for analyses, which surround an area of 20 km by 20 km, where the *in situ measurements* were performed. The 1 km resolution Terra 16-day NDVI on doy 161 in 2011 is also shown, where we can clearly distinguish the irrigated and dryland areas.

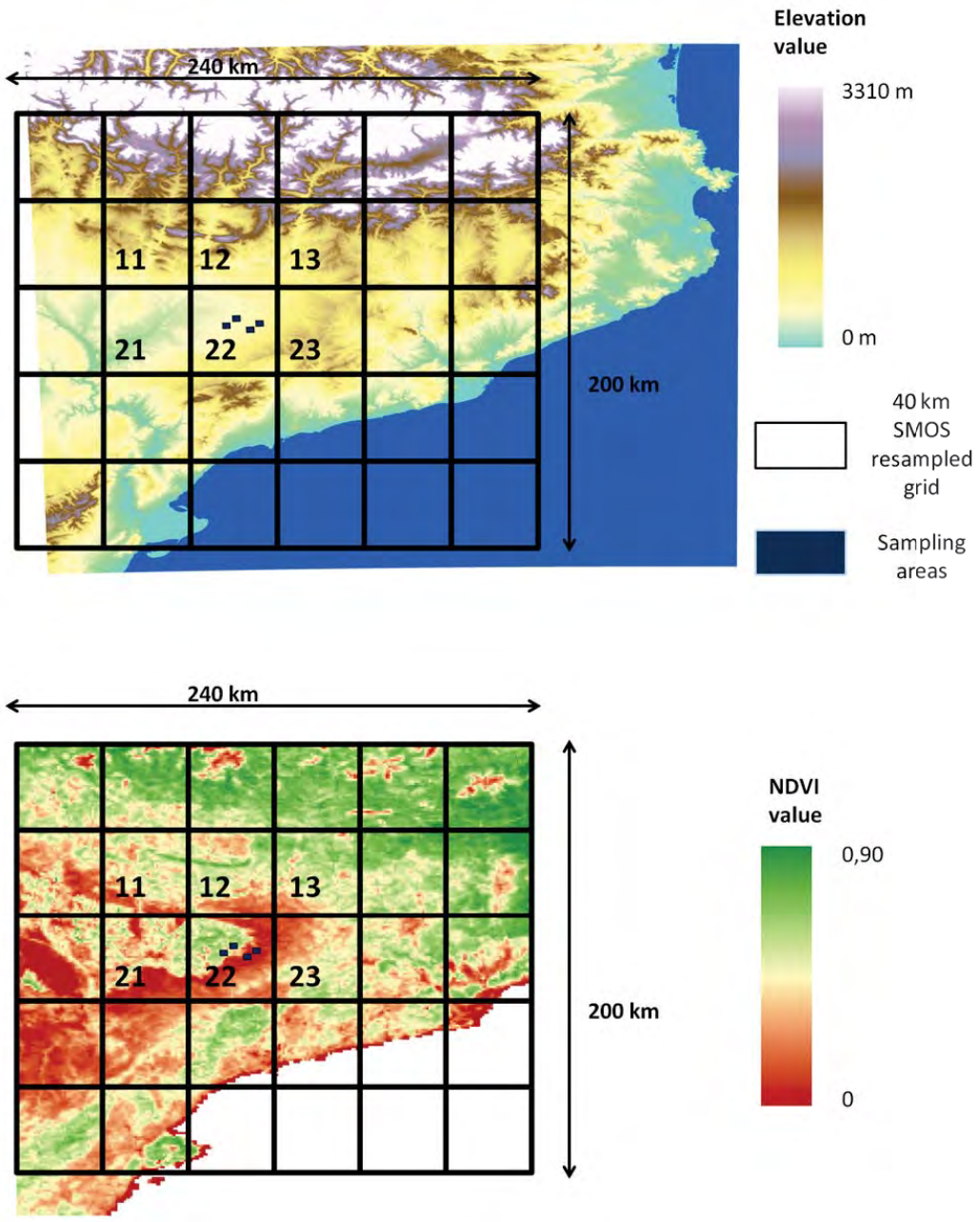


Figure 5.4: The Urgell study area on which both DISPATCH and DISPATCH–E were applied.

5.3.1.1 Influence of soil/vegetation partitioning

As a first analysis, the influence of the partitioning method described at Section 5.2.2 is analyzed with respect to the SEE behavior. In practice, SEE is computed by taking into account points corresponding to different areas of the $LST - f_{vg}$ polygon, as follows:

- zones ABCD - all points within the $LST - f_{vg}$ space are considered.
- zones ABC - points corresponding to just three zones are considered, by excluding points corresponding to zone D.
- zone A - points associated to just one zone are taken into account.

A hypothesis is made, that since LST points corresponding to zone A are controlled essentially by soil evaporation, they have an optimal sensitivity with respect to SM, and thus the corresponding SEE estimates should best describe the shape of the theoretical SEE with respect to SM (values ranging from 0 to 1, with an inflexion point around 0.5 (*Merlin et al.*, 2011)).

In order to have a more representative behavior of the SEE over the entire time series, the analysis takes into account SEE values averaged per 0.01 SMOS SM bins, with at least 3 values per bin. In a first step, the EBsoil model used to derive T_{ext} uses both the Monin-Obukhov (MO) and Richardson (RI) formulations of aerodynamical resistance. Results are then analyzed and a choice is made with respect to the best r_{ah} formulation to be used for future analyses.

Figures 5.5, 5.6 and 5.7 plot the temperature–based SEE against the SMOS SM, for the entire times series, for 6 pixels surrounding the study area, for the Aqua overpass, and for the three different $LST - f_{vg}$ zones respectively. Figures 5.8, 5.9 and 5.10 plot the same thing, but for the Terra overpass. The number of MODIS days taken into account when running DISPATCH is 1 day, with a CNT value equal to 4.

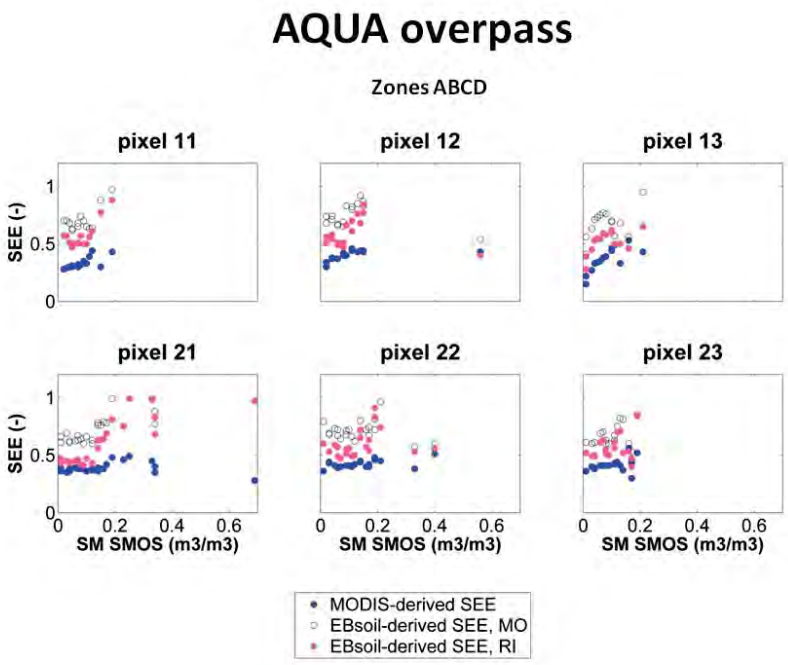


Figure 5.5: The temperature–based SEE is plotted against the SMOS SM, for the entire time series, for the Aqua overpass, for the ABCD zones defined in the $LST - f_{vg}$ space. Temperature–based SEE is derived either from MODIS, or from EBsoil, with its two aerodynamic resistance formulations: MO and RI. The number of MODIS days taken into account is equal to 1.

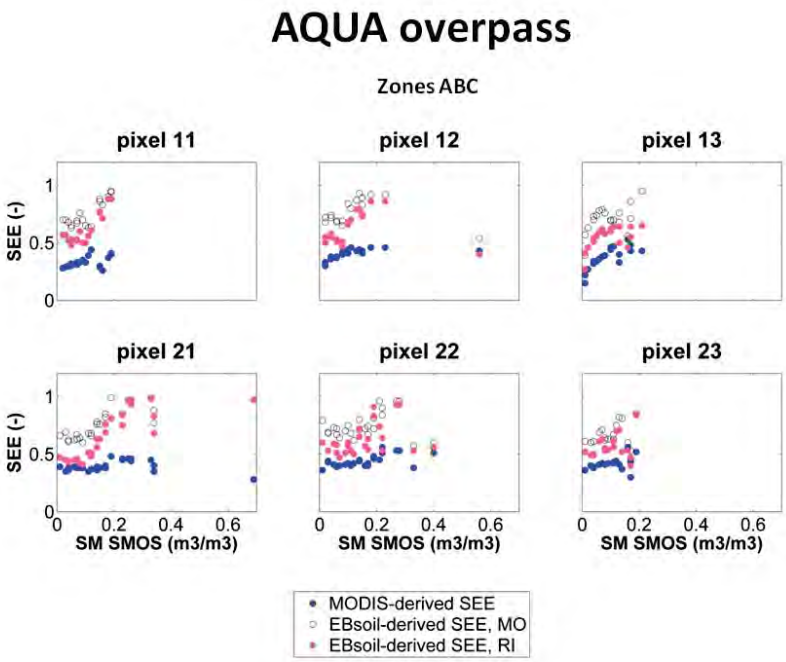


Figure 5.6: Same as Figure 5.5, but for the ABC zones.

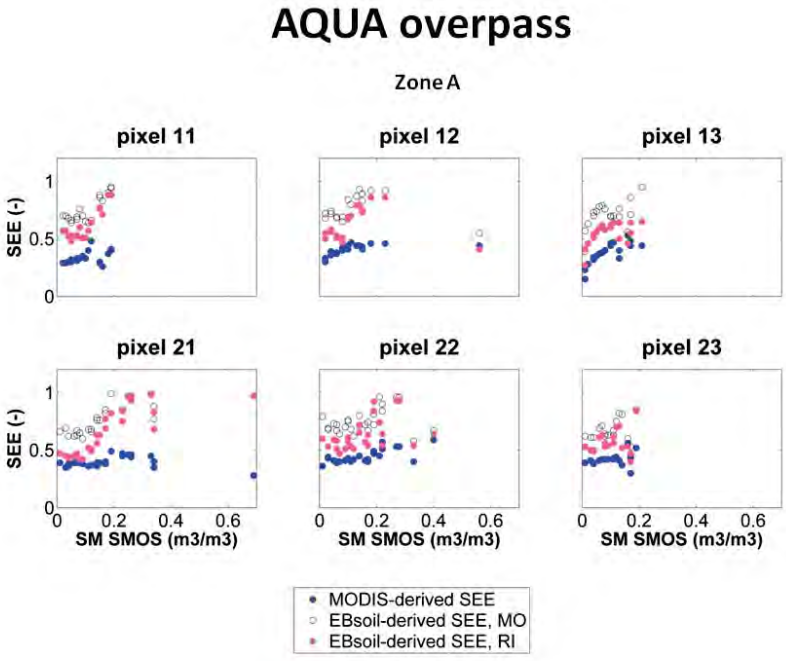


Figure 5.7: Same as Figure 5.5, but for zone A.

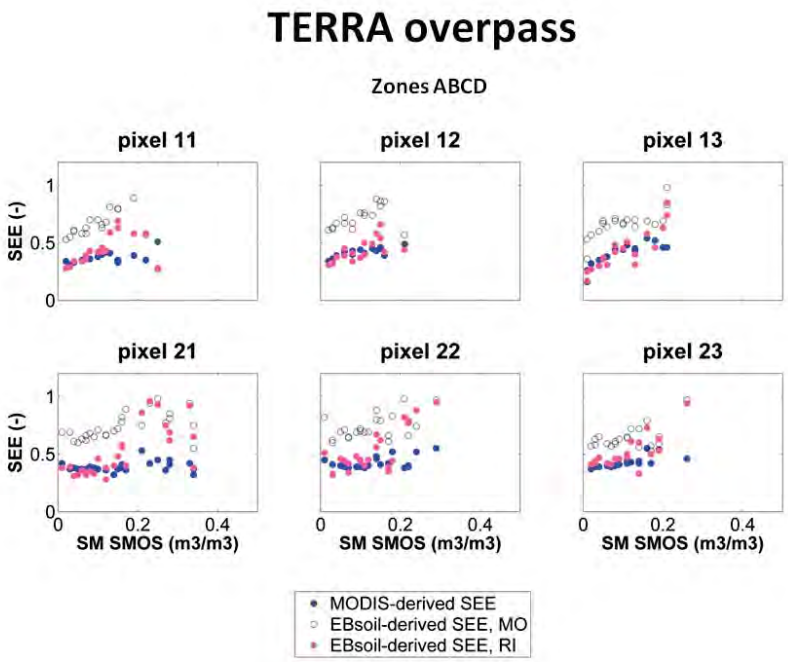


Figure 5.8: Same as Figure 5.5, but for the Terra overpass.

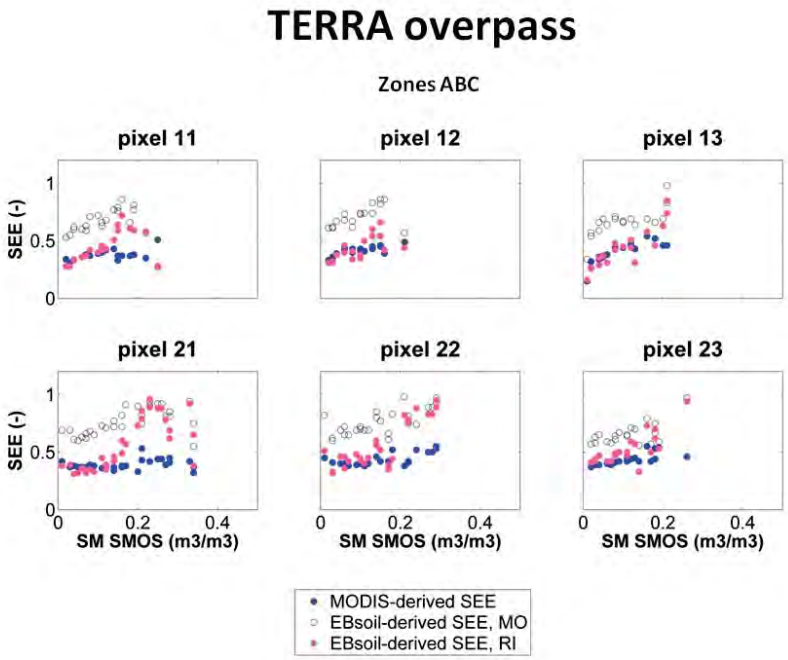


Figure 5.9: Same as Figure 5.6, but for the Terra overpass.

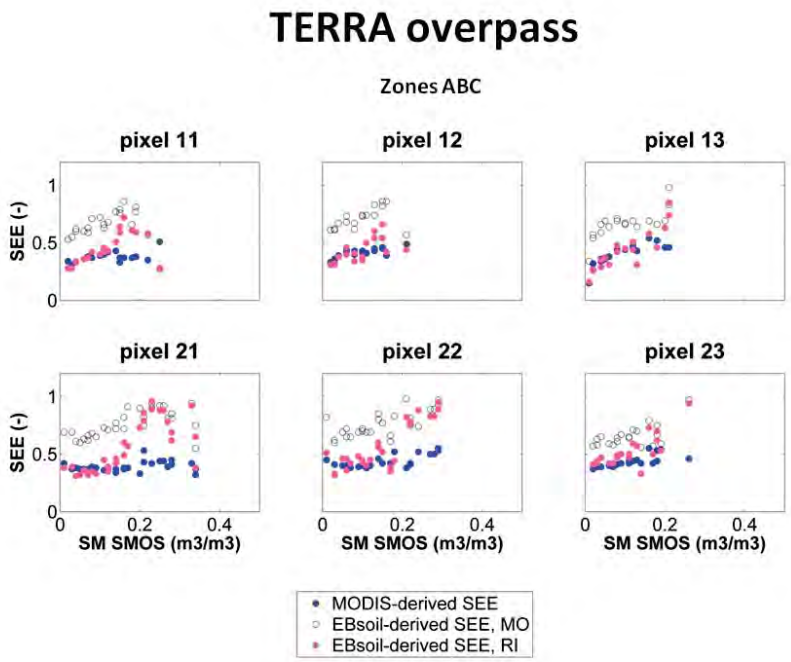


Figure 5.10: Same as Figure 5.7, but for the Terra overpass.

When individually analyzing the figures per MODIS overpass, one can notice that SEE behaves slightly different with respect to each zone. However, no significant contrast can be seen between simulations for zone A and for the ABC zones, the sole difference being just some values of the SEE estimates being slightly lower for the ABC zones. When looking at simulations for the ABCD zones, one can notice that the number of SEE estimates is decreased for certain pixels, even though the general shape of SEE with respect to SM is preserved. The same conclusion can be drawn when analyzing results for the Terra overpass.

When comparing the results obtained for the same configuration (same zone) but for different MODIS overpasses, one can observe a noticeable difference in the shape of the SEE. This can be explained by taking into account the dependence of SEE with respect to evaporative demand. As the evaporative demand changes throughout the day (thus having different evaporative demands for the two MODIS overpasses), the shape of SEE also changes accordingly.

Hence, the hypothesis made in the begining was proven to be false, as there is no significant difference between zone A, zones ABC and zones ABCD. Possible compensation effects might occur, and thus the influence of the distribution of LST data points within the $LST - f_{vg}$ space follows roughly the same pattern. However, in order to have a compromise between the LST dataset used for deriving T_{ext} and a good representativity of the SEE with respect to SM, points associated to zone D (controlled primarily by vegetation transpiration, leading to no sensitivity to SM) are removed from the analysis. Hence, points only within the ABC zones will be used in the next analyses.

When looking at the SEE estimates obtained using the image–based T_{ext} *versus* the model–based T_{ext}, differences can be seen with respect to the range of values. The range of values is increased for the EBsoil–derived SEE, while the MODIS–derived SEE exhibits a rather flat behavior. This relatively flat behavior kept by the MODIS–derived SEE in the higher range of SMOS SM could be explained by the poorer estimates of the MODIS–derived $T_{s,dry}$ and $T_{s,wet}$,

which are not reflective of the true extreme temperatures if the true wet and dry conditions within the scene are not met at the 1 km resolution. At a 1 km resolution in fully dry conditions, MODIS–derived $T_{s,wet}$ tends to be overestimated, which results in a narrower polygon and corresponding SEE values ~ 0.5. In a similar manner, at a 1 km resolution in fully wet conditions, MODIS–derived $T_{s,dry}$ tends to be underestimated, which leads to SEE values ~ 0.5. Therefore, using MODIS data, which implies that extreme conditions are not well met at the 1 km resolution, the model gives estimates of SEE around 0.5, resulting in its relatively flat behavior.

Hence, when deriving SEE using poorer estimates of T_{ext}, the range of values tends to be smaller. An important aspect to mention is that the EBsoil–derived SEE computed from T_{ext} that use the $r_{ah,MO}$ formulation do not capture the same range as the EBsoil–derived SEE computed from T_{ext} that use the $r_{ah,RI}$ formulation. They also tend to produce very high estimates of the SEE, not going below ~ 0.5.

A different behavior is observed for each pixel, possibly due to different soil properties, vegetation covers, land use (irrigated/dry), and topography effects.

5.3.1.2 Influence of number of MODIS days and CNT

In an effort to increase the number of input datasets associated with a better temporal coverage, DISPATCH and DISPATCH-E were run by taking into consideration 3 MODIS days. The impact of the CNT value is then assessed, by taking into consideration three different values in the aggregation of the HR SM products: 4, 8 and 12.

Figures 5.11 , 5.12 and 5.13 plot the temperature–based SEE against the SMOS SM, for the entire times series, for the same 6 pixels, for the Aqua overpass and for three different CNT values: 12, 8 and 4. The temperature–based SEE is estimated either from image–based T_{ext}, or from EBsoil–derived $Text$, with its two aerodynamic resistance formulations. Figures 5.14, 5.15 and 5.16 plot the same, but for the Terra overpass.

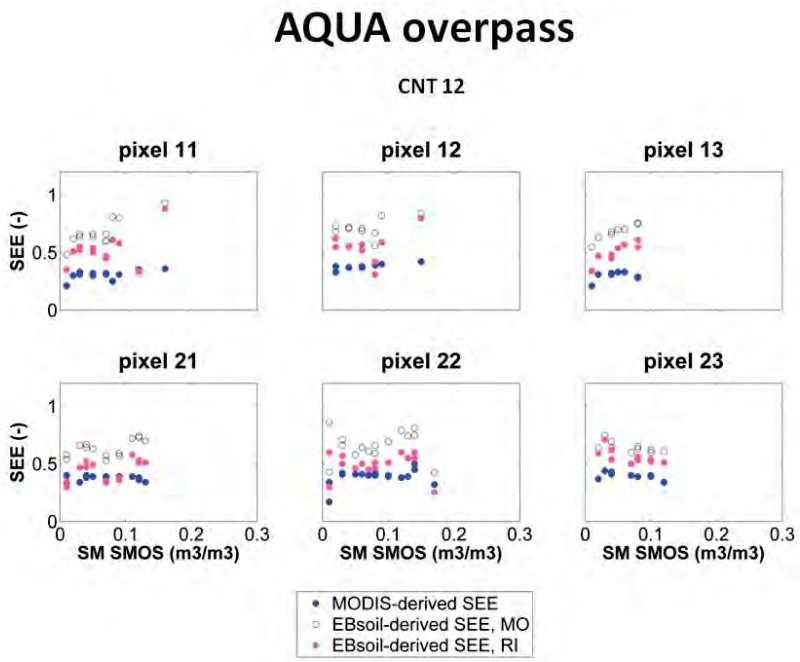


Figure 5.11: The temperature–based SEE is plotted against the SMOS SM, for the entire time series, for the Aqua overpass, for the ABC zones defined in the $LST - f_{vg}$ space. Temperature–based SEE is derived either from MODIS, or from EBsoil, with its two aerodynamic resistance formulations: MO and RI. The CNT value considered is 12.

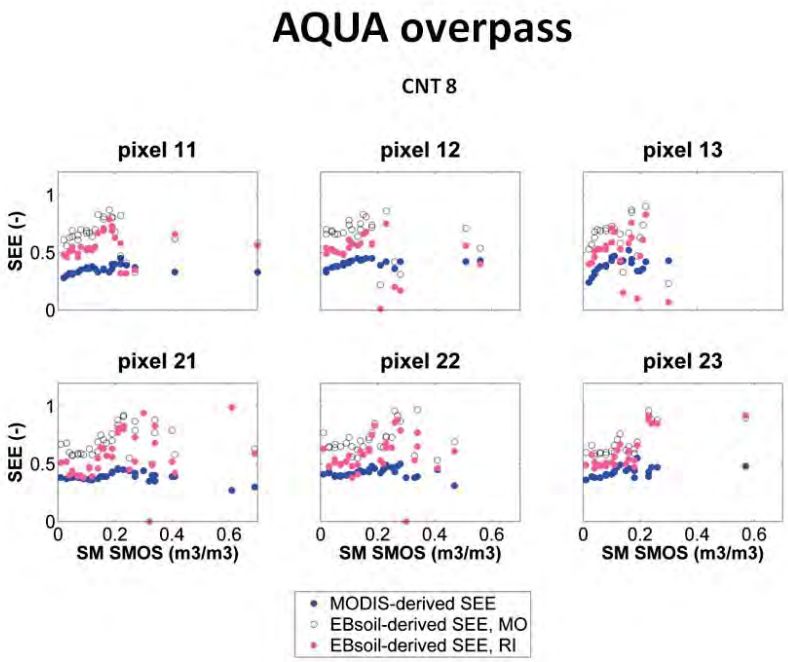


Figure 5.12: Same as Figure 5.11, but for a CNT value equal to 8.

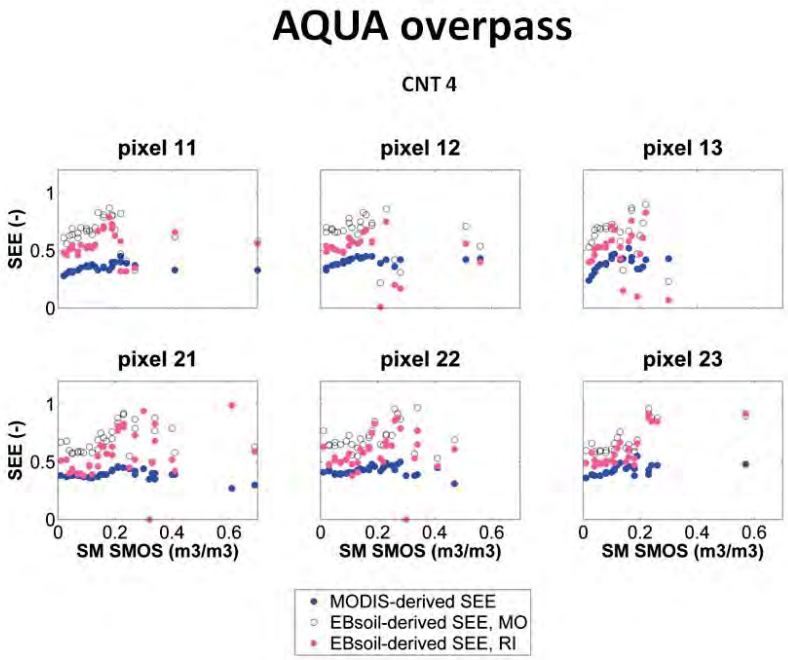


Figure 5.13: Same as Figure 5.11, but for a CNT value equal to 4.

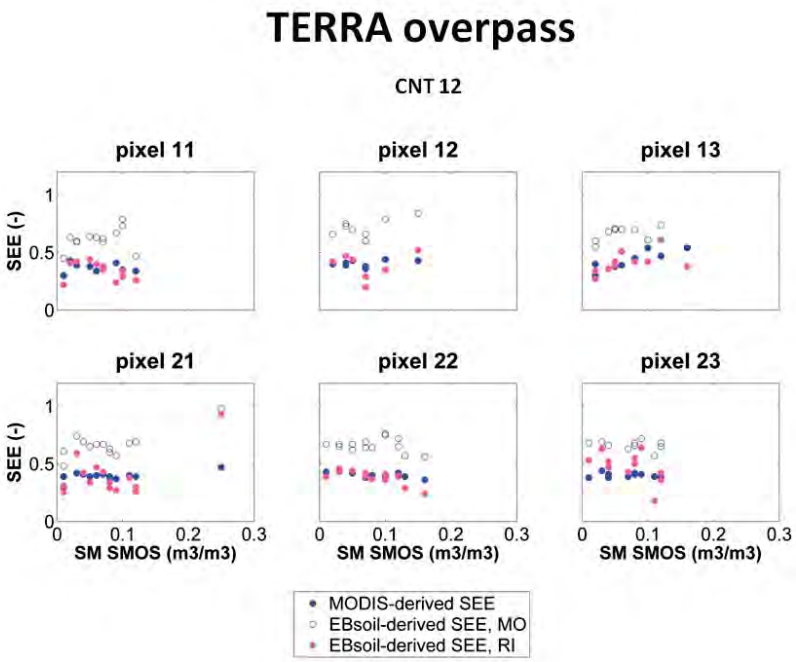


Figure 5.14: Same as Figure 5.11, but for the Terra overpass.

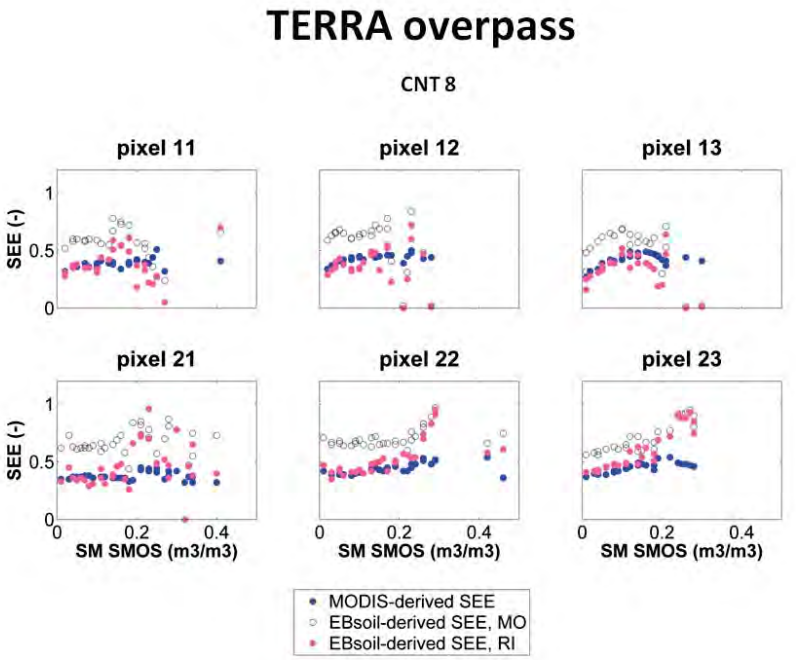


Figure 5.15: Same as Figure 5.12, but for the Terra overpass.

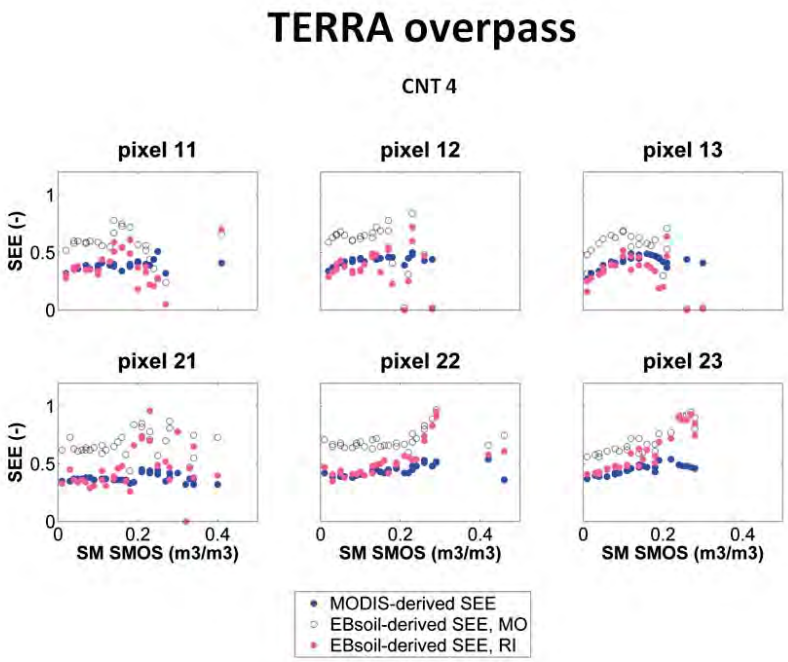


Figure 5.16: Same as Figure 5.13, but for the Terra overpass.

When taking into consideration a lower value for CNT, and thus a higher number of input datasets, the number of datapoints in the SEE estimates is increased, while the shape of the SEE tends to follow the same pattern. It is to be expected, since with an increasing CNT, less available data to generate outputs is present (due to cloud presence, results won't be generated on a daily basis). The same conclusions with respect to the different shape per pixel, per MODIS overpass and per image–based/model–derived T_{ext} are kept. A compromise between a good temporal coverage and a higher uncertainty associated with the temporal stability assumption (that can be infringed by precipitation and irrigation events), and the configuration of MODIS 3 days with a CNT value equal to 4 is chosen for the next simulations. This configuration was also chosen for the simulations performed in Section 4 of the article.

5.3.2 Temperature–based *versus* soil moisture–based SEE estimates

The observed (temperature–based) SEE and the simulated (SM–based) SEE against the SMOS SM are shown in Figures 5.17 and 5.18, for the entire time series, for the same six pixels, and for the Aqua and Terra overpasses separately. The SM–based model used to estimate SEE is the non–linear one. The standard deviation within each bin is also computed. The temperature-based SEE is used to fit the SM–based SEE, by calibrating the SM_c parameter for each SMOS pixel (11, 12, 13, 21, 22 and 23) covering the study area, as previously explained in Section 5.2.3.2. The extreme temperatures used in deriving the temperature–based SEE are derived from either MODIS images or EBsoil. The values of the calibrated SM_c parameter are 0.22 $m^3 \cdot m^{-3}$ (for Aqua) and 0.24 $m^3 \cdot m^{-3}$ (for Terra) when taking into account the MODIS–derived SEE as a reference for the fitting. When taking into account the EBsoil–derived SEE as a reference, the reported SM_c values are 0.14 $m^3 \cdot m^{-3}$ (for Aqua) and 0.20 $m^3 \cdot m^{-3}$ (for Terra). These calibration

values are estimated over pixel 22 (where *in situ* SM measurements are available) and used as a proxy for other pixels when deriving SM_{HR}.

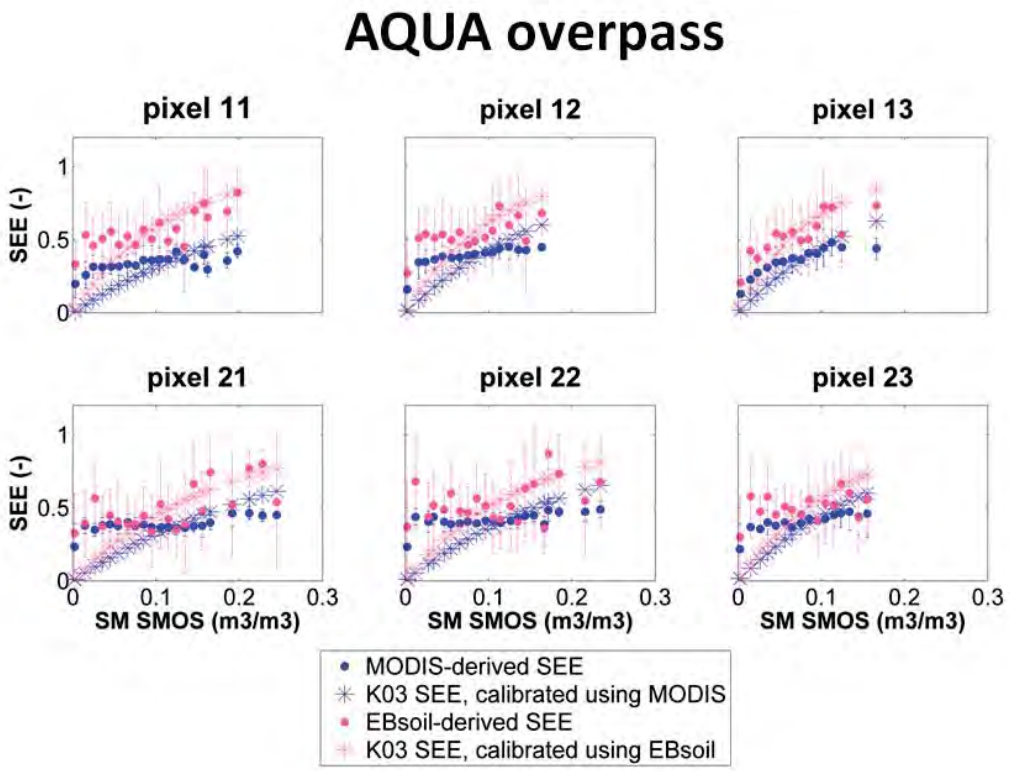


Figure 5.17: The observed (temperature–based) SEE and the simulated (SM–based) SEE is plotted against the SMOS SM, for the entire time series, for the Aqua overpass. The observed SEE is derived either from MODIS, or from EBsoil. The SM–based SEE uses the non–linear K03 model and is calibrated from the observed SEE.

With respect to the different behavior of the SEE as a function of SM, observed for different pixels, extensive studies have been made which proved that the shape of SEE depends on soil properties whose common proxy is given by sand and clay percentages (*Merlin et al.*, 2016). The pixels that have a high coverage consisting of crops are pixels 21 and 22, while the rest of the pixels share a certain percentage between crops and grass/shrubs, with small scattered forests. Irrigation is present for some of crops, which leads to poorer estimates of extreme temperatures in fully dry conditions, and therefore the retrieved SEE. Pixels containing intense canopies present poorer image–based estimates than pixels containing only grass or shrubs. Since mountainous areas are also close to our study site (more specifically in the vecinity of pixels 11, 12 and 13), topography effects could induce errors in the estimates. The nominal behavior of the SEE with respect to SM should be observed over flat irrigated pixels, where the dominant land cover is grass/shrubs (notably pixels 21 and 22). However, the best behavior is observed over pixel 13, which is affected by topography and is half covered by forests. Topographic and cloud effects tend to introduce a positive bias in SEE (attributed to los MODIS T_s), which translates into the SEE not having values below 0.15-0.2.

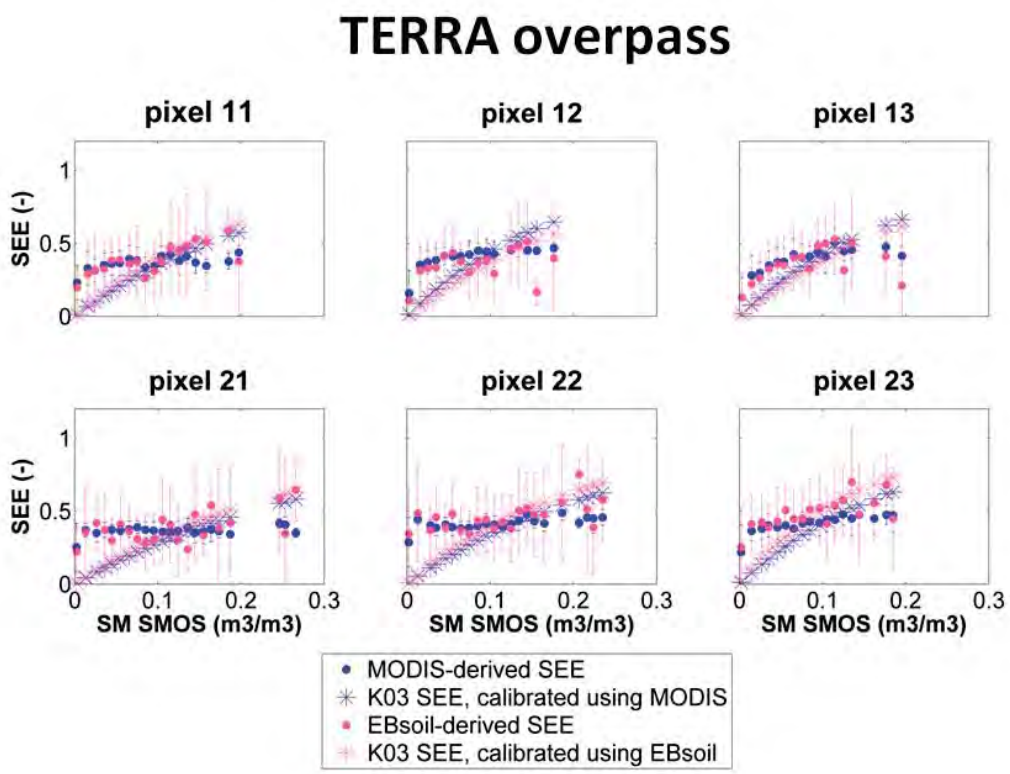


Figure 5.18: Same as Figure 5.17 but for the Terra overpass.

5.3.3 Validation of disaggregation methodology

The performance of DISPATCH–E in terms of disaggregated SM values is evaluated in this section. Figures 5.19 and 5.20 plot the SMOS SM and the DISPATCH–E outputs with respect to *in situ* measurements, for Aqua and Terra overpass times respectively. The *in situ* point measurements were aggregated at a 3 km scale by simple linear averaging. A combination between either the MODIS–derived extreme temperatures and a linear/non–linear SEE(SM) model, or the EBsoil–derived extreme temperatures and a linear/non–linear SEE(SM) model in the downscaling relationship is used in order to derive DISPATCH/DISPATCH–E outputs. As a reminder, the "classical" DISPATCH case corresponds to MODIS–derived SEE, while DISPATCH–E corresponds to EBsoil–derived SEE. Both are tested in the linear and non-linear case scenarios.

Statistical results in terms of correlation coefficient, slope of linear regression, bias and RMSD are reported in Table 5.1.

When looking at the LR SM and the HR SM compared with the 3 km *in situ* data, one observes that DISPATCH improves the fine–scale precision. When using MODIS–/EBsoil–derived SEE in linear mode for the Aqua overpass, the reported correlation coefficients between HR SM and *in situ* data are equal to 0.30 and 0.25 respectively, as opposed to -0.16 when comparing SMOS SM with *in situ* data. Better values of these metrics are obtained for the Terra overpass, with a correlation coefficient equal to 0.69 and 0.70, as opposed to the 0.47 correlation value calculated for SMOS SM. Improvements are also made with respect to the slope of the linear regression, whose value is increased from -0.063 to 0.16 and 0.10 (Aqua overpass, linear mode)

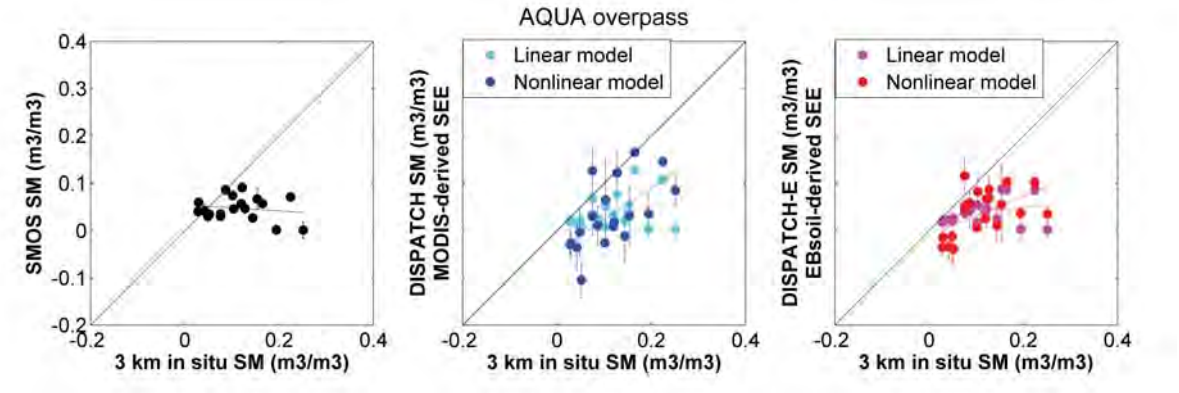


Figure 5.19: The SMOS, DISPATCH and DISPATCH–E SM are plotted against 3 km aggregated *in situ* measurements, for the Aqua overpass.

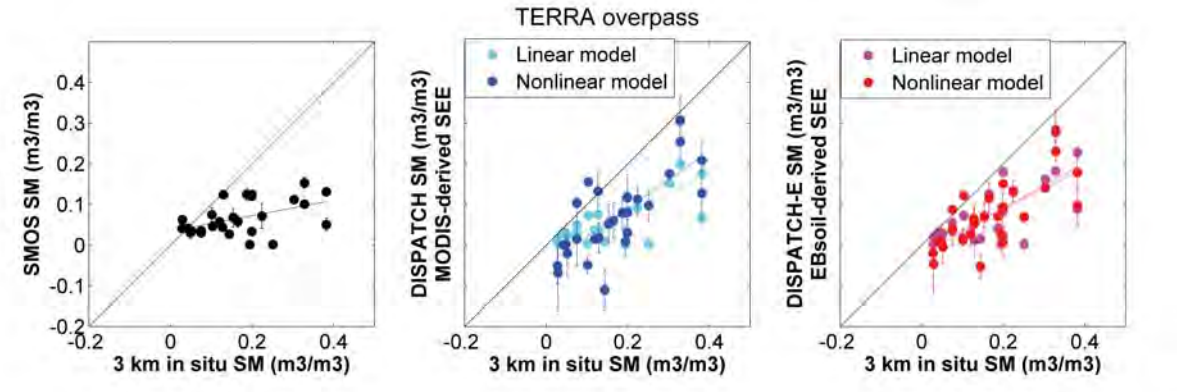


Figure 5.20: Same as Figure 5.19, but for the Terra overpass.

Table 5.1: Correlation coefficient, slope of the linear regression, bias and RMSD between SMOS SM/DISPATCH SM/DISPATCH-E SM and 3 km *in situ* data are reported for each MODIS overpass individually, as well as both overpasses combined. Both DISPATCH and DISPATCH-E are tested in linear and nonlinear modes.

Pixel ID	**AQUA overpass**					**TERRA overpass**				
	SMOS	**MODIS-derived**		**EBsoil-derived**		**SMOS**	**MODIS-derived**		**EBsoil-derived**	
		Linear	**Non–linear**	**Linear**	**Non–linear**		**Linear**	**Non–linear**	**Linear**	**Non–linear**
R	−0.16	0.30	0.61	0.25	0.53	0.47	0.69	0.75	0.70	0.74
S	−0.063	0.16	0.66	0.10	0.41	0.19	0.45	0.70	0.50	0.56
Bias	−0.065	−0.074	−0.081	−0.072	−0.075	−0.10	−0.10	−0.10	−0.095	−0.096
RMSD	0.096	0.096	0.099	0.095	0.093	0.14	0.13	0.12	0.12	0.12

and from 0.19 to 0.45 and 0.50 (Terra overpass, linear mode). When looking at the performance of the two temperature–based SEE models in linear mode, a similar performance in terms of correlation and slope of linear regression, for both MODIS overpasses can be observed. With respect to the bias, the negative bias present in the SMOS SM is also found in all disaggregated SM products. The RMSD values seem to be consistent, independently of the SM resolution, with lower values ($\sim$0.09 $m^3 \cdot m^{-3}$) reported for the Aqua overpass, as opposed to the Terra overpass ($\sim$0.12/0.13 $m^3 \cdot m^{-3}$).

When comparing the linear and the non–linear modes, the non–linear SEE model significantly enhances the correlation coefficient and the slope. The correlation coefficient is significantly increased for the Aqua overpass: from 0.30 to 0.61 (MODIS–derived SEE) and from 0.25 to 0.53 (EBsoil–derived SEE). The values reported for the slope of the linear regression are also greatly increased, from 0.16 to 0.66 (MODIS–derived SEE) and from 0.10 to 0.41 (EBsoil–derived SEE) – for the Aqua overpass over a smaller SM range.

The non–linear mode leads to better results than the linear mode. Since in the non–linear mode the partial derivative of SM with respect to SEE is decreased in the lower SM ranges and increased in the higher SM ranges, this leads to an overall better precision and accuracy of the corresponding disaggregated products when compared to the 3 km aggregated *in situ* measurements.

An important aspect to mention is the negative values of disaggregated SM when using the non–linear SEE(SM) model in the downscaling relationship. If we assume that the disaggregation is efficient, then this could point out an underestimation of SMOS SM in very dry areas (with SM close to zero). A negative bias in SMOS SM has been reported by various calibration and validation studies (*Collow et al.*, 2012; *Zeng et al.*, 2015; *Al Bitar et al.*, 2012; *Djamai et al.*, 2016). Biases in the brightness temperatures lead to biases in the retrieved SM. One possible cause for the biases in brightness temperatures can be the RFI. According to *Dente et al.* (2012), a positive bias in the observed brightness temperature would introduce a negative bias in the SM products. Since SMOS SM is used when calibrating the SM_c parameter, the parameter retrieval is also affected by the negative bias. The downscaled SM data obtained in the non–linear mode is thus affected by both potential bias in retrieved SM_c values, as well as the negative bias in SMOS data. The negative bias can also be partly induced by the difference in the sampling depth between the SMOS measurements and the *in situ* measurements. The negative bias in the SMOS SM is however found for large SM values; in the lower ranges of the SM, SMOS SM is coherent with *in situ* measurements.

5.4 Summary and Conclusions

This part focuses on characterizing both the temporal and spatial behaviors of the soil evaporative efficency (SEE), by combining both a LST– and a SM–based model of SEE. The technique is validated with respect to SM estimates. DISPATCH is a downscaling algorithm that combines the two LST– and SM–based SEE models in order to estimate 1 km resolution data from 40 km resolution SMOS and 1 km resolution MODIS data. One model relies on a temperature–based expression of the SEE in order to determine SEE estimates from MODIS images solely, while the other model expresses SEE as a linear function of the SM. A new version (DISPATCH–E) is developed, by making modifications to both the temperature– and SM–based models. A soil energy balance (EBsoil) model is now used as an alternative temperature–based approach to derive SEE. The goal behind using EBsoil is to enhance the physics and robustness of the temperature–based SEE model and thus the calibration strategy of the SM–based SEE model.

The methodology is tested by comparing the temperature–based and SM–based SEE estimates, and disaggregated SM values derived by both DISPATCH and DISPATCH–E with *in situ* measurements over a Spanish site. The behavior of the temperature–based SEE with respect to SMOS SM is assessed independently for each SMOS pixel covering the study area. A compromise between precision and robustness is obtained: the SEE estimates based on the MODIS–derived extreme soil temperatures are more precise, while the EBsoil model provides more accurate SEE estimates. With respect to SM estimates, SMOS SM and 1 km resolution disaggregated SM are both compared to 3 km resolution *in situ* measurements. Statistics are greatly improved for the downscaled SM products. Moreover, by integrating a non–linear SM–based SEE model, further improvements are obtained for the downscaled SM products.

The potential of DISPATCH–E is strong when applied to areas representative of relatively dry or relatively wet conditions, i.e. both wet and dry conditions are not met at the same time: desert areas, temperate areas or rainfed areas. By taking into account the two separate temperature–based models in future analyses, a better representation of SEE could potentially be accomplished. Moreover, the SEE modeling based on the non–linear SM model could be integrated into existing evapotranspiration models, which are based on a combination of thermal and microwave data. From a downscaling point of view, disaggregation is a way of detecting a bias in the low resolution SM data, which cannot be easily detected without taking into account the spatial heterogeneity within the scene. Future studies could try and correct the bias reported by literature in SMOS data and recalibrate the SM–based SEE model taking into consideration new corrected data. Any improvement in the SEE model could lead to refinements of the SEE estimates and of the associated HR SM products.

5.5 Article

Improving the physics and robustness of an evaporation-based disaggregation method of SMOS soil moisture data

V. Stefan[1,*], O. Merlin[1, 2], B. Molero[1], B. Aït-Hssaine[2], M.-J. Escorihuela[3], P. Quintana-Seguí[4], S. Er-Raki[2]

1 CESBIO, Université de Toulouse, CNES/CNRS/IRD/UPS, Toulouse, France

2 Université Cadi Ayyad, Marrakech, Morocco

3 isardSAT, Barcelona, Spain

4 Ebro Observatory, Catalonia, Spain

* vivien.stefan@cesbio.cnes.fr; tel.: +33-5-6155-8829

Abstract: Surface soil moisture (SM) is one of the main factors controlling soil evaporation under non-energy limited conditions. This has been the basis for developing DISPATCH (Disaggregation based on Physical and Theoretical scale Change). DISPATCH attempts to improve the resolution of SMOS (Soil Moisture and Ocean Salinity) soil moisture data using the soil evaporative efficiency (SEE, defined as the ratio of actual to potential soil evaporation) derived from MODIS (MODerate resolution Imaging Spectroradiometer) data. Specifically DISPATCH relies on two separate SEE models: a temperature-based SEE model driven by MODIS data and a SM-based SEE model driven by SMOS data. The objective of this paper is to improve the physics and calibration of both SEE models by integrating within DISPATCH a soil energy balance model forced by available meteorological data. The approach is tested using SAFRAN (Système d'analyse fournissant des renseignements atmosphériques à la neige) meteorological data and *in situ* SM measurements for validation over a mixed dry and irrigated area in Catalunya, Spain, in 2011 and 2012. As a first

step, the temperature-based SEE model is temporally compared to SMOS SM for each SMOS pixel. The temperature-based SEE model generally has a better slope of the linear regression but worse root mean square difference (RMSD) and correlation coefficient when meteorological data are used as input. As a second step, the DISPATCH algorithm is run using the SM-based SEE model fitted to the temperature-based SEE model. The mean correlation and slope of the linear regression between disaggregated and *in situ* SM is increased from 0.65 and 0.41 to 0.72 and 0.54 when integrating meteorological data within DISPATCH, respectively, all the while keeping stable values of the RMSD. Such an algorithm (DISPATCH-E) is expected to significantly improve the robustness of the evaporation-based disaggregation of SMOS data in a range of SM conditions.

Keywords: downscaling, disaggregation, soil moisture, SMOS, MODIS, evaporation

1. Introduction

Soil moisture (SM) is one of the main factors controlling the soil evaporation under non-energy limited conditions. Hence, the majority of climatological, meteorological, hydrological and agricultural applications require high resolution (HR) SM data. Surface SM observations are nowadays provided on a global basis using remote sensing data. Among all existing satellites, passive L-band microwave sensors are widely used to derive SM thanks to the strong physical link between the brightness temperature and the 0-5 cm SM profile (Kerr et al. 2007, Wagner et al. 2007). The downside to the operational retrieval of SM from microwave is given by the low resolution (LR) of the products, which ranges from 40 to 60 km (Njoku et al. 2003, Kerr et al. 2012), a resolution that is too coarse for most hydrological and agricultural applications.

On the other hand, optical sensors have the advantage of providing data at high and medium resolutions. In particular, Landsat and ASTER (Advanced Spaceborne Thermal Emission

and and Reflection radiometer) have resolutions of several tens of meters, while MODIS (MODerate Resolution Imaging Spectroradiometer) has a resolution of 1 km. Even though optical data could be used to derive SM, the main drawback in deriving a retrieval methodology is given by the sensors' sensitivity to meteorological conditions (cloud presence) and vegetation cover.

However, a synergy between the LR microwave and HR optical data (Zhan et al. 2002) can be used in order to derive SM at various spatial scales. This is achieved by using soil temperature and vegetation cover data (provided by the optical sensors), which are linked to the soil water content (Fang et al. 2013). Most of the methods based on the synergy between microwave and optical data generally use the triangle (Carlson et al. 1994) or trapezoid approach (Moran et al. 1994), in which the variations in land surface temperature (LST) are linked to variations in soil water content and vegetation cover (Carlson 2007, Petropoulos et al. 2009).

One category of methods that is based on the synergy between microwave and optical data regroups purely empirical algorithms, based on a polynomial fitting between LST, NDVI (Normalized Difference Vegetation Index) and SM. Studies like Piles et al. (2011) also take into account the brightness temperature in the polynomial fitting model in order to derive SM at 10 km and 1 km resolutions from SMOS observations. This approach was proven to reduce the bias but overall degrade the spatio-temporal correlation between SMOS SM and in situ measurements (Piles et al. 2011).

Another category based on the triangle/trapezoid approaches is represented by semi-physical methods (evaporation-based). These methods replace the polynomial function with a physical model that uses evaporation to explain the variability of SM (Merlin et al. 2008). In particular, they represent the spatial link between optical-derived soil evaporative efficiency (SEE, defined as the ratio of actual to potential evaporation) and surface SM. Their main

advantages over the polynomial fitting methods are that i) they are self calibrated and ii) the average of the estimated HR SM is equal to the LR observed SM.

Merlin et al. (2012b, 2013) have improved the algorithm presented in Merlin et al. (2008). DISPATCH (DISaggregation based on a Physical and Theoretical scale CHange) converts HR MODIS-derived SEE fields into HR SM fields by expanding a first order Taylor series of a SEE model around the LR SMOS SM. The HR SEE fields are derived using HR LST and NDVI data and the low resolution extreme temperatures (T_{ext}) which are estimated from HR optical data. Optical data is then linked to SM by using a self calibrated SM-based SEE model. To sum up, DISPATCH relies on two different SEE models: a temperature-based model, used to derive HR SEE from MODIS data, and a SM-based model, used to link the temperature-based SEE to SM. The current version of DISPATCH is contextual, meaning that the MODIS-derived SEE is a function of T_{ext}, which are determined from the image-based trapezoid method. Since contextual methods involve determining the wet and dry boundaries of LST, which may or may not be present within the scene at the observation resolution, limitations concerning the estimation of T_{ext} arise when fully dry and fully wet conditions are not met at the observation resolution. In particular, the image-based algorithms should provide good estimations of T_{ext} over semi-arid irrigated areas, when using HR data (de Tomas et al. 2014), such that three main requirements are fulfilled: i) the study area is relatively flat, ii) the atmospheric conditions are uniform and iii) extreme temperatures are actually observed at the thermal sensor resolution within the area of interest (Long & Singh 2011, Long & Singh 2012a, Long & Singh 2012b, Long & Singh 2013, Timmermans et al. 2007, Yang & Shang 2013). The image-based T_{ext} algorithms are less adequate for homogeneous rainfed areas, when using medium (1 km) to low resolution thermal remote sensing data (Djamai et al. 2015b).

In an effort to remove some of the above limitations, algorithms that estimate T_{ext} independently of surface conditions present within the study area (and hence independent of the spatial resolution of LST data) have been developed. T_{ext} are thus modeled by running a soil energy balance model forced by available meteorological data in both dry and wet conditions. (Stefan et al. 2015, Merlin et al. 2016).

In this context, this study aims to develop a new version of the DISPATCH algorithm, named DISPATCH-E, by integrating a soil energy balance (EBsoil) model forced by meteorological data at low resolution in order to derive T_{ext}. EBsoil aims to improve the physics and robustness of the temperature-based SEE model and hence the calibration strategy of the SM-based SEE model.

The approach is tested using SAFRAN (Système d'analyse fournissant des renseignements atmosphériques à la neige) meteorological data and *in situ* SM measurements for validation over a mixed dry and irrigated area in Catalunya, Spain, in 2011 and 2012. Firstly, the behavior of the temperature-based SEE model as a function of SMOS SM is analyzed. Secondly, the DISPATCH algorithm is run using the SM-based SEE model fitted to the temperature-based SEE model. Both DISPATCH and DISPATCH-E temperature-based SEE models are compared in time with SMOS SM. Thirdly, results in terms of disaggregated SM are then compared to *in situ* measurements.

The study site along with all the data are presented in Section 2. In Section 3, the DISPATCH algorithm is briefly presented, along with the two SEE models (temperature-based and SM-based) and the development methodology for DISPATCH-E. Section 4 covers the results in terms of T_{ext}, SEE estimates and disaggregated SM.

2. Data description

2.1. Validation site and *in situ* data

The study area is located in the Urgell region, in Catalunya, northeast of Spain. It presents a semiarid Mediterranean climate, with a mean yearly precipitation of 400 mm, 60 days of rain, and a mean yearly temperature of 16°C. Various irrigated crops (wheat, maize, alfalfa, apple and pear trees) and dryland crops (barley, olive trees, vineyards and almond trees) are present within the area.

The Urgell field experiment, conducted in 2011 and 2012, focused on a 20 km by 20 km area to collect 0-5 cm soil moisture data using the gravimetric technique. The sampling spanned a total of ten days in 2011: on day of year (doy) 97, doy 98, doy 146, doy 147, doy 165, doy 196, doy 228, doy 229, doy 244 and doy 277, and four days in 2012: doy 87, doy 103, doy 151 and doy 167. The sampling covered a total of four areas, each of 3 km by 3 km. Two are situated in the dryland area and two in the irrigated area, with ten points per sampling area. Three separate measurements per sampling point were performed, which means that a total of 120 measurements were taken within the entire area. As reported in Merlin et al. (2013), soil particle analysis at each sampling point, with a mean clay fraction of 0.24 and a mean sand fraction of 0.37, was used to derive soil texture. The approach described in Saxton et al. (1986) was then used to convert the gravimetric measurements to volumetric values.

In situ measurements serve as a validation dataset for the downscaled SM products. Since they are point measurements, they were aggregated (simple average) to a 3 km resolution.

2.2. Remote Sensing data

The SMOS Level-3 1-day global SM product (MIR CLF31A/D), version 2.72 (in 220 reprocessing mode RE02) product is used. SMOS data are extracted over a 200 km by 240 km area, and re-sampled at a resolution of ~40 km.

The re-sampling methodology described in Merlin et al. (2010c) and Molero et al. (2016) was used in order to obtain re-sampled SMOS data which overlap four times over the study area. The final downscaled soil moisture product is generated on the intersection of these four SMOS grids.

The MODIS version-5 land surface temperature (LST) products onboard Terra (MOD11A1) and Aqua (MYD11A1) and normalized difference vegetation index NDVI (MOD13A) were downloaded using the NASA Land Processes Distributed Active Archive Center (LP DAAC). Both the LST and NDVI products are re-sampled on a 1 km grid.

Elevation data extracted from the GTOPO30 digital elevation model (DEM) are also required.

Fig. 1 shows the study area on which the downscaling algorithm is applied, consisting of 30 re-sampled SMOS pixels (40 km by 40 km each). We focus our attention on six SMOS pixels in particular (hereby numbered as 11, 12, 13, 21, 22 and 23), which surround an area of 20 km by 20 km, where the *in situ* measurements were performed. A visual of the 1 km resolution Terra 16-day NDVI on doy 161 in 2011 is also represented, where we can clearly distinguish the irrigated and dryland areas.

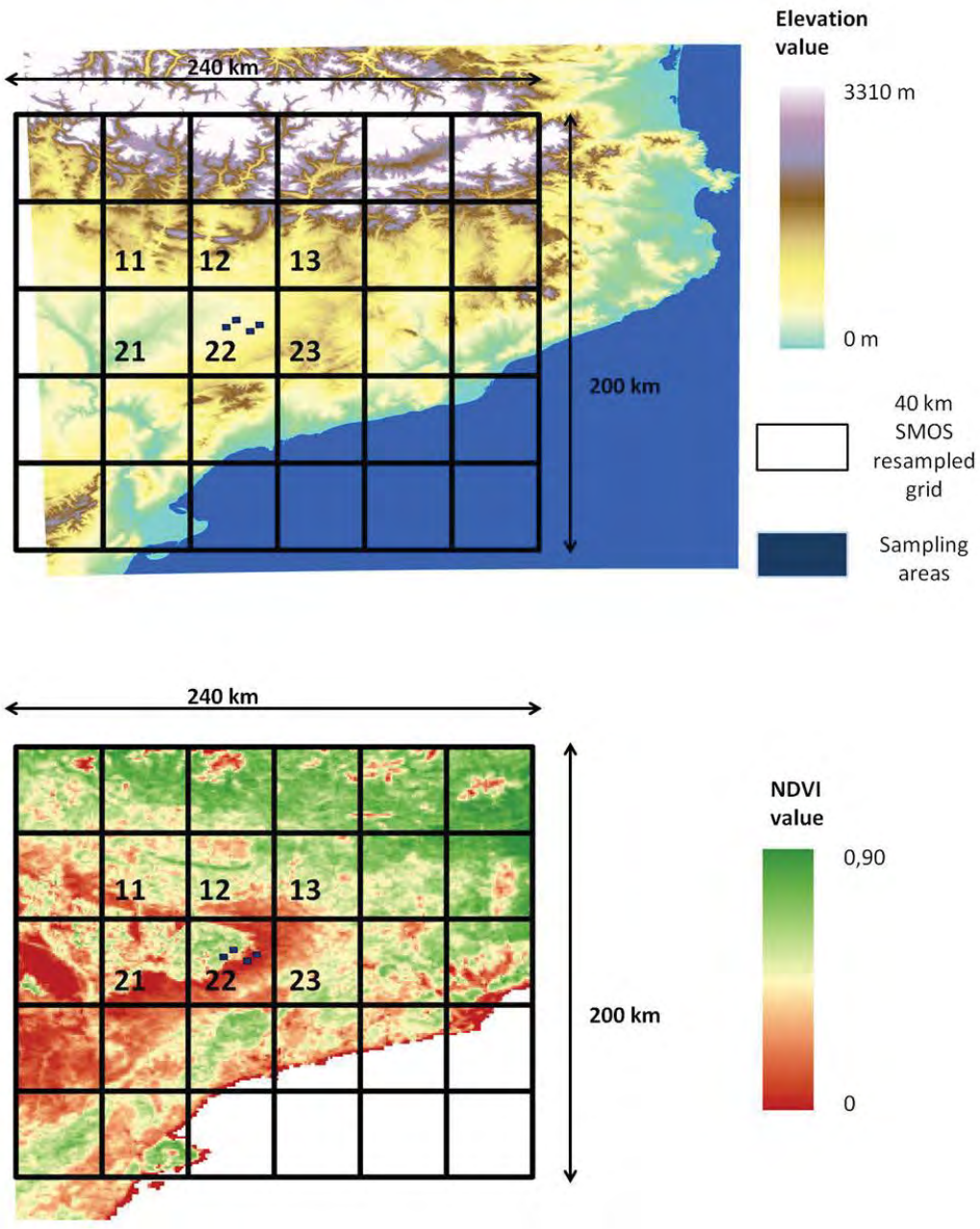


Fig. 1. Study area.

2.3. Meteorological data

Meteorological data consisting of wind speed, relative air humidity, air temperature and solar radiation are given by the SAFRAN (Système d'analyse fournissant des renseignements atmosphériques à la neige) model, which is an atmospheric analysis system for surface variables (Quintana Segui et al. 2016). It consists of daily data, at an hourly time-step. A pre-processing of the data was first performed, with data extracted at the two distinct MODIS overpasses as follows: data corresponding to 10:00 am and 11:00 am were averaged together in order to obtain values corresponding to the Terra overpass time over the area (10:30 am). Similarly, data from 1:00 pm and 2:00 pm were averaged together to obtain values corresponding to the Aqua overpass time (1:30 pm). The second step in the pre-processing consisted in extracting the meteorological data corresponding to the SMOS resampled grids that cover the study area. As a final step, the data, which are provided at a 5 km resolution, were upscaled at the SMOS resolution by simple linear averaging.

The data used in this study are available at the HyMex database (http://mistrals.sedoo.fr/10.14768/MISTRALS-HYMEX.1388).

3. Methodology

3.1. DISPATCH

The operational version of the DISPATCH algorithm (Merlin et al. 2013) is fully described in Molero et al. (2016). Only the essential aspects are reminded herein. Briefly, DISPATCH converts high resolution (HR) MODIS-derived SEE (Soil Evaporative Efficiency) fields into HR SM fields by expanding a first order Taylor series of a SEE model around the low resolution (LR) SMOS SM value. Different combinations of SMOS SM (ascending 6 am and descending 6 pm) and MODIS (Terra overpass 10:30 am and Aqua 1:30 pm from one

day before until one day after the SMOS overpass) data are used to generate an input ensemble. The DISPATCH product is the average at 1 km resolution of the disaggregated output ensemble.

The downscaling relationship is written as:

$$SM_{HR} = SM_{LR} + dSM_{mod}/dSEE(SEE_{LR}) * (SEE_{HR}-SEE_{LR}) \quad (1)$$

With SM_{HR} being the 1 km disaggregated SM, SM_{LR} the SMOS observation, $dSM_{mod}/dSEE(SEE_{LR})$ the partial derivative of SM with respect to SEE, evaluated at SMOS scale, SEE_{HR} the MODIS-derived SEE, and SEE_{LR} its average within the SMOS pixel.

DISPATCH thus relies on two different SEE models: a temperature-based model, used to derive SEE_{HR} from MODIS data, and a SM-based model (partial derivative of SM with respect to SEE in eq. (1)), used to link the temperature-based SEE to SM. Both models are presented below.

3.2. Temperature-based SEE

The estimation of SEE at HR, supposed to be approximately constant during the day, given clear sky conditions (Shuttleworth et al. 1989, Nichols & Cuenca 1993, Crago & Brutsaert 1996), is attained by:

$$SEE_{HR} = (T_{s,dry} - T_s)/(T_{s,dry} - T_{s,wet}) \quad (2)$$

Where T_s is the MODIS-derived soil temperature, $T_{s,dry}$ the surface soil temperature corresponding to bone dry conditions (SEE = 0) and $T_{s,wet}$ the surface soil temperature in fully humid conditions (SEE = 1). The assumed linear relationship between soil temperature and SEE in eq. (2) was recently verified using a physically-based soil energy balance model (Merlin et al. 2016).

The soil temperature in eq. (2) is obtained by decomposing the MODIS LST into its two basic components – soil and vegetation. The partitioning method is based on the "hourglass" approach of Moran et al. (1994) and is fully described in Merlin et al. (2012).

Note that MODIS LST is corrected beforehand for topographic effects (specifically decrease of air temperature with altitude) as in Merlin et al. (2013).

3.2.1. Image-derived extreme temperatures

The current version of DISPATCH estimates the extreme temperatures (T_{ext}) of eq. (2) using MODIS data solely. In practice, $T_{s,dry}$ and $T_{s,wet}$ are derived from the LST-f_v (fractional vegetation cover) feature space (Molero et al. 2016):

- $T_{s,dry}$ is set to the maximum surface temperature.
- $T_{s,wet}$ is calculated by extrapolating the wet edge passing through the well-watered full-covering green vegetation (having a temperature $T_{v,wet}$) until the bare soil line. Explicitly, it is the intercept at f_v=0 of the line passing through the point (1,$T_{v,wet}$) and the point with f_v<0.5, such that the slope of the line is maximum (all points with f_v<0.5 being located above the wet edge). $T_{v,wet}$ is set to the minimum surface temperature.

The image-based model is well adapted for semi-arid irrigated areas, provided that HR data are used (de Tomas et al. 2014), such that the heterogeneity is resolved at the observation scale (Long & Singh 2011, Long & Singh 2012a, Long & Singh 2012b, Long & Singh 2013, Timmermans et al. 2007, Yang & Shang 2013). They are less adequate for homogeneous rainfed areas, and when using medium (~1 km) to low resolution thermal remote sensing data, when the wet and dry conditions (Djamai et al. 2015b) are poorly estimated (Stefan et al. 2015).

3.2.2. Modeling of extreme temperatures based on soil energy balance (EBsoil)

In an effort to improve the algorithm's robustness, a new approach is implemented in order to estimate extreme temperatures independently of the surface conditions within the study area.

Extreme soil temperatures are estimated using a soil energy balance model (EBsoil) forced by meteorological data as in Stefan et al. (2015) and Merlin et al. (2016).

By prescribing a soil evaporation resistance r_{ss} equal to 0 and infinity (in practice a very large number), the minimum and maximum soil temperatures are estimated for a given atmospheric forcing. More specifically, EBsoil initializes Ts with the air temperature T_a and iterates on T_s until thermal equilibrium is reached.

More details about the EBsoil model can be found in Appendix A.

Meteorological conditions within each SMOS pixel are supposed to be uniform, and the extreme temperatures, which are calculated using data at a 1 km resolution, but are defined at the SMOS resolution of 40 km, retain the uniformity hypothesis.

3.3. SM-based SEE

3.3.1. Linear model

The current version of DISPATCH links the SM and the SEE using a semi-empirical linear model (Budyko 1956, Manabe 1969):

$$SEE_{HR} = SM_{HR}/SM_p \quad (3)$$

With SM_p a parameter estimated at LR as:

$$SM_p = SM_{LR}/SEE_{LR} \quad (4)$$

The performance of this model has been successfully assessed within DISPATCH in Merlin et al. (2013). The linearity assumption was proven to be adequate at the kilometric scale. Moreover, the reasons behind using a linear model as in eq. (3) were the potential robustness it could provide over a nonlinear model with an erroneous behavior, and that the SM_p parameter calibration could contribute to the description of the real behavior of SEE (Merlin et al. 2013). However, SEE is known to have a strongly nonlinear behavior over the full SM range (Lee & Pielke 1992, Chanzy et al. 1993, Merlin et al. 2011), which represents a fundamental limitation of the SEE model in equation (3). In fact, the derivative of SEE with respect to SM is influenced by the SM range, and generally decreases with SM. (Komatsu 2003).

3.3.2. Nonlinear model

Previous versions of the DISPATCH algorithm (Merlin et al. 2008, 2010a, 2012b) have tested a range of nonlinear models of SEE derived from Komatsu (2003), Lee & Pielke (1992) and Noilhan & Planton (1989). As reported in Merlin et al. (2010a), a greater stability (slope is more stable around 1) of the algorithm is obtained using the Komatsu (2003) model than with the other two models, stability which is already present in the linear mode. Merlin et al. (2013) have studied the impact of the nonlinearity effects on the disaggregated SM and they concluded that when using HR (~100 m) data to derive SEE, a nonlinear model is a more adequate choice, as it increases the slope of the linear regression between downscaled products and *in situ* measurements, thus improving the spatial representativeness of SM.

Regarding nonlinear SEE representations, the model presented in Komatsu (2003) presents a particular interest, and the reason is twofold: i) it lowers the derivative in eq. (1) ($dSM_{mod}/dSEE$) in the lower ranges of the SM_{LR} and ii) it increases $dSM_{mod}/dSEE$ in the higher ranges of the SM_{LR}. It is important not to have large values of the derivative at SM_{LR}

283 ~0, as a small derivative implies uniform 1 km SM data be approximately equal to $SM_{LR,}$
284 and in dry conditions will compensate for errors in the estimated T_{ext}.
285 Especially, the models presented in Noilhan & Planton (1989) and Lee & Pielke (1992) do
286 not allow a low derivative in the low ranges of the $SM_{LR,}$ as the simulated SEE slowly
287 increases with SM in the lower ranges of SM. This implies that a very large derivative is
288 obtained, leading to erroneous large disaggregated values.
289 For the above mentioned reasons, the exponential form of Komatsu (2003) (hereby
290 mentioned as K03) is chosen as a nonlinear SEE model:
291 $$SEE_{sim} = 1 - \exp(SM/SM_c) \quad (5)$$
292 where SM_c is a semi-empirical parameter. In our case, SM_c is supposed to be constant in
293 space (within each LR pixel) and is calibrated at LR for each SMOS pixel from the times
294 series of SMOS, MODIS and SAFRAN data.
295 An iterative loop is run on SM_c (initialized at 0) until the simulated SEE approaches the
296 observed SEE. In other words, until the cost function $||SEE_{sim} - SEE_{obs}||^2$ is minimum. The
297 observed SEE is the SEE_{LR} (spatial linear average of SEE_{HR}).
298

299 **3.4. DISPATCH-E**

300 The introduction of the EBsoil model in the computation of T_{ext} allows for the development of a
301 new version of DISPATCH, called DISPATCH-E (where E stands for evaporation). The goal is
302 to improve the physics and robustness of the MODIS temperature-based SEE model and hence
303 the calibration strategy of the (SMOS) SM-based SEE model. This could be of particular
304 interest over areas that do not include both fully dry and fully wet conditions at the MODIS
305 resolution, where and when the extreme temperatures may be poorly estimated from MODIS
306 data alone. In practice, the extreme temperatures of the temperature-based SEE model are first
307 estimated by the EBsoil model, independently of MODIS data. The SM_c parameter of the SM-

based SEE model is then calibrated for each LR pixel from the time series of LR (aggregated) temperature-based SEE and LR SMOS SM. Note that EBsoil is forced at LR by using SAFRAN meteorological data aggregated (linear average) at SMOS resolution. A schematic diagram of both DISPATCH and DISPATCH-E methods is proposed in Fig. 2.

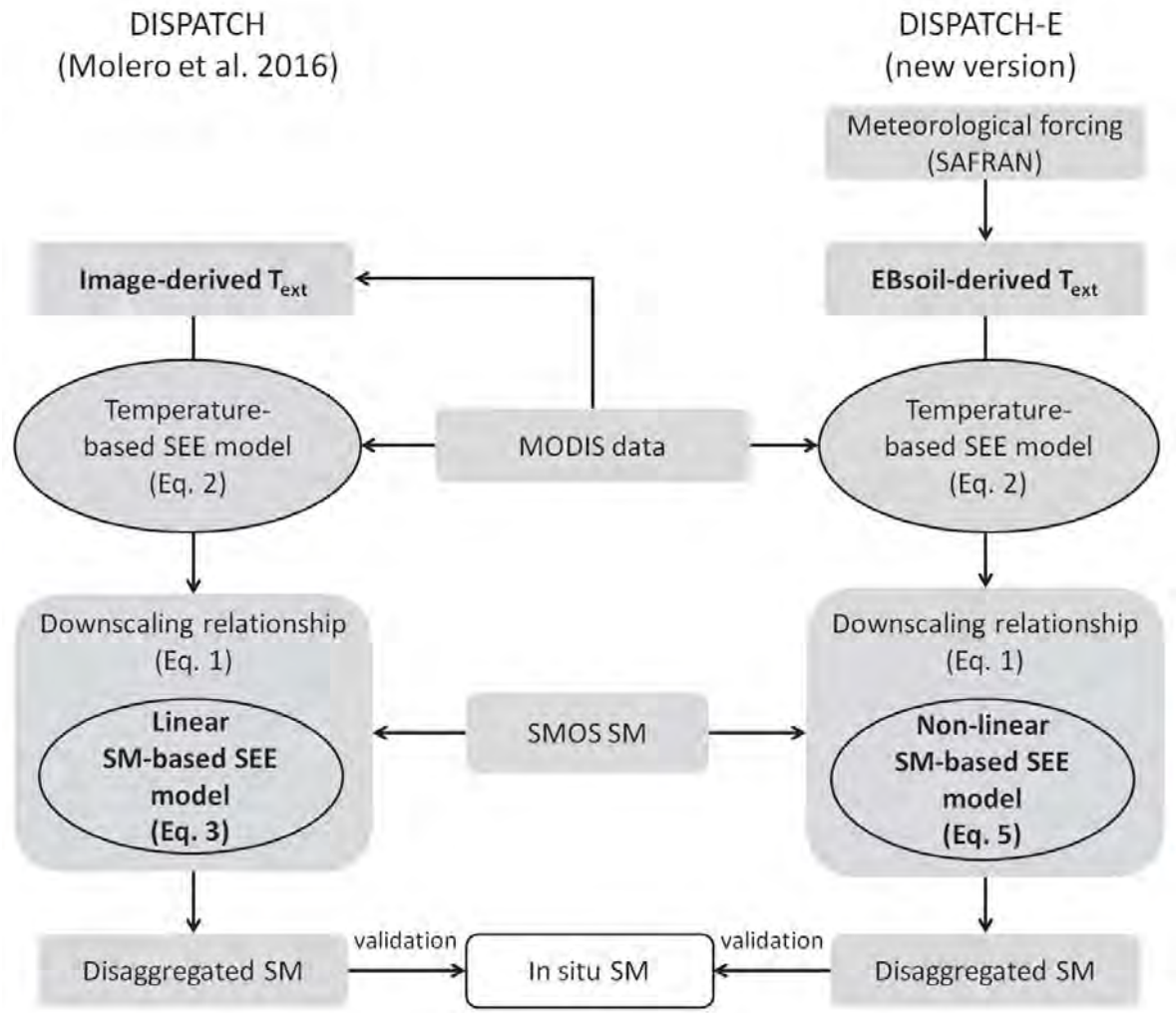


Fig. 2. Schematic diagram of DISPATCH and DISPATCH-E.

4. RESULTS AND DISCUSSION

In this section, the performance of the EBsoil-DISPATCH coupling scheme (DISPATCH-E) is assessed for the study area in terms of extreme temperatures in dry/wet conditions, temperature-based and SM-based SEE estimates, and disaggregated SM values. First, the extreme temperatures estimated by both EBsoil and the image-based algorithm are analyzed. Then, the temperature-based SEE estimates obtained from either the MODIS- or EBsoil-derived extreme temperatures are compared to the SM-based SEE estimates. Finally, the SMOS SM disaggregated by DISPATCH and DISPATCH-E in both linear and nonlinear modes is evaluated against the *in situ* measurements collected in 2011 and 2012.

4.1. Extreme soil temperatures

Fig. 3 plots the MODIS-derived extreme soil temperatures with respect to the EBsoil-derived extreme soil temperatures, for pixels 22 and 13, for the time series spanning 2011 till April 2012, and for Aqua and Terra overpasses separately. These two pixels were chosen as they are representative of the wet and dry (mixed) conditions over the study area.

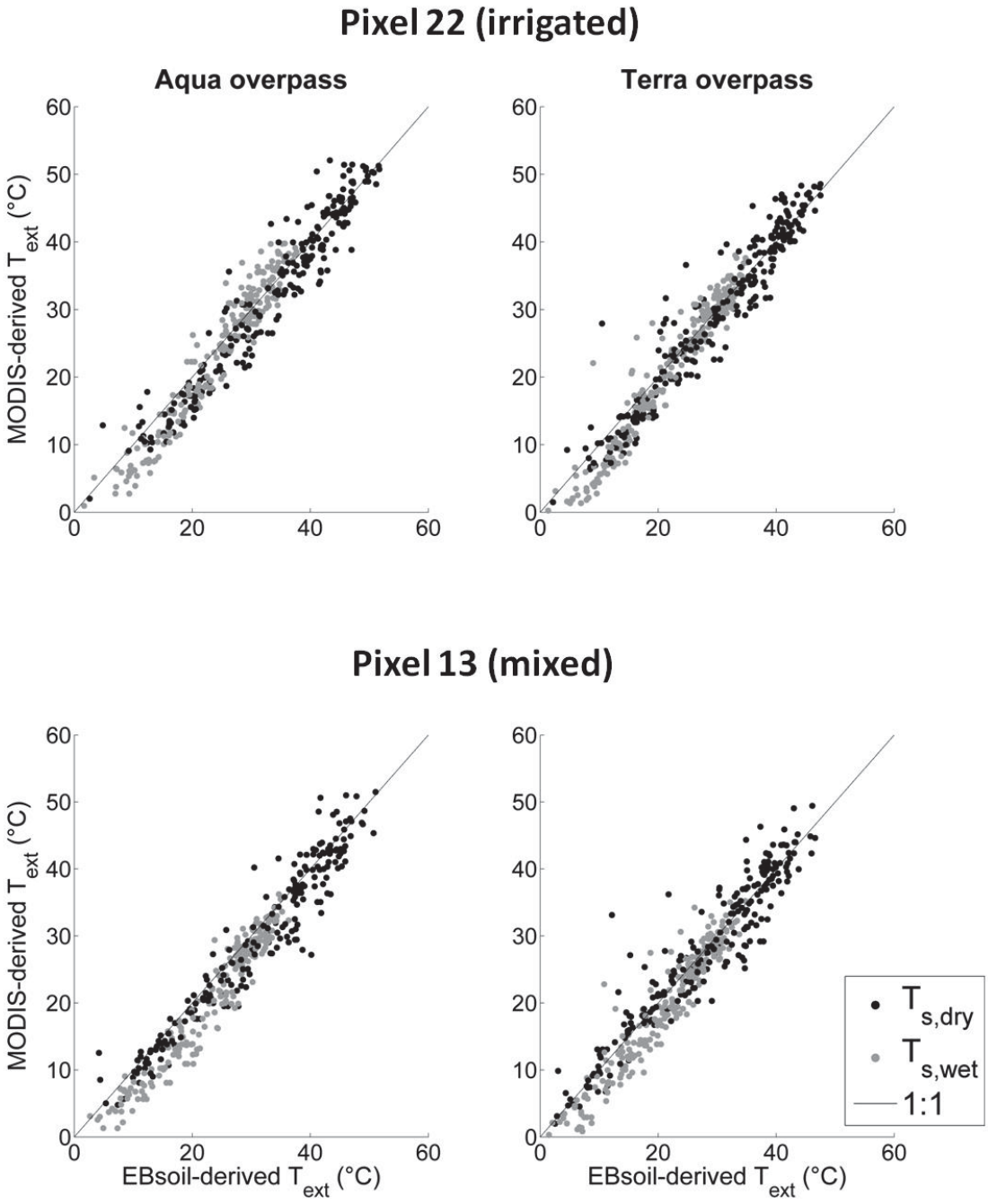


Fig. 3. The MODIS-derived soil T_{ext} are plotted against the EBsoil-derived soil T_{ext} for two pixels representative of irrigated and mixed conditions.

A strong consistency can be observed between modeled and image-based extreme temperatures, for both pixels. With respect to $T_{s,dry}$, the data points are generally more

337 evenly scattered along the 1:1 regression line than the $T_{s,wet}$ data points, which tend to be
338 underestimated in the lower range when derived from MODIS images.
339 Statistical results in terms of slope of the linear regression, correlation coefficient, RMSD
340 and bias between modeled and remotely-sensed extreme soil temperatures are presented in
341 Table 1.
342 A correlation coefficient of 0.96 is obtained for pixel 22, for both MODIS overpasses and
343 for both $T_{s,dry}$ and $T_{s,wet}$. In the case of pixel 13, the correlation coefficient is slightly
344 reduced from 0.96 ($T_{s,wet}$ and $T_{s,dry}$ respectively) for the Aqua overpass to 0.95 and 0.94 for
345 the Terra overpass. In addition, the RMSD between modeled and image-based temperatures
346 is relatively low (near 3°C), compared to the range of $T_{s,wet}$ and $T_{s,dry}$, which coupled with
347 good correlations is characteristic of better accuracy. Irrigated pixel 22 presents both dry
348 and wet conditions, so the estimation of extreme temperatures using solely MODIS data is
349 more accurate than in the case of pixel 13, which is mostly represented by dry conditions.
350 In general, lower values of the RMSD are obtained for the Terra overpass than the Aqua
351 overpass. One reason could be the drying of the soil surface later in the day (near the Aqua
352 overpass), so the minimum soil temperatures derived from the images might not be
353 reflective of the fully humid condition. An additional explanation could be the advection
354 effects, as the atmospheric conditions are more stable at 10:30 am (Terra overpass) than at
355 1:30 pm (Aqua overpass).
356 In general, the slope of the linear regression is closer to 1 when analyzing the results over
357 the 13 pixel, which implies that the sensitivity of the modeling is closer to the observations.
358

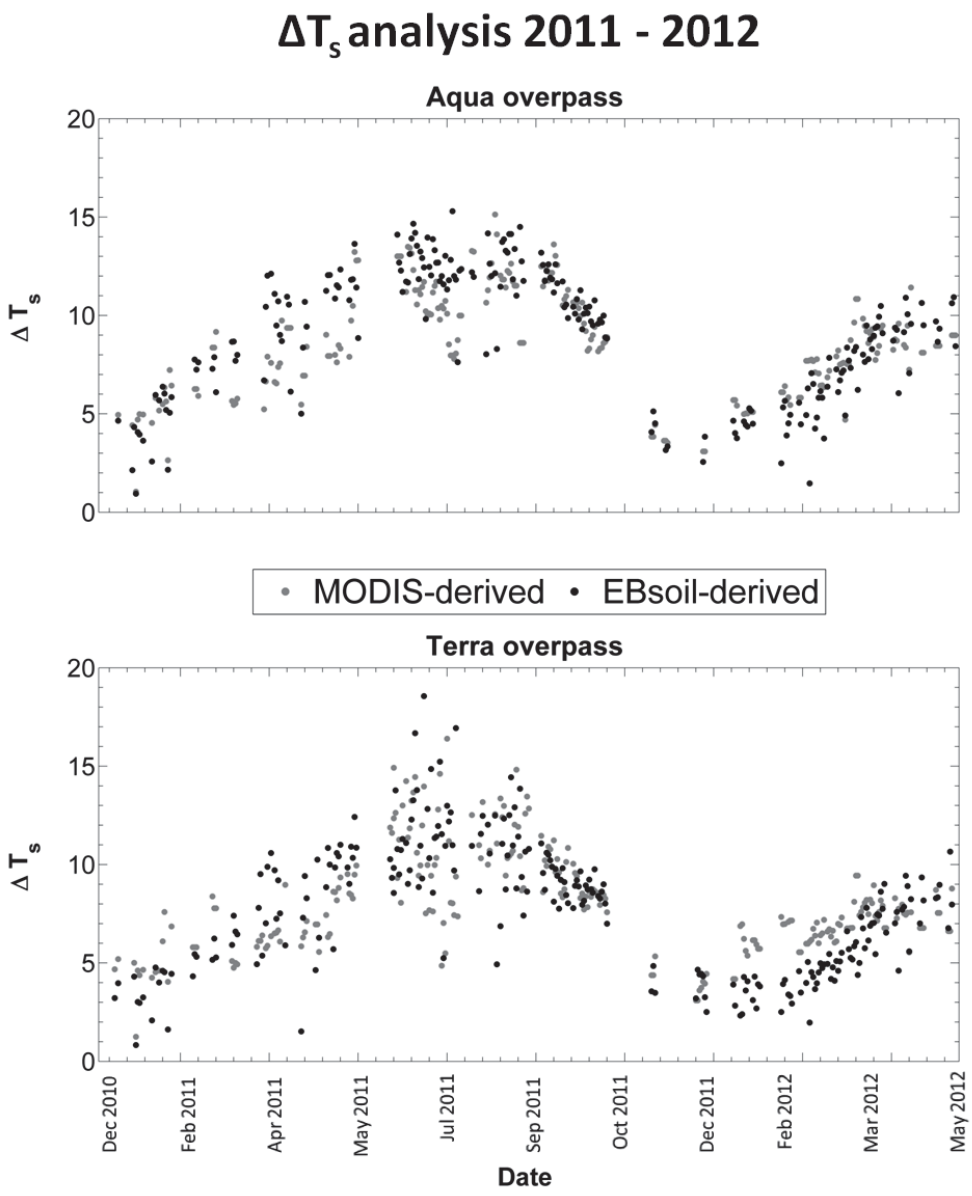
Pixel ID	**AQUA overpass**								**TERRA overpass**							
	$T_{s,wet}$				$T_{s,dry}$				$T_{s,wet}$				$T_{s,dry}$			
	R	S	RMSD (°C)	Bias (°C)	R	S	RMSD (°C)	Bias (°C)	R	S	RMSD (°C)	Bias (°C)	R	S	RMSD (°C)	Bias (°C)
22	0.96	1.19	3.20	-0.67	0.96	1.05	3.57	-1.19	0.96	1.18	3.06	-0.89	0.96	1.04	3.41	-0.52
13	0.96	1.08	3.78	-2.74	0.96	0.98	3.64	-1.46	0.95	1.06	3.13	-1.48	0.94	0.94	3.84	0.19

Table 1. Correlation coefficient (R), slope of the linear regression (S), root mean square difference (RMSD) and bias between MODIS-derived and EBsoil-derived soil T_{ext}, for AQUA and TERRA overpasses separately.

All in all, consistent results over the two (irrigated and dryland) pixels are obtained, at the two MODIS overpass times, which reflects the robustness and precision of the EBsoil-derived extreme temperatures over highly heterogeneous areas, and validates both the physical approach as well as the image-based algorithm.

The sensitivity of the difference (noted as ΔT_s) between $T_{s,dry}$ and $T_{s,wet}$ to atmospheric conditions (especially evaporative demand) is important to consider as it impacts the performance of DISPATCH. For instance, when ΔT_s is below the MODIS uncertainty threshold in LST (ranging from 1°C to 4.5° C according to Hulley et al. 2012, Wan et al. 2002, Wan et al. 2004, Coll et al. 2005), disaggregation results are expected to be less reliable.

As a step further, Fig. 4 plots ΔT_s over the entire time series, for pixel 22, taking into account the two separate MODIS overpass times.



382

383 Fig. 4. The difference (ΔT_s) between $T_{s,dry}$ and $T_{s,wet}$ is plotted over the entire time series for pixel

384 22, for Aqua and Terra separately.

385 Therefore, the algorithm should be optimal in summer, when ΔT_s reaches 15° C and minimal

386 in winter. The physical reason is that during wintertime, the soil evaporation is controlled by

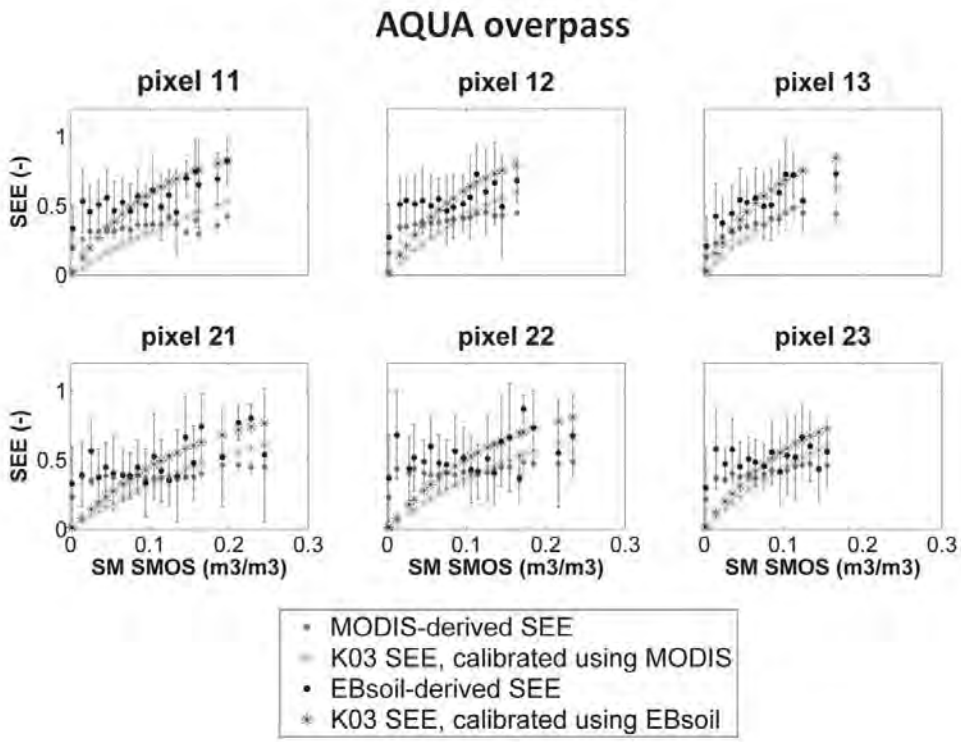
387 potential evaporation (energy-limited evaporation), whereas during the summer season, it is

388 controlled by soil moisture (moisture-limited evaporation).

We can also note that the temperature-based models give different results in terms of ΔTs, with the EBsoil-derived ΔT_s appearing to be more sensitive for the Terra overpass during summertime. The higher values of ΔT_s obtained for the Terra overpass during summertime could be explained by the higher incoming solar radiation heating the soil around noon. There is also a noticeable difference between EBsoil-derived and MODIS-derived ΔT_s in wintertime, when the EBsoil-derived ΔT_s is inferior to the image-derived ΔT_s. This could be explained by the soil temperatures in wintertime being controlled by potential evaporation, which in turn is influenced by the incoming net radiation, low in wintertime. As the incoming net radiation is an input to the energy balance model used in EBsoil, this directly impacts the ΔT_s.

4.2. SEE behavior

Figures 5 and 6 show the observed (temperature-based) SEE and the simulated (SM-based) SEE against the SMOS SM, for the entire time series, for six pixels surrounding the study area, and for the Aqua and Terra overpasses separately. The SM-based model here is the nonlinear one. In order to have a more representative behavior of the SEE over the time series, the analysis takes into account SEE values averaged per 0.01 SMOS SM bins, with at least 3 values per bin. The standard deviation within each bin is also computed. The SM-based SEE is fitted (via the calibration of the SM_c parameter) to the temperature-based SEE, whose extreme temperatures are derived from either MODIS images or EBsoil, for each SMOS pixel (11, 12, 13, 21, 22 and 23) covering the study area. The values (estimated over pixel 22 and used as a proxy for other pixels when deriving SM_{HR}) of the calibrated SM_c parameter are 0.22 m^3m^{-3} (Aqua overpass) and 0.24 m^3m^{-3} (Terra overpass) when taking into account the MODIS-derived SEE as a reference for the fitting. When considering the EBsoil-derived SEE as a reference, the obtained SM_c values are 0.14 m^3m^{-3} (Aqua overpass) and 0.20 m^3m^{-3} (Terra overpass).



414

415 Fig. 5. The observed (temperature-based) SEE and the simulated (SM-based) SEE is plotted
416 against the SMOS SM, for the entire time series, for the Aqua overpass. The observed SEE
417 is derived either from MODIS, or from EBsoil. The SM-based SEE uses the nonlinear K03
418 model and is calibrated from the observed SEE.

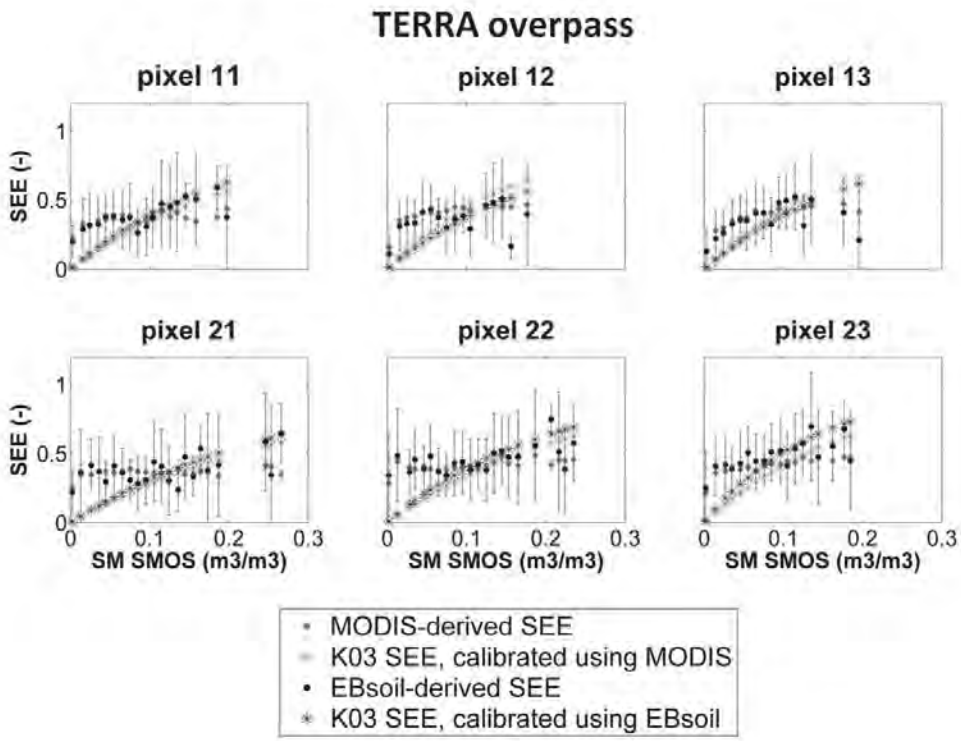
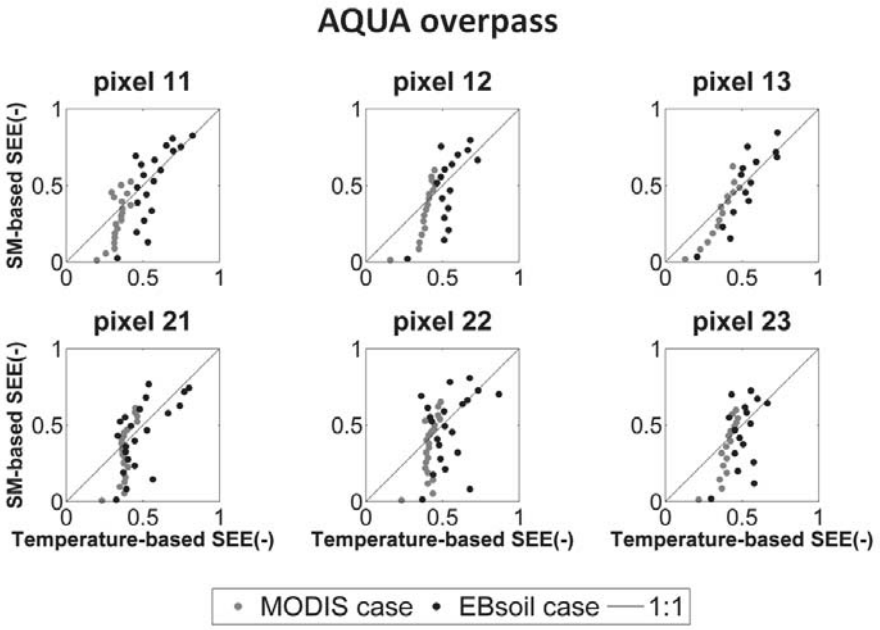


Fig. 6. Same as Fig. 5, but for the Terra overpass.

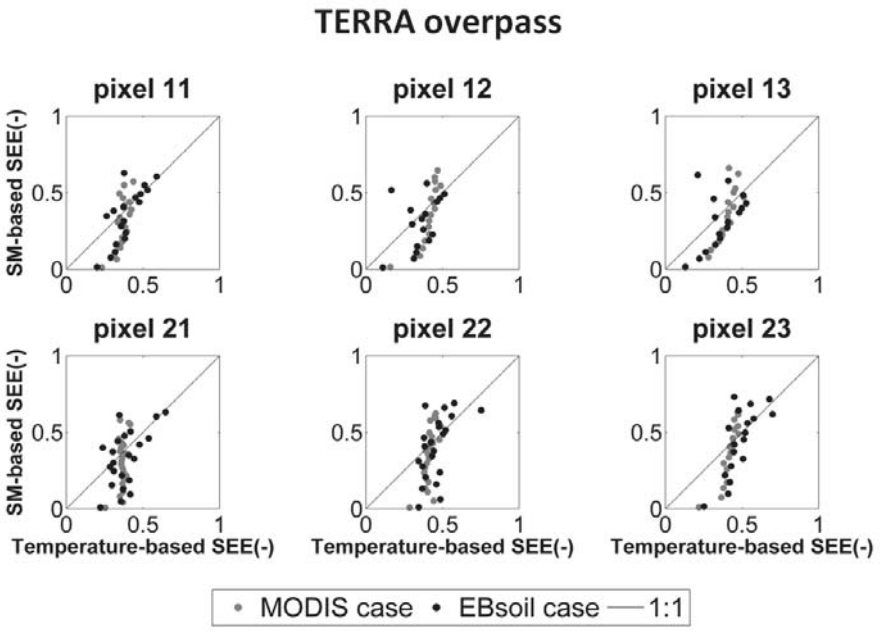
By comparing the two temperature-based SEE, one observes that as opposed to the relatively flat behavior of the MODIS-derived SEE, the EBsoil-derived SEE's range of values is wider. The MODIS-derived SEE in the higher range of SMOS SM retains a rather flat behavior compared to the EBsoil-derived SEE. This could be explained by the poorer estimates of the MODIS-derived $T_{s,dry}$ and $T_{s,wet}$, which are not reflective of the true extreme temperatures under certain conditions. The SEE calculated from these estimates hence tends to maintain a smaller range of values. In addition, one observes that the SEE behavior is different for each pixel, possibly due to different soil properties, vegetation covers, land use (irrigated/dry), and topography effects. It is well known that the shape of SEE depends on soil properties, whose common proxy is given by sand and clay percentages (Merlin et al. 2016). Pixels 21 and 22 are mostly covered by crops, whereas the rest of the pixels share a certain percentage between crops and grass/shrubs, with small

433 scattered forests. Some of the crops are irrigated, which makes the estimates of extreme
434 temperatures in fully dry conditions, and therefore the retrieved SEE, poor. The image-
435 based estimates retrieved over intense canopies are also poorer with respect to estimates
436 over grass or shrubs. Errors could also appear due to topography effects, caused by the
437 mountainous areas close to our study site, close to pixels 11, 12 and 13, respectively. The
438 nominal behavior of the SEE with respect to SM should be observed over flat irrigated
439 pixels, where the dominant land cover is grass/shrubs (notably pixels 21 and 22). However,
440 the best behavior is observed over pixel 13, which is affected by topography and is half
441 covered by forests.
442 One can see that there are no values of the SEE in the lower ranges (below 0.15-0.2), which
443 could be explained by topographic and cloud effects that tend to induce a positive bias in
444 SEE due to low MODIS T_s.
445 Figures 7 and 8 show the simulated SEE as a function of the observed SEE, for the same
446 pixels, and for Aqua and Terra overpasses respectively.



447

448 Fig. 7. The simulated (SM-based) SEE is plotted against the observed (temperature-based)
449 SEE, for the Aqua overpass.



450

451 Fig. 8. Same as Fig. 7, but for the Terra overpass.

452 Statistical results per SMOS pixel are reported in Table 2, in terms of correlation coefficient
453 (R), slope of the linear regression (S), and RMSD. The metrics are computed using the SEE
454 values averaged per 0.01 SMOS SM bins.

Pixel ID	**AQUA overpass**						**TERRA overpass**					
	MODIS case			EBSoil case			MODIS case			EBsoil case		
	R	S	RMSD	R	S	RMSD	R	S	RMSD	R	S	RMSD
11	0.72	2.08	0.12	0.73	1.46	0.17	0.69	2.44	0.14	0.77	1.40	0.12
12	0.83	2.15	0.13	0.67	1.51	0.18	0.80	2.04	0.14	0.40	0.63	0.16
13	0.93	1.71	0.09	0.87	1.46	0.13	0.83	1.73	0.12	0.49	0.75	0.15
21	0.71	2.60	0.15	0.61	0.94	0.17	0.41	2.28	0.17	0.54	0.93	0.16
22	0.64	2.27	0.16	0.33	0.60	0.22	0.68	3.03	0.17	0.54	1.20	0.18

23	0.86	2.53	0.13	0.40	0.99	0.20	0.86	2.68	0.14	0.73	1.52	0.15
Mean	0.78	2.22	0.13	0.60	1.16	0.18	0.71	2.36	0.15	0.58	1.07	0.15

Table 2. Correlation coefficient (R), slope of the linear regression (S) and root mean square difference (RMSD) between simulated and observed SEE, for each pixel individually, and per Aqua/Terra overpasses separately. Mean metrics are also computed (average over the six pixels).

The simulated (SM-based) versus observed (temperature-based) SEE gives comparable results for both formulations of the observed SEE (MODIS-derived or EBsoil-derived). However, for the MODIS-derived SEE scenario, better results are obtained in terms of R and RMSD, with reported mean values (average for all pixels) for the Aqua overpass of 0.78 and 0.13 as opposed to 0.60 and 0.18 obtained for the EBsoil-derived SEE. For the Terra overpass, the mean R and mean RMSD are 0.71 and 0.15 for the MODIS-derived SEE, while for the EBsoil-derived SEE, they are equal to 0.58 and 0.15.

The slope of the linear regression is however closer to 1 for the EBsoil-derived case scenario, with an average value for the Aqua and Terra overpasses of 1.16 and 1.07 as opposed to 2.22 and 2.36 in the MODIS-derived case.

To sum up, the MODIS-derived SEE is more precise, with better mean correlation values, while the EBsoil-derived SEE has a better sensitivity, with mean slope values approaching 1. We accept a compromise between the increase in RMSD for the EBsoil case because of the improvement in the slope.

4.3. Validation of SM disaggregation methodology

In this section, we investigate the performance of DISPATCH-E in terms of disaggregated SM values. Figures 9 and 10 plot the SMOS SM and the DISPATCH-E output with respect

to *in situ* measurements (aggregated at a 3 km scale by simple linear averaging), for Aqua and Terra overpass times respectively.

In each case, the DISPATCH-E outputs consist in using either the MODIS-derived extreme temperatures and a linear/nonlinear SEE(SM) model, or the EBsoil-derived extreme temperatures and a linear/nonlinear SEE(SM) model in the downscaling relationship. The DISPATCH case corresponds to MODIS-derived SEE, while DISPATCH-E to EBsoil-derived SEE. Both are tested in the linear and nonlinear case scenarios.

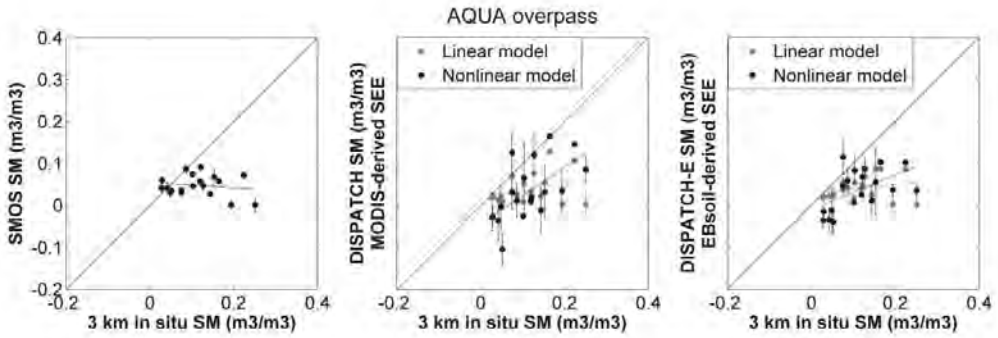


Fig. 9. The SMOS, DISPATCH and DISPATCH-E SM are plotted against 3 km aggregated *in situ* measurements, for the Aqua overpass.

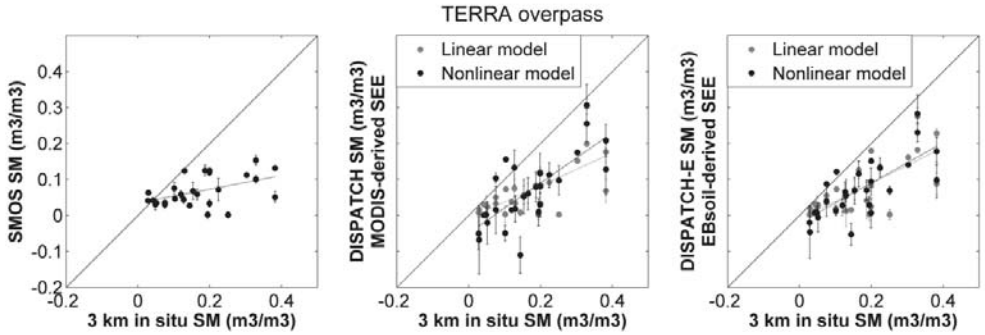


Fig. 10. Same as Fig. 9, but for the Terra overpass.

Statistical results in terms of correlation coefficient, slope of linear regression, bias, RMSD, unbiased RMSD (uRMSD) and the average of individual standard deviations (Mean_σ) are reported in Table 3.

	AQUA overpass					TERRA overpass					AQUA TERRA combined				
	SMOS	MODIS-derived		EBsoil-derived		SMOS	MODIS-derived		EBsoil-derived		SMOS	MODIS-derived		EBsoil-derived	
		Linear mode	Nonlinear mode	Linear mode	Nonlinear mode		Linear mode	Nonlinear mode	Linear mode	Nonlinear mode		Linear mode	Nonlinear mode	Linear mode	Nonlinear mode
R	-0.16	0.30	0.61	0.25	0.53	0.47	0.69	0.75	0.70	0.74	0.40	0.65	0.73	0.67	0.72
S	-0.063	0.16	0.66	0.10	0.41	0.19	0.45	0.70	0.50	0.56	0.16	0.41	0.69	0.43	0.54
Bias	-0.065	-0.074	-0.081	-0.072	-0.075	-0.10	-0.10	-0.10	-0.095	-0.096	-0.087	-0.089	-0.092	-0.085	-0.087
RMSD	0.096	0.096	0.099	0.095	0.093	0.14	0.13	0.12	0.12	0.12	0.12	0.11	0.11	0.11	0.11
uRMSD	0.071	0.062	0.058	0.061	0.055	0.091	0.075	0.071	0.073	0.069	0.085	0.071	0.067	0.069	0.064
Mean_σ	0.0093	0.016	0.044	0.017	0.030	0.011	0.016	0.044	0.019	0.032	0.011	0.017	0.042	0.018	0.030

Table 3. Correlation coefficient, slope of the linear regression, bias, RMSD, unbiased RMSD (uRMSD) and the average of individual standard deviations (Mean_σ) between SMOS SM/DISPATCH SM/DISPATCH-E SM an 3 km *in situ* data are reported for each MODIS overpass individually, as well as both overpasses combined. Both DISPATCH and DISPATCH-E are tested in linear and nonlinear modes.

When comparing the LR SM and the HR SM with the 3 km *in situ* data, one can see that DISPATCH improves the fine-scale precision. A correlation coefficient equal to 0.30 and 0.25 when using Aqua MODIS/EBsoil-derived SEE (linear mode) is obtained, as opposed to -0.16 when comparing SMOS SM with *in situ* data. The values for the same metrics are better for the Terra overpass than for the Aqua overpass, with a correlation coefficient equal to 0.69 and 0.70, as opposed to the 0.47 correlation value calculated for SMOS SM. Moreover the correlation seems to be stronger for days when rainfall occurred (doy 244 in 2011), proving that the algorithm is suited for detecting small scale changes, thanks to its 1 km disaggregation resolution.

The slope of the linear regression is also improved from -0.063 to 0.16 and 0.10 (Aqua overpass, linear mode) and from 0.19 to 0.45 and 0.50 (Terra overpass, linear mode).

When comparing the two temperature-based SEE models (in linear mode), one can see a relative similar performance in terms of correlation and slope of linear regression, for both MODIS overpasses. In terms of bias, all disaggregated SM products retain the negative bias found in the SMOS SM. The RMSD values seem to be consistent, independently of the SM resolution, with lower values (~0.09 m3/m3) reported for the Aqua overpass, as opposed to the Terra overpass (~0.12/0.13 m3/m3).

When comparing the linear and the nonlinear modes, one can conclude that the nonlinear SEE model significantly improves the correlation coefficient and the slope. The correlation coefficient is significantly increased for the Aqua overpass: from 0.30 to 0.61 (MODIS-derived SEE) and from 0.25 to 0.53 (EBsoil-derived SEE). The slope of the linear regression is also significantly improved, from 0.16 to 0.66 (MODIS-

527 derived SEE) and from 0.10 to 0.41 (EBsoil-derived SEE) – for the Aqua overpass
528 over a smaller SM range.

529 When looking at the statistics obtained when taking into account both Aqua and Terra
530 overpasses, the trend remains the same. The correlation coefficient is improved from
531 0.40 (SMOS SM) to 0.65 and 0.67 (MODIS/EBsoil-derived SEE) in linear mode. The
532 correlation values obtained in nonlinear mode are 0.73 and 0.72 (MODIS /EBsoil-
533 derived SEE). Similarly, the slope is increased from 0.16 to 0.41 and 0.43 (MODIS-
534 /EBsoil-derived SEE) in linear mode. In nonlinear mode, the slope is equal to 0.69
535 and 0.54 (MODIS /EBsoil-derived SEE). The uRMSD values are systematically
536 lower when using the EBsoil-derived SEE, with reported values (Aqua and Terra
537 combined) of 0.069 (linear mode) and 0.064 (nonlinear mode), as opposed to 0.071
538 and 0.067 when using the MODIS-derived SEE. In addition, the mean of individual
539 standard deviations is consistently smaller in nonlinear mode when using EBsoil-
540 derived SEE as opposed to MODIS-derived SEE, with values of 0.030 (Aqua
541 overpass) and 0.032 (Terra overpass) as opposed to 0.044 (Aqua and Terra
542 overpasses respectively).

543 Better results are reported in nonlinear mode than in linear mode. In fact, in the
544 nonlinear mode, the partial derivative of SM with respect to SEE is diminished in the
545 lower SM ranges and increased in the higher SM ranges, which entails an overall
546 better precision and accuracy of the corresponding disaggregated products when
547 compared to the 3 km aggregated *in situ* measurements.

548 An important aspect to mention is the negative values of disaggregated SM when
549 using the nonlinear SEE(SM) model in the downscaling relationship. If we assume
550 that the disaggregation is efficient, then this could point out that SMOS underestimates

SM in very dry areas (with SM close to zero). Various calibration and validation studies of SMOS SM products have reported a negative bias (Collow & Robock 2012, Zeng & all. 2015, Al Bitar et al. 2015, Djamai et al. 2016). The bias in retrieved SM is inflicted by biases in the brightness temperatures, which could be provoked by the RFI (Radio Frequency Interference) for instance. According to (Dente et al. 2012), a positive bias in the observed brightness temperature would imply a negative bias in the SM products. Since SMOS SM is used when calibrating the SM_c parameter, the parameter retrieval is also affected by the negative bias. The downscaled SM data obtained in the nonlinear mode is thus affected by both potential bias in retrieved SM_c values, as well as the negative bias in SMOS data.

Summary and Conclusions

DISPATCH provides 1 km resolution SM data from 40 km resolution SMOS and 1 km resolution MODIS data by combining two SEE models based on either MODIS LST or SMOS SM. Both models are originally self-calibrated from MODIS and SMOS data alone. While the extreme temperatures of the temperature-based SEE model are determined from MODIS images, the SM-based SEE model is calibrated from quasi-simultaneous MODIS and SMOS observations by assuming a linear SEE(SM) representation.

This paper develops a new version – named DISPATCH-E – of the DISPATCH method by integrating a soil energy balance (EBsoil) model forced by meteorological data at low (SMOS) resolution. EBsoil aims to improve the physics and robustness of the (MODIS) temperature-based SEE model and hence the calibration strategy of a (SMOS) SM-based SEE model. In practice, the extreme temperatures of the

575 temperature-based SEE model are now estimated by EBsoil (independently of MODIS
576 data), while the SM-based SEE model is now derived from the (observed) temporal
577 behavior of the temperature-based SEE as a function of SMOS SM.

578 The approach is tested by comparing the extreme temperatures in dry/wet conditions,
579 temperature-based and SEE-based estimates, and disaggregated SM values obtained
580 by DISPATCH and DISPATCH-E respectively, over a mixed irrigated/dry land area
581 in Catalunya (Spain) during 2011 and 2012. SAFRAN meteorological data aggregated
582 at SMOS resolution are used as an input to DISPATCH-E, and *in situ* SM
583 measurements are used for validation at a 3 km resolution.

584 Both DISPATCH and DISPATCH-E temperature-based SEE models are compared in
585 time with SMOS SM, for each SMOS pixel independently. A compromise between
586 precision and robustness is obtained: the SEE estimates based on the MODIS-derived
587 extreme soil temperatures are more precise than the SEE estimates based on the
588 EBsoil-derived extreme soil temperatures, with a mean correlation coefficient (average
589 of correlation coefficients corresponding to each pixel) equal to 0.78 (Aqua overpass)
590 and 0.71 (Terra overpass) as opposed to 0.60 (Aqua overpass) and 0.58 (Terra
591 overpass). However, the latter provides more accurate SEE estimates, with a mean
592 slope of the linear regression (average of the values corresponding to each pixel) equal
593 to 1.16 (Aqua overpass) and 1.07 (Terra overpass) as opposed to 2.22 (Aqua overpass)
594 and 2.36 (Terra overpass). Results indicate that EBsoil is a robust method to estimate
595 bone dry and fully wet soil temperatures within the 40 km SMOS pixel. Moreover, an
596 independent estimation of $T_{s,dry}$ and $T_{s,wet}$ provides a useful assessment of the relative
597 uncertainty in the temperature-based SEE at the time of MODIS overpass.

In terms of disaggregated SM values, SMOS SM and 1 km resolution disaggregated SM are both compared to 3 km resolution *in situ* measurements. The correlation coefficient (for Aqua and Terra combined) is increased from 0.40 for the SMOS SM to 0.65 and 0.67 for the disaggregated SM (obtained using either MODIS-derived or EBsoil-derived SEE).

The integration of a nonlinear SM-based SEE model, fitted separately to the two temperature-based SEE models, significantly improves the statistics. For the Aqua overpass, correlations of 0.61 (MODIS-derived SEE) and 0.53 (EBsoil-derived SEE) are obtained as opposed to 0.30 and 0.25 obtained in the linear mode. For the Terra overpass, correlations of 0.75 (MODIS-derived SEE) and 0.74 (EBsoil-derived SEE) are obtained as opposed to 0.69 and 0.70 obtained in the linear mode. The slope of the linear regression is increased from 0.16 to 0.66 (MODIS-derived SEE) and from 0.10 to 0.41 (EBsoil-derived SEE) – for the Aqua overpass. Coherent results are obtained for the Terra overpass: slope is increased from 0.45 to 0.70 (MODIS-derived SEE) and from 0.50 to 0.56 (EBsoil-derived SEE).

DISPATCH-E has a strong potential when applied to areas representative of either fully dry or fully wet conditions (desert areas, temperate areas or rainfed areas).

A better representation of SEE could potentially be achieved by taking into account the two separate temperature-based models in future studies. Additionally, the SEE modeling based on the nonlinear SM model could be integrated into existing evapotranspiration models, which are based on a combination of thermal and microwave data.

From a downscaling point of view, disaggregation is a way of detecting a bias in the low resolution SM data, which is not easily distinguishable without taking into

622 account the spatial heterogeneity within the scene. Future studies could try and correct
623 the bias in SMOS data and recalibrate the SM-based SEE model taking into
624 consideration new corrected data. Any improvement in the SEE model could lead to
625 refinements of the SEE estimates and of the associated HR SM products.

626

627 **Appendix A**

628 The EBsoil model used in the computation of $T_{s,dry}$ and $T_{s,wet}$ - corresponding to r_{ss}
629 equal to infinity and 0 respectively - iterates on T_s until thermal equilibrium is
630 established:

631 $$Rn_s - G = H_s + LE_s \quad (6)$$

632 Where Rn_s (Wm^{-2}) is the soil net radiation, G (Wm^{-2}) the ground heat flux, H_s (Wm^{-2})
633 the soil sensible heat flux and LE_s (Wm^{-2}) the soil latent heat flux.

634 The soil net radiation is estimated as:

635 $$Rn_s = (1- \alpha_s) * Rg + \varepsilon_s * (Ra - \sigma T_s^4) \quad (7)$$

636 Where α_s(unitless) is the soil albedo, Rg (Wm^{-2}) the incident solar radiation at short
637 wavelengths, ε_s (untiless) the soil emissivity, Ra (Wm^{-2}) the incident thermal radiation
638 at large wavelengths and σ ($Wm^{-2}K^{-4}$) the Stefan-Boltzmann constant.
639 The ground heat flux is approximated as:

640 $$G = 0.32Rn_s \quad (8)$$

641 The soil sensible heat flux is given by:

642 $$H_s = \rho * C_p * (T_s - T_a)/r_{ah} \quad (9)$$

643 with ρ ($kg*m^{-3}$) being the air density, C_p ($J*kg^{-}1*K^{-1}$) the specific heat of air at
644 constant temperature and r_{ah} ($s*m^{-1}$) the aerodynamic resistance to heat transfer.

The soil latent heat flux is calculated as:

$$LE_s = \rho * C_p * [e_{sat}(Ts) - e_{sat}(T_a)] / [\gamma * (r_{ss} + r_{ah})] \quad (10)$$

with $e_{sat}(Ts)$ (Pa) being the saturated vapor pressure at soil temperature, $e_{sat}(Ta)$ (Pa) the saturated vapor pressure at air temperature and γ (unitless) the psychrometric constant.

The aerodynamic resistance in equations 9 and 10 is estimated using the Richardson formulation:

$$r_{ah} = r_{ah0}/(1+R_i)^{\eta} \quad (11)$$

with R_i (untiless) being the Richardson number, η (unitless) a coefficient set to 0.75 in unstable conditions ($T_s > T_a$) and to 2 in stable conditions ($T_s < T_a$) and r_{ah0} the aerodynamic resistance which neglects natural convection, expressed as:

$$r_{ah0} = \log^2(Z_r/Z_{0m})/(k^2 u_a) \quad (12)$$

Z_r (m) is the reference height at which the wind speed was measured, Z_{0m} (m) the roughness length for momentum transfer over bare soil, k (unitless) the von Kármán constant, and u_a ($m*s^{-1}$) the wind speed.

The Richardson number reflects the influence of the atmospheric stability on the flux-gradient relationship in the surface layer and is given by:

$$R_i = 5*g*Z_r*(T_s-T_a)/(T_a*u_a^2) \quad (13)$$

With g ($m*s^{-2}$) being the gravitational constant.

References

Chapter 6

CONCLUSIONS AND PERSPECTIVES

In the setting of nowadays' increasing awareness of climate change and limited water ressources, many research studies focus on developing models to estimate different hydrological variables that are essential to a better management of Earth's resources. A key variable in the hydrological cycle, especially in semi–arid areas, is evapotranspiration (ET), with its two main components: soil evaporation (E) and plant transpiration (T). Monitoring E at multiple scales is crucial to many agronomic, hydrologic, meteorological and climatic applications. Remote sensing is in this respect a promising approach to estimate E at multiple scales, on a global basis. Two E–related variables that can be retrieved through remote sensing are essential in approaches based on remote sensing: the land surface temperature (LST) - derived from thermal sensors such as MODIS and ASTER, and surface soil moisture (SM) - derived from microwave sensors such as SMOS.

This PhD thesis focused on improving the spatio–temporal representation of soil evaporation by deriving a multi–scale modeling technique based on a synergy between existing evaporation models and remote sensing data. Amongst all the possible models that can be used to derive soil evaporation, ranging from mechanistic and phenomenological models to remote sensing–based approaches, one method was of particular interest: the so-called "contextual" approach. Contextual approaches are derived from thermal remote–sensing and they estimate evaporation as a function of LST. They can also be referred to as temperature–based approaches. Their appeal lies within the possibility to be applied to different areas around the world, thanks to the availability of input data. However, the main constraint of contextual methods lies within determining the physical wet and dry boundaries of LST, which may not be present within the scene at the thermal sensor resolution.

In this context, one major objective of this PhD was to introduce a new way of deriving the wet and dry boundaries independently of remote sensing data and implicitly, of the observation resolution of the thermal sensor. In practice, a soil energy balance model (EBsoil) forced by meteorological data was introduced. The model was then tested over three different semi–arid sites: Morocco, Mexico and Spain.

As a first step, EBsoil was implemented within a contextual ET model (SEB–1S). ET esti-

mates derived either from remote sensing data solely or by using EBsoil were intercompared at multiple resolutions with a reference ET over the Mexican site. Results indicated that integrating EBsoil in SEB–1S improves ET estimates at a medium ($\sim$ 1 km) resolution. This provided a spatial characterization of ET from LST data. Note that these results could be applied to soil evaporation through the soil evaporative efficiency (SEE), which is defined as the ratio of actual to potential evaporation.

One limitation concerning EBsoil is related to wind speed, which intervenes in the computation of the aerodynamic resistance. More specifically, a sensitivity study showed that when wind speed is inferior to 2 $m{\cdot}s^{-1}$, the extreme soil temperatures corresponding to wet and dry conditions are impacted; this implies that the wet and dry boundaries that are then used in estimating ET are also impacted. Future studies should keep this limitation in mind. Another important limitation to bear in mind regarding the application of EBsoil is the availability and accuracy of the atmospheric forcing at multiple resolutions.

The next step consisted in focusing on the temporal dynamics of SEE. Since SEE can be retrieved from remote sensing data, either using LST or SM, there is a particular interest in focusing on this variable, due to the synergy between the two modeling approaches. Therefore, efforts have been made to characterize the temporal dynamics of SEE by using the global modeling approach. There is a novelty aspect related to this approach, since this technique has been used for a very small number of environmental contexts, with even fewer models obtained that are able to reproduce the original dynamics. For the first time ever, a four dimensional chaotic model was obtained for the daily cycle of SEE, able to reproduce the dynamics of the variable with a good approximation in the phase domain. However, limitations regarding the model representation in the time domain arise: the obtained time series does not reproduce daily cycles and exhibits faster oscillations than the original series, which means multiple cycles per day. Even though the difference is expected, due to the chaotic regime of the dynamics, the model cannot be considered satisfying in the time domain, since the reproduced series presents faster oscillations than the expected daily signal. The "nominal" behavior of the SEE (one smooth cycle, with a maximum value registered in the morning) can be strongly affected by cloud presence, which can induce oscillations in the daily cycle. These oscillations are then taken into account in the global modeling approach, which may explain the faster oscillations of the reconstructed SEE series. Moreover, in the modeling, values of the SEE during night–time were taken into account. This implies values for a non–defined SEE (since the evaporative demand is equal to 0 at night), which could lead to singularities in the model. Nevertheless, one should keep in mind that the global model obtained is autonomous, so no forcing was used, which means that no information regarding meteorological data (cloud presence) are taken into consideration to synchronize the model to the daily cycle. Future analyses could take the meteorological forcing into account when applying the global modeling technique. Remote sensing data can be used to derive SEE series as input to the approach. Coupling remote sensing data and meteorological information could provide a way of modeling the SEE dynamics from a spatially distributed point of view. The global modeling approach may be used to model the dynamics of the surface cover. The model does not allow for long–term forecasting but it nonetheless allows the anticipation of complex behaviors. It also provides a good representation of the dynamical behavior in the phase space. Although it has no direct applicability (because of its low predictability), this result is very encouraging because it proves that the global modeling approach is viable, since a model could be obtained from a variable that doesn't offer a good observability of the dynamics. A future perspective could be choosing another variable or using multiple variables that could offer a better observability of the dynamics, and thus better results in terms of predictability could be obtained.

The two modeling approaches used in deriving SEE - as a function of LST or SM - were

regrouped at the third step. DISPATCH is a disaggregation algorithm that provides high resolution SM data. It is the only link between the two approaches of deriving SEE, because it uses both a LST– and SM–based model in order to downscale SM data by combining microwave and optical data. The EBsoil model developed at the first step was integrated within DISPATCH, thus developing a new algorithm, DISPATCH–E. The influence of implementing EBsoil is assessed in terms of SM products over the Spanish site. It is an indirect validation of the SEE, because no available direct measurements of this variable are available to compare against. Nevertheless, this indirect validation in terms of SM is an indication of the impact behind the SEE modeling technique. Implementing EBsoil significantly improves the downscaled SM. It proves to be a robust way of deriving downscaled SM products, as well as being an independent way of assessing uncertainty in the satellite data.

Various research perspectives in connection with the work carried in this PhD thesis can be considered for the future:

1. Further improvements to DISPATCH–E
2. Coupling DISPATCH–E with ET models including soil/vegetation components

1. Further improvements to DISPATCH–E

Regarding the possible developments of DISPATCH–E, meteorological data provided by an interpolation model (namely SAFRAN) were used to force EBsoil and consequently in the derivation of high resolution SM products. Other meteorological data - issued from weather stations or from ECMWF reanalysis could be used as an input to the soil energy balance model. This could provide an independent assessement of the robustness of the methodology, as well as extending its applications to a global scale. One should keep in mind though the uncertainty of the meteorological data. For example, SAFRAN data in the north–east of Spain (where DISPATCH–E was applied) presents a mean RMSE of about 1.5 °C in temperature and of 7 % in relative air humidity. There is a 1.3 $\mathrm{m{\cdot}s^{-1}}$ RMSE in the wind speed estimations, and 114 $\mathrm{W{\cdot}m^{-2}}$ RMSE in visible radiation (*Quintana-Seguí et al.*, 2008). This uncertainty in the data is expected to influence the output of the EBsoil model, and therefore, the SEE estimates. Using large scale data such as from ECMWF reanalysis is very attractive from the point of view of a global scale application, but could present large uncertainties in the meteorological data, and subsequently, on the SEE estimates. High resolution meteorological data should be used in order to account for heterogeneities, which can be obtained by interpolating *in situ* measurements at a range of spatial resolutions, or by atmospheric surface analysis at the atmospheric model grid resolution.

Moreover, the application of the original DISPATCH algorithm is limited by the contextual nature of the SEE modeling as a function of LST (derived solely from satellite imagery). In particular, DISPATCH works well in semi–arid areas, provided that the dry and wet boundaries of the LST can be observed. Therefore, its application is restricted in desert and rainfed areas, where one of the two boundaries cannot be captured using medium resolution ($\sim$ 1 km) data. In this context, the new DISPATCH–E algorithm could be implemented over these areas since the limitations regarding the wet and dry boundaries do not exist when using an energy balance model. For example, the SMELLS (Soil Moisture for dEsert Locust earLy Survey, `http://smells.isardsat.com/`) project is an Innovators-III project funded by ESA to develop innovative Earth Observation products and services in response to authoritative requirements from end–user organisations. It investigates the use of soil moisture to preventive management of desert locust in Africa. Desert locust have been linked to a variety of plagues, plagues which might be forecasted by using high resolution soil moisture products. The project aims at implementing an innovative approach to combine Sentinel–1 SAR data with thermal–disaggregated SMOS soil moisture (derived from DISPATCH) to derive a high spatio–temporal resolution SM

product. This is in order to provide a new tool for decision makers in the desert locust preventive management system. Since the study area covers the western and northern Africa, which is mostly a desert area, there are limitations regarding the DISPATCH SM product. This is explained by the wet boundary of the LST - fractional vegetation cover space retrieved from satellite data being overestimated. DISPATCH–E could provide better disaggregated SM estimates, since it estimates the boundaries independently of the remote sensing data.

Moreover, because DISPATCH–E estimates extreme temperatures independently of MODIS data, one could identify the sensitivity of LST to SM by quantifying DISPATCH–E's uncertainty in given meteo/climatic conditions. This is essential in view of assimilating disaggregated SM data in land surface models (*Merlin et al.*, 2006b; *Bandara et al.*, 2015; *Dumedah et al.*, 2015). Any advances in the LST–based SEE models would foster further developments of DISPATCH–E and associated improvements in the refined soil moisture products.

In addition, another aspect that one could focus on is the impact of soil texture on SEE. By trying and characterizing the hydrodynamical soil properties, one could study the impact of soil variability on the DISPATCH–E output. The recent model developed in *Merlin et al.* (2016) introduces a SEE model that is tightly linked to soil texture. The model was tested for regionalization capabilities by using a parameterization depending on clay and sand percentages. Moreover, the model could be calibrated by using thermal and microwave remote sensing data. This model could be included at low resolution in DISPATCH–E: either using existing maps of soil texture or by inverting the clay and sand percentages derived from observed (SMOS and MODIS) data. Because a significant relationship was found between the strongly non–linear behavior of SEE as a function of SM and soil texture, a possible approach to deriving soil texture from multi–sensor/multi–spectral remote sensing could be developed.

2. Coupling DISPATCH–E with ET models including soil/vegetation components

By combining multiple (remote sensing and meteorological) data, one could improve the multi–scale spatial representation of ET. In particular, DISPATCH–E has the potential to be applied to other satellite data: C–band radiometers (AMSR–E, AMSR–2), scatterometers (ASCAT) and thermal sensors such as LANDSAT. Even though one might argue that disaggregated data are less reliable than the observed data at high resolution, taking into account nowadays limitations regarding the coarse resolution of SM products, disaggregation allows to extend applications of currently available remote sensing data. Also, multi–spectral based approaches allow testing the complementarity of multi–source data before these data become available at high resolution.

One main aspect to be taken into account is the partitiong of ET into E and T. Remotely–sensed estimates of LST–derived SEE and microwave–derived SM data would be crucial in a better constraining of the partitioning of LST–derived ET into soil evaporation and plant transpiration (*Merlin et al.*, 2014), which is key in enhacing water resource management over semi–arid areas. In theory, combining LST and SM data should allow the decoupling over mixed pixels through DISPATCH–E. It could also allow for the calibration of semi–empirical parameters of E through remote sensing means. NDVI data from visible and near infrared sensors have been extensively used to estimate ET on surfaces covered by vegetation, not specifically on bare or partially covered soils (*Allen et al.*, 1998). Methods based on thermal infrared do not allow for a complete partition between E and T over mixed pixels by using only visible/near infrared and LST data. The reason is because there is an equifinality between E and T rates and the LST integrated at the pixel scale. One solution to solve this equifinality problem would be to introduce further data regarding the SEE derived from microwave data. Approaches combining thermal infrared and microwave data have a great potential to constrain the E process (*Chanzy et al.*, 1995; *Li et al.*, 2006). Combining LST and SM data would allow i) partitioning ET over mixed

pixels and ii) calibrating semi–empirical parameters from remote sensing data. The derived E/T products could be validated with *in situ*–derived measurements. In particular, eddy covariance and/or lysimeter measurements of ET could be partitioned into E and T measurements using the statistical method based on marginal distribution sampling - MDS (*Béziat et al.*, 2009). Transpiration measurements could also be taken separately, using a sapflow system, and further used to validated the MDS method. Moreover, one could imagine a way of deriving *in situ* measurements of the SEE. For that, one would need measurements of potential evaporation. This can be measured indirectly, using climatic factors. Usually, a value can be calculated at a nearby climate station on a reference surface, which can then be converted to potential evaporation by multiplying it with a surface coefficient. Once both evaporation and potential evaporation measurements are available, one could derive the SEE as the ratio of the two. This could help directly validating the SEE products retrieved through remote sensing, at a range of scales.

DISPATCH–E therefore has a real potential of decoupling ET into its two main components. It could provide a measure of the improvement of SEE modeling on SM estimates and reciprocally, the improvement in SM products on the representation of multi–scale E. One could imagine the possibility of deriving E products through DISPATCH–E, which would be a first, since currently there is no remote sensing algorithm dedicated to monitoring this flux.

CONCLUSIONS ET PERSPECTIVES

Dans le cadre actuel de la sensibilisation aux changements climatiques et de la limitation des ressources en eau, de nombreuses études de recherche se concentrent sur le développement de modèles pour estimer les différentes variables hydrologiques qui sont essentielles à une meilleure gestion des ressources de la Terre. Une variable clé dans le cycle hydrologique, en particulier dans les zones semi–arides, est l'évapotranspiration (ET), avec ses deux composantes principales: l'évaporation du sol (E) et la transpiration des plantes (T). La surveillance de l'E à des échelles multiples est cruciale pour beaucoup d'applications agronomiques, hydrologiques, météorologiques et climatiques. La télédétection est à cet égard une approche prometteuse pour estimer l'E à des échelles multiples, sur une base mondiale. Deux variables liées à l'E sont essentielles dans les approches fondées sur la télédétection: la température de surface (Land Surface Temperature - LST) - dérivée de capteurs thermiques tels que MODIS et ASTER, et l'humidité du sol en surface (Soil Moisture - SM) - dérivée de des capteurs micro–ondes, tels que SMOS.

Cette thèse a porté sur l'amélioration de la représentation spatio–temporelle de l'évaporation du sol en dérivant une technique de modélisation multi–échelle basée sur une synergie entre les modèles d'évaporation existants et les données de télédétection. Parmi tous les modèles possibles qui peuvent être utilisés pour calculer l'évaporation du sol, qu'ils soient mécanistes, phénoménologiques ou basés sur la télédétection, une méthode dénote d'un intérêt tout particulier: l'approche dite "contextuelle". Les approches contextuelles sont dérivées de la télédétection thermique et elles estiment l'évaporation en fonction de la LST. Elles peuvent également être appelées approches basées sur la température. Leur intérêt se trouve dans la possibilité d'être appliquées à différentes régions du monde, grâce à la disponibilité des données d'entrée. Cependant, leur principale contrainte se situe dans la détermination des limites physiques humides et sèches de la LST, qui peuvent ne pas être présentes sur la scène à la résolution du capteur thermique.

Dans ce contexte, l'un des principaux objectifs de cette thèse était d'introduire une nouvelle façon de dériver les limites humides et sèches indépendamment des données de télédétection et, implicitement, de la résolution d'observation du capteur thermique. Dans la pratique, un modèle de bilan d'énergie du sol (EBsoil) contraint par des données météorologiques a été introduit. Le modèle a ensuite été testé sur trois sites semi–arides différentes, situés au Maroc, au Mexique et en Espagne.

Dans un premier temps, EBsoil a été implémenté au sein d'un modèle contextuelle de l'ET (SEB–1S). Les estimations de l'ET ont été obtenues soit uniquement à partir des données de

télédétection, soit en utilisant EBsoil pour ensuite être comparées à de résolutions multiples avec une ET de référence sur le site mexicain. Les résultats indiquent que l'intégration du EBsoil dans SEB–1S améliore les estimations ET à une résolution moyenne ($\sim$ 1 km). Cela a fourni une caractérisation spatiale de l'ET à partir des données LST. On notera que ces résultats pourraient être appliqués à l'évaporation du sol grâce à l'efficacité évaporative du sol (Soil Evaporative Efficiency - SEE), qui est définie comme étant le rapport entre l'évaporation réelle et l'évaporation potentielle.

Une limitation concernant EBsoil est liée à la vitesse du vent, qui intervient dans le calcul de la résistance aérodynamique. Plus précisément, une étude de sensibilité a montré que lorsque la vitesse du vent est inférieure à 2 $m{\cdot}s^{-1}$, les températures du sol extrêmes correspondant aux conditions sèches et humides sont impactées; cela implique que les limites humides et sèches qui sont ensuite utilisées dans l'estimation de l'ET sont également impactées. Pour de futures études, il est nécessaire de garder cette limitation à l'esprit. Une autre limitation importante à garder en tête en ce qui concerne l'application du EBsoil est la disponibilité et la précision du forçage atmosphérique à des résolutions multiples.

L'étape suivante s'est concentrée sur la dynamique temporelle de la SEE. Comme la SEE peut être récupérée à partir des données de télédétection, soit en utilisant la LST, soit la SM, il y a un intérêt particulier à se concentrer sur cette variable, en raison de la synergie entre les deux approches de modélisation. Par conséquent, une étude de caractérisation de la dynamique temporelle de la SEE en utilisant l'approche de modélisation globale a été menée. Cette technique n'a été appliquée à ce jour qu'à un nombre très restreint de contextes environnementaux, avec encore moins de modèles obtenus qui sont en mesure de reproduire la dynamique d'origine. C'est pourquoi cette approche est nouvelle à son genre. Pour la première fois, un modèle chaotique de dimension quatre a été obtenu pour le cycle journalier de la SEE, capable de reproduire la dynamique de la variable avec une bonne approximation dans l'espace de phase. Cependant, on observe des limitations relatives à la représentation du modèle dans le domaine temporel. En effet, la série temporelle obtenue ne reproduit pas les cycles journaliers et elle montre des oscillations plus rapides que la série originale, ce qui signifie des cycles journaliers multiples. Cette différence est prévisible, du fait du régime chaotique de la dynamique. Cependant, comme la série reproduite présente des oscillations plus rapides que le signal journalier prévu, le modèle ne peut pas être considéré comme satisfaisant dans le domaine temporel. Le "comportement" nominal de la SEE (un cycle lisse, avec une valeur maximale enregistrée le matin) peut être fortement affectée par la présence de nuages, ce qui peut induire des oscillations dans le cycle quotidien. Ces oscillations sont alors prises en compte dans l'approche de modélisation globale, ce qui peut expliquer les oscillations plus élevées de la série SEE reconstruite. En outre, dans la modélisation, les valeurs de la SEE au cours de la nuit ont été prises en compte. Cela implique que l'on obtient des valeurs alors que la SEE n'est pas définie (car la demande d'évaporation est égal à 0 dans la nuit), ce qui pourrait conduire à des singularités dans le modèle. Néanmoins, il faut garder à l'esprit que le modèle global obtenu est autonome, ce qui signifie qu'aucun forçage n'a été utilisé. Aucune information concernant les données météorologiques (présence de nuages) n'est prise en considération pour synchroniser le modèle au cycle journalier. Les analyses futures pourraient prendre le forçage météorologique en compte lors de l'application de la technique de modélisation globale. Les données de télédétection peuvent être utilisées pour obtenir des séries de la SEE comme entrée à l'approche. Le couplage des données de télédétection et des informations météorologiques pourrait fournir un moyen de modélisation de la dynamique de la SEE à partir d'un point de vue spatialement distribué. L'approche de modélisation globale peut être utilisée pour modéliser la dynamique du couvert de surface. Le modèle ne permet pas des prévisions à long terme, mais il permet néanmoins l'anticipation des comportements complexes. Il fournit également une bonne représentation du comportement dynamique dans l'espace de

phase. Bien qu'il n'a pas d'applicabilité directe (en raison de sa faible prévisibilité), ce résultat est très encourageant car il prouve que l'approche de modélisation globale est viable, car un modèle a pu être obtenu à partir d'une variable qui ne propose pas une bonne observabilité de la dynamique. Une perspective future pourrait être de choisir une autre variable ou d'utiliser plusieurs variables qui pourraient offrir une meilleure observabilité de la dynamique, et donc des meilleurs résultats en termes de prévisibilité pourraient être obtenus.

Les deux approches de modélisation utilisées pour dériver la SEE - en fonction de la LST ou de la SM - ont été regroupées lors de la troisième étape. DISPATCH est un algorithme de désagrégation qui fournit des données SM à haute résolution. Il est le seul lien entre les deux approches de dérivation de la SEE, car il utilise à la fois un modèle basé sur la LST et un modèle basé sur la SM pour désagréger des données SM en combinant des données micro–ondes et des données optiques. Le modèle EBsoil développé à la première étape a été intégré au sein de DISPATCH, développant ainsi un nouvel algorithme, DISPATCH–E. L'influence de l'implémentation du EBsoil est évaluée en termes de produits SM sur le site espagnol. Il s'agit d'une validation indirecte de la SEE, car aucune mesure directe de cette variable n'est disponible pour la comparaison. Néanmoins, cette validation indirecte en termes de SM est une indication de l'impact derrière la technique de modélisation de la SEE. L'intégration du EBsoil améliore significativement la SM désagrégée. Il se révèle être un moyen robuste de dériver des produits de SM désagrégés, tout en étant une manière indépendante d'évaluation de l'incertitude dans les données satellitaires.

Diverses perspectives de recherche en relation avec le travail effectué dans cette thèse peuvent être envisagées pour l'avenir:

1. D'autres améliorations à DISPATCH–E
2. Un couplage entre DISPATCH–E et les modèles de l'ET incluant les composants sol/végétation

1. D'autres améliorations à DISPATCH–E

En ce qui concerne les développements possibles de DISPATCH–E, les données météorologiques fournies par un modèle d'interpolation (à savoir SAFRAN) ont été utilisés pour forcer EBsoil et par conséquent dans la dérivation de produits SM à haute résolution. Autres données météorologiques - fournies par des stations météo ou issues de la ré–analyse de l'ECMWF pourraient être utilisées comme une entrée au modèle du bilan d'énergie du sol. Cela pourrait fournir une évaluation indépendante de la robustesse de la méthodologie, ainsi que l'extension de ses applications à l'échelle mondiale. Néanmoins, il faut garder à l'esprit l'incertitude des données météorologiques. Par exemple, les données SAFRAN dans le nord–est de l'Espagne (où DISPATCH–E a été appliqué) présente une RMSE moyenne d'environ 1.5 °C de la température et de 7 % en humidité relative de l'air. LA RMSE est de 1.3 $m \cdot s^{-1}$ dans les estimations de la vitesse du vent, et de 114 $W \cdot m^{-2}$ pour le rayonnement visible (*Quintana-Seguí et al.*, 2008). Il est facilement prévisible que cette incertitude dans les données va influencer la sortie du modèle EBsoil et, par conséquent, les estimations de la SEE. L'utilisation de données à grande échelle tels que les données issues de la ré–analyse de l'ECMWF est très attrayant du point de vue d'une application à l'échelle mondiale, mais ces données pourraient présenter de grandes incertitudes, et par la suite, pourraient influencer les estimations de la SEE. Des données météorologiques à haute résolution doivent être utilisées afin de tenir compte des hétérogénéités, qui peuvent être obtenus par une interpolation de mesures *in situ* à une variété de résolutions spatiales, ou par l'analyse de la surface atmosphérique à la résolution de la grille atmosphérique du modèle.

En outre, l'application de l'algorithme original DISPATCH est limitée par la nature contextuelle de la modélisation de la SEE en fonction de la LST (provenant exclusivement de l'imagerie par satellite). En particulier, DISPATCH fonctionne bien dans les zones semi–arides, à

condition que les limites sèches et humides de la LST puissent être observées. Ce qui signifie que son application est limitée dans le désert et les zones pluviales, où l'une des deux limites ne peut être capturée en utilisant des données à une résolution moyenne ($\sim$ 1 km). Dans ce contexte, le nouvel algorithme DISPATCH–E pourrait être implémenté dans ces zones puisque les limitations concernant les limites humides et sèches n'existent pas lors de l'utilisation d'un modèle de bilan d'énergie. Par exemple, le projet SMELLS (Soil Moisture for dEsert Locust earLy Survey, `http://smells.isardsat.com/`) est un projet Innovateurs-III financé par l'ESA pour développer des produits et des services innovants d'observation de la Terre en réponse aux exigences des organisations utilisatrices. Il étudie l'utilisation de l'humidité du sol afin de prévenir les déplacements du criquet pèlerin en Afrique. Le criquet pèlerin est lié à plusieurs variétés des pestes. Elles pourraient ainsi être anticipées en utilisant des produits de l'humidité du sol à haute résolution. Le projet vise à développer une approche innovante qui combine des données SAR de Sentinel–1 avec des données désagrégées de l'humidité du sol SMOS (dérivées de DISPATCH) pour dériver un produit SM à haute résolution spatio–temporelle. Ceci dans le but de fournir un nouvel outil au système de prévention du criquet pèlerin. Comme la zone d'étude couvre l'Afrique de l'Ouest et du Nord, qui sont principalement des zones désertiques, l'utilisation du produit DISPATCH SM s'en trouve limité. Ceci s'explique par la limite humide de l'espace LST - couvert végétal estimée à partir des données satellitaires étant surestimée. DISPATCH–E pourrait fournir de meilleures estimations SM désagrégées, car il estime les limites indépendamment des données de télédétection.

De plus, parce que DISPATCH–E estime les températures extrêmes indépendamment des données MODIS, on pourrait déterminer la sensibilité de la LST à la SM en quantifiant l'incertitude de DISPATCH–E dans des conditions météorologiques/climatiques données. Cela est essentiel en raison de l'assimilation des données SM désagrégées dans les modèles de surface (*Merlin et al.*, 2006b; *Bandara et al.*, 2015; *Dumedah et al.*, 2015). Tout progrès dans les modèles de la SEE basés sur la LST favoriserait d'autres développements de DISPATCH–E et des améliorations associées dans les produits d'humidité du sol raffinés.

Enfin, un autre aspect sur lequel il est possible de se concentrer est l'impact de la texture du sol sur la SEE. Grâce à une caractérisation des propriétés hydrodynamiques du sol, il est possible d'étudier l'impact de la variabilité du sol sur les sorties de DISPATCH–E. Le modèle récemment développé dans *Merlin et al.* (2016) présente un modèle de la SEE qui est étroitement liée à la texture du sol. Il a été testé pour les capacités de régionalisation en utilisant une paramétrisation en fonction de pourcentages d'argile et de sable. En outre, le modèle pourrait être étalonné à l'aide des données de télédétection thermiques et micro–ondes. Ce modèle pourrait être inclus à basse résolution dans DISPATCH–E: soit en utilisant les cartes existantes de la texture du sol ou en inversant les pourcentages d'argile et de sable provenant de données observées (SMOS et MODIS). Grâce à la relation significative trouvée entre le comportement fortement non linéaire de la SEE en fonction de la SM et la texture du sol, une approche consistant à dériver la texture du sol à partir des données de télédétection multi–capteurs/multi–spectrale pourrait être développée.

2. Un couplage entre DISPATCH–E et les modèles de l'ET incluant les composants sol/végétation

En combinant les données multiples (météorologiques et de télédétection), il est possible d'améliorer la représentation spatiale de l'ET à des échelles multiples. En particulier, DISPATCH–E a le potentiel d'être appliqué à d'autres données satellitaires: radiomètres en bande C (AMSR–E, AMSR–2), diffusiomètres (ASCAT) et capteurs thermiques tels que LANDSAT. Même si l'on peut soutenir que les données désagrégées sont moins fiables que les données observées à haute résolution, en tenant compte des limitations actuelles en ce qui concerne la résolution grossière des produits SM, la désagrégation permet d'étendre les applications des données de télédétec-

tion actuellement disponibles. En effet, des approches multi–spectrales permettent de tester la complémentarité des données multi–source avant que ces données soient disponibles à haute résolution.

Un élément principal à prendre en compte est la partition de l'ET dans E et T. Des estimations à partir de la télédétection de la SEE basée sur la LST et sur la SM micro–onde seraient cruciales pour mieux contraindre la partition de l'ET dérivée de la LST dans l'évaporation du sol et la transpiration des plantes (*Merlin et al.*, 2014), qui est la clé pour une meilleure gestion des ressources en eau sur les zones semi–arides. En théorie, la combinaison des données LST et SM devraient permettre le découplage sur les pixels mixtes à travers DISPATCH–E. Cela pourrait également permettre la calibration des paramètres semi–empiriques de l'E par des moyens de télédétection. Les données NDVI des capteurs visible et proche infrarouge ont été largement utilisées pour estimer l'ET sur les surfaces couvertes par la végétation, et non uniquement pour des sols nus ou partiellement couverts (*Allen et al.*, 1998). Les méthodes basées sur l'infrarouge thermique ne permettent pas une partition complète entre E et T sur les pixels mixtes lorsqu'on utilise seulement des données visibles/proche–infrarouge et la LST. En effet, apparaît dans ce cas une équifinalité entre les taux d'E et de T et la LST intégrée à l'échelle du pixel. Une solution pour résoudre ce problème d'équifinalité serait d'introduire des données supplémentaires concernant la SEE dérivée des données à micro–ondes. Des approches combinant les données infrarouges thermiques et micro–ondes ont un grand potentiel pour limiter le processus d'E (*Chanzy et al.*, 1995; *Li et al.*, 2006). La combinaison de données LST et SM permettraient i) le partitionnement ET sur les pixels mixtes et ii) l'étalonnage de paramètres semi–empiriques à partir des données de télédétection. Les produits dérivés d'E/T pourraient être validés avec des mesures *in situ* dérivées. En particulier, l'ET issue de l'eddy covariance et/ou de mesures lysimétriques pourraient être partitionné en mesures d'E et de T en utilisant la méthode statistique basée sur l'échantillonnage de distribution marginale (marginal distribution sampling) - MDS (*Béziat et al.*, 2009). Les mesures de transpiration pourraient également être prises séparément, à l'aide d'un système de sapflow, et ainsi être utilisées pour valider la méthode MDS. De plus, on pourrait imaginer une façon de dériver des mesures *in situ* de la SEE. Pour cela, il faudrait des mesures de l'évaporation potentielle. Cela peut être mesurée indirectement, en utilisant des facteurs climatiques. De manière générale, une valeur peut être calculée à partir d'une station climatique de proximité, sur une surface de référence, qui peut ensuite être convertie en un potentiel d'évaporation en la multipliant par un coefficient de surface. Une fois les mesures de l'évaporation et de l'évaporation potentielle disponibles, il est possible de dériver la SEE comme le rapport entre les deux. Cela pourrait aider à valider directement les produits SEE récupérés grâce à la télédétection, à une gamme d'échelles.

DISPATCH–E a donc un réel potentiel de découplage de l'ET dans ses deux principales composantes. Il pourrait fournir une mesure de l'amélioration de la modélisation de la SEE sur les estimations de la SM et réciproquement, l'amélioration des produits SM sur la représentation multi–échelle de l'E. On pourrait imaginer la possibilité d'en tirer des produits d'E par DISPATCH–E, qui serait une première, car il n'y a actuellement aucun algorithme de télédétection dédié au suivi de ce flux.

Annex

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2015WR018233

Modeling soil evaporation efficiency in a range of soil and atmospheric conditions using a meta-analysis approach

O. Merlin[1,2], V. G. Stefan[1], A. Amazirh[2], A. Chanzy[3], E. Ceschia[1], S. Er-Raki[2], P. Gentine[4], T. Tallec[1], J. Ezzahar[2], S. Bircher[1], J. Beringer[5], and S. Khabba[2]

[1]CESBIO, Université de Toulouse, CNES, CNRS, IRD, UPS, Toulouse, France, [2]Faculté des Sciences et Techniques, University of Cadi Ayyad, Marrakech, Morocco, [3]EMMAH, INRA, Université d'Avignon et des Pays de Vaucluse, 84000, Avignon, France, [4]Department of Earth and Environmental Engineering, Columbia University, New York City, New York, USA, [5]School of Earth and Environment, University of Western Australia, Perth, Australia

Abstract A meta-analysis data-driven approach is developed to represent the soil evaporative efficiency (SEE) defined as the ratio of actual to potential soil evaporation. The new model is tested across a bare soil database composed of more than 30 sites around the world, a clay fraction range of 0.02–0.56, a sand fraction range of 0.05–0.92, and about 30,000 acquisition times. SEE is modeled using a soil resistance (r_{ss}) formulation based on surface soil moisture (θ) and two resistance parameters $r_{ss,ref}$ and $\theta_{efolding}$. The data-driven approach aims to express both parameters as a function of observable data including meteorological forcing, cut-off soil moisture value $\theta_{1/2}$ at which SEE=0.5, and first derivative of SEE at $\theta_{1/2}$, named $\Delta\theta_{1/2}^{-1}$. An analytical relationship between $(r_{ss,ref};\theta_{efolding})$ and $(\theta_{1/2};\Delta\theta_{1/2}^{-1})$ is first built by running a soil energy balance model for two extreme conditions with $r_{ss}=0$ and $r_{ss}\sim\infty$ using meteorological forcing solely, and by approaching the middle point from the two (wet and dry) reference points. Two different methods are then investigated to estimate the pair $(\theta_{1/2};\Delta\theta_{1/2}^{-1})$ either from the time series of SEE and θ observations for a given site, or using the soil texture information for all sites. The first method is based on an algorithm specifically designed to accomodate for strongly nonlinear SEE(θ) relationships and potentially large random deviations of observed SEE from the mean observed SEE(θ). The second method parameterizes $\theta_{1/2}$ as a multi-linear regression of clay and sand percentages, and sets $\Delta\theta_{1/2}^{-1}$ to a constant mean value for all sites. The new model significantly outperformed the evaporation modules of ISBA (Interaction Sol-Biosphère-Atmosphère), H-TESSEL (Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land), and CLM (Community Land Model). It has potential for integration in various land-surface schemes, and real calibration capabilities using combined thermal and microwave remote sensing data.

Key Points:
- The new soil resistance model is based on soil moisture and two observable parameters
- Models are tested using a data set composed of more than 30 contrasted sites
- One resistance parameter is significantly correlated with both sand and clay fractions

Correspondence to:
O. Merlin,
olivier.merlin@cesbio.cnes.fr

Citation:
Merlin, O., et al. (2016), Modeling soil evaporation efficiency in a range of soil and atmospheric conditions using a meta-analysis approach, *Water Resour. Res.*, *52*, 3663–3684, doi:10.1002/2015WR018233.

Received 13 OCT 2015
Accepted 12 APR 2016
Accepted article online 20 APR 2016
Published online 14 MAY 2016

© 2016. American Geophysical Union. All Rights Reserved.

1. Introduction

Evaporation accounts for approximately 20–40% of the global evapotranspiration [*Lawrence et al.*, 2007; *Schlesinger and Jasechko*, 2014] and originates mainly (65%) from soils rather than surface waters [*Good et al.*, 2015]. As an important boundary condition between the soil and atmosphere, soil evaporation is explicitly represented in a range of agronomic, hydrological, meteorological and climate models at multiple scales: from the crop field [e.g., *Allen*, 2000], to the catchment [e.g., *Wood et al.*, 1992] and to the mesoscale [e.g., *Sakaguchi and Zeng*, 2009]. Accurate estimations of soil evaporation are notably needed to quantify the partitioning of evapotranspiration into soil evaporation and plant evaporation and transpiration [*Williams et al.*, 2004; *Kool et al.*, 2014]. Such partitioning is fundamental to monitor vegetation water uptake and stress [*Porporato et al.*, 2001; *Er-Raki et al.*, 2010] within an environment of scarse water resources, and to better understand land-atmosphere interactions [*Feddes et al.*, 2001; *Er-Raki et al.*, 2010]. This is especially true in sparsely vegetated areas such as arid and semi-arid regions, and agricultural fields during bare or partially covered soil periods.

The evaporation of unsaturated soils is a complex process due to the coupling of the energy and mass transfers at the soil-atmosphere interface, which involves liquid and vapor transport in the near-surface soil pores, incident solar energy for phase change, and vapor transfer across the boundary layer [*Philip and de Vries*, 1957; *Milly*, 1984; *Chanzy and Bruckler*, 1993; *Bittelli et al.*, 2008; *Smits et al.*, 2012; *Or et al.*, 2013]. The soil control on evaporation originates from two main processes: the difference in water vapor concentrations (Cw) between the evaporative surface and atmosphere, and the soil vapor diffusion when the

evaporative sources are located below the soil surface. The vapor diffusion depends on the depth, degree of saturation and temperature of the evaporative site. Comprehensive physical models as those based on Philip and de Vries equations can represent both temperature and water potential gradients as well as vapor diffusion. Such models are driven by standard climatic conditions but are, however, very sensitive to soil hydraulic properties (SHPs), initialization and bottom boundary conditions [*Chanzy et al.*, 2008]. These characteristics hamper the implementation of comprehensive models to represent large areas, as it requires numerous simulation units and the capability of characterizing soil parameters in every unit. The most common alternative is to use evaporation models related to the surface soil moisture. Considering the physical processes mentioned above, soil moisture is only a proxy of the soil quantities that control the evaporation rate. For instance, we have to consider all evaporation sites -which may be located at different levels in the soil- as well as the water potential/water content relationships. These properties are linked to the soil surface wetness but can also be governed by other factors such as the climatic demand or the soil surface structure. Moreover the thickness of the layer considered to characterize moisture has also an impact on the evaporation models' parameters.

There are numerous evaporation models that are based on soil moisture. They all try to represent the limitation of evaporation by soil moisture (water) and evaporative demand, generally using empirical or semi-empirical approaches [*Viterbo and Beljaars*, 1995; *Pitman*, 2003]. Historically, the evaporation module of the so-called Bucket model [*Manabe*, 1969; *Robock et al.*, 1995] has been:

$$\mathrm{SEE}=\theta/\theta_c \tag{1}$$

with SEE being the soil evaporative efficiency defined as the actual to potential soil evaporation ratio, θ (m^3 m^{-3}) the surface soil moisture, and θ_c (m^3 m^{-3}) a parameter equal to 0.75 times the soil moisture at field capacity. Since the development of the Bucket model, various attempts have been made to improve the above representation, notably by separating soil and vegetation components using dual-source formulations [*Shuttleworth and Wallace*, 1985]. Soil evaporation is now typically modeled using one of the four following methods, namely the soil surface resistance (r_{ss}) formulation:

$$LE(r_{ss})=\frac{\rho C_P}{\gamma}\times\frac{e_{sat}(T)-e_a}{r_{ah}+r_{ss}} \tag{2}$$

the α formulation:

$$LE(\alpha)=\frac{\rho C_P}{\gamma}\times\frac{\alpha e_{sat}(T)-e_a}{r_{ah}} \tag{3}$$

the β formulation:

$$LE(\beta)=\beta\times\frac{\rho C_P}{\gamma}\times\frac{e_{sat}(T)-e_a}{r_{ah}} \tag{4}$$

or the threshold (LE_{max}) formulation:

$$LE(LE_{max})=\min\left(LEp, LE_{max}\right) \tag{5}$$

with LE (W m^{-2}) being the soil latent heat flux, r_{ss} (s m^{-1}) the resistance to the diffusion of vapor in large soil pores, α a factor (typically ranging from 0 to 1) that scales the saturated vapor pressure down to the actual vapor pressure at the soil surface, β a factor (typically ranging from 0 to 1) that scales potential evaporation down to actual evaporation, ρ (kg m^{-3}) the density of air, C_P (J kg^{-1} K^{-1}) the specific heat capacity of air, γ (Pa K^{-1}) the psychrometric constant, $e_{sat}(T)$ (Pa) the saturated vapor pressure at the soil surface, T (K) the soil surface temperature, e_a (Pa) the vapor pressure of air, r_{ah} (s m^{-1}) the aerodynamic resistance to heat transfer, LEp (W m^{-2}) the potential soil evaporation, and LE_{max} (W m^{-2}) the maximum soil-limited water flux from below the surface. Depending on the authors, the threshold method is also called demand-supply or Priestley-Taylor method and LEp is estimated using the aerodynamic, Penman, or Priestley-Taylor methods. The LE_{max} formulation is equivalent to the β formulation if LE_{max} is parameterized as a fraction of LEp. Note that LE can also be modeled by combining both r_{ss} and α formulations:

$$LE(r_{ss},\alpha)=\frac{\rho C_P}{\gamma}\times\frac{\alpha e_{sat}(T)-e_a}{r_{ah}+r_{ss}}, \tag{6}$$

or both β and α formulations:

$$LE(\beta,\alpha)=\beta\times\frac{\rho C_P}{\gamma}\times\frac{\alpha e_{sat}(T)-e_a}{r_{ah}}. \quad (7)$$

Comprehensive overview of the α, β, r_{ss}, and LE_{max} methods can be found in *Mahfouf and Noilhan* [1991], *Lee and Pielke* [1992], *Ye and Pielke* [1993], *Mihailovic et al*. [1995], *Dekic et al*. [1995] and *Cahill et al*. [1999]. The form of α, β, r_{ss} or LE_{max} is obtained either physically or empirically. Physically based expressions are derived from thermodynamical considerations [*Philip and de Vries*, 1957] or by simplifying the Fick's law of diffusion [e.g., *Dickinson et al*., 1986; *Wetzel and Chang*, 1988; *Sakaguchi and Zeng*, 2009]. All of them simplify the physics underlying the evaporation process and require some empirism to overcome the assumptions. For instance, simplifications of the theoretical diffusion equation require some empirical parameters in addition to SHPs [*Sakaguchi and Zeng*, 2009]. Empirical models are based on ad hoc expressions [e.g., *Manabe*, 1969; *Noilhan and Planton*, 1989] or curve fitting using limited experimental data [e.g., *Sun*, 1982; *Sellers et al*., 1992]. Although many different formulations have been developed since the 60's, there is still no consensus on the best way to parameterize evaporation over large areas [*Desborough et al*., 1996; *Sakaguchi and Zeng*, 2009]. Nevertheless the literature has indicated that (1) existing θ-based formulations differ in four main aspects: the θ lower and upper threshold values, the nonlinearity of the relationship between evaporation and θ, the required input data other than θ, and the sensing depth of θ data, (2) simple empirical expressions may provide better evaporation simulations than physically derived formulations [*Dekic et al*., 1995; *Mihailovic et al*., 1995; *Yang et al*., 1998], (3) the β formulation seems to be more robust than the α one [*Cahill et al*., 1999; *Van den Hurk et al*., 2000], and (4) very little work has been done to evaluate the above formulations with observations over a range of soil and atmospheric conditions.

Phenomenological models are distinct from the above simplified models because they are not derived from theory and they are not built on ad hoc assumptions. Phenomenological models are based on observational data rather than theoretical considerations [*Sivapalan et al*., 2003], but they provide a physical or semi-physical interpretation of model parameters. *Komatsu* [2003] made a first attempt to relate an experimental parameter to soil texture and aerodynamic conditions. However, their study was based on a surface layer of several millimeters, which is much thinner than the top soil thickness (typically several cm) represented by most land-surface models. Moreover, one major difficulty in parameterizing SEE with sufficient generality is the drying (usually around noon) of the top few millimeters of soil which inhibits evaporation, regardless of the availability of soil water underneath [*Mahrt and Pan*, 1984; *Dickinson et al*., 1986; *Soarès et al*., 1988; *Wetzel and Chang*, 1988; *Van de Griend and Owe*, 1994; *Heitman et al*., 2008; *Shahraeeni et al*., 2012]. This was the rationale for developing a new SEE formulation with a shape that adapts to the depth of θ measurements. The study in *Merlin et al*. [2011] provides an insight into ways of taking into account the soil moisture gradient in the topsoil using a simple parameterization as a function of potential evaporation. Their SEE model was evaluated at the daily time scale at two sites located in the same area (southwestern France).

In the vein of *Komatsu* [2003] and *Merlin et al*. [2011], this paper aims to develop a formulation of quasi-instantaneous SEE that builds upon a multi-site data set including a range of soil and atmospheric conditions. This study notably takes advantage of local, regional and global monitoring networks (e.g., AmeriFlux, European Flux Database), which allow to improve models. A new evaporation model is evaluated in terms of SEE estimates over the wide soil texture range observed within the multi-site data set, and is compared with the evaporation modules of three reference land-surface schemes: ISBA (Interaction Sol-Biosphère-Atmosphère) [*Noilhan and Planton*, 1989], CLM (Community Land Model) [*Oleson et al*., 2013], and H-TESSEL (Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land) [*ECMWF*, 2014]. Note that all evaporation modules are implemented in the same energy budget model, using the same forcing data, to ensure that the four models are run in identical conditions.

2. Sites and Data Description

The data set comprises 34 sites distributed in 13 countries (see Table 1). Those sites were or have been developed in the frame of national and international flux station networks (AmeriFlux, FluxNet, European Flux Database, OzNet), long term observatories such as AMMA (African Monsoon Multidisciplinary Analysis), HOBE (Danish Hydrological Observatory) and SudMed (South Mediterranean Observatory), or short term

Table 1. Flux Sites Including One or Several "Bare Soil" Periods

Site	Exp./Net.	Lat;lon	Land cover	θ (cm)	f_{clay}	f_{sand}	Reference
AUStu	OzFlux	−17.15;133.35	grass	5	0.145	0.343	*Beringer et al.* [2011]
BELon	GHGEurope	50.55;4.74	crop	5	0.20	0.075	*Papale et al.* [2006]
CHOe2	GHGEurope	47.29;7.73	crop	5	0.43	0.095	*Alaoui and Goetz* [2008]
DEGeb	GHGEurope	51.10;10.91	crop	5	0.30	0.095	*Kutsch et al.* [2010]
DEKli	GHGEurope	50.89;13.52	crop	5	0.557	0.215	*Kindler et al.* [2011]
DESeh	GHGEurope	50.87;6.45	crop	5	0.122	0.168	*Papale et al.* [2006]
DKVou	HOBE	56.04;9.16	crop	2.5	0.02	0.92	*Bircher et al.* [2012]
ESEFE	EFEDA	39.07; −2.11	bare	10	0.20	0.19	*Braud et al.* [1993]
ESES2	GHGEurope	39.28; −0.32	crop	5	0.475	0.104	*Kutsch et al.* [2010]
FRAur	GHGEurope	43.55;1.11	crop	5	0.323	0.206	*Béziat et al.* [2009]
FRAvi	GHGEurope	43.92;4.88	crop	5	0.328	0.132	*Garrigues et al.* [2015]
FRGri	GHGEurope	48.84;1.95	crop	5	0.189	0.098	*Van den Hoof et al.* [2011]
FRLam	GHGEurope	43.50;1.24	crop	5	0.543	0.12	*Béziat et al.* [2009]
FRRe1	ReSeDa	43.79;4.74	crop	2.5	0.40	0.05	*Olioso et al.* [2002]
FRRe2	ReSeDa	43.79;4.74	crop	2.5	0.40	0.05	*Olioso et al.* [2002]
IECa1	GHGEurope	52.86; −6.92	crop	5	0.17	0.57	*Walmsley et al.* [2011]
ITBCi	GHGEurope	40.52;14.96	crop	5	0.46	0.32	*Denef et al.* [2013]
ITCas	GHGEurope	45.20;9.67	crop	5	0.22	0.25	*Denef et al.* [2013]
ITRo4	GHGEurope	42.37;11.92	crop	5	0.382	0.301	*Marchesini et al.* [2008]
MEYaq	Yaqui'08	27.27; −109.88	crop	5	0.44	0.36	*Chirouze et al.* [2014]
MOSR1	SudMed	31.67; −7.59	crop	5	0.47	0.185	*Er-Raki et al.* [2007]
MOSR2	SudMed	31.67; −7.61	crop	5	0.47	0.185	*Jarlan et al.* [2015]
NIDeg	AMMA	13.65;2.64	bare	10	0.08	0.77	*Pellarin et al.* [2009]
NIHAP	HAPEX	2.24;13.20	bare	5	0.057	0.93	*Wallace et al.* [1993]
NIMil	AMMA	13.64;2.63	crop	10	0.08	0.77	*Pellarin et al.* [2009]
NISav	AMMA	13.65;2.63	fallow	10	0.08	0.77	*Pellarin et al.* [2009]
USArm	AmeriFlux	36.61; −97.49	crop	5	0.43	0.28	*Fischer et al.* [2007]
USDk1	AmeriFlux	35.97; −79.09	grass	10	0.09	0.48	*Novick et al.* [2004]
USFwf	AmeriFlux	35.45; −111.77	grass	2	0.13	0.30	*Dore et al.* [2012]
USIb1	AmeriFlux	41.86; −88.22	crop	2.5	0.35	0.10	*Wu et al.* [2012]
USIHO	IHOP	36.47; 100.62	bare	5	0.28	0.58	*Lemone et al.* [2007]
USMo1	Monsoon'90	31.74; −110.05	shrub	5	0.10	0.66	*Santanello et al.* [2007]
USMo7	Monsoon'90	31.72; −110.01	shrub	5	0.06	0.80	*Santanello et al.* [2007]
USSGP	SGP'97	35.54; −98.06	bare	5	0.24	0.26	*Timmermans et al.* [2007]

intensive field campaigns such as EFEDA (Echival Field Experiment in a Desertification Threatened Area), ReSeDa (Remote Sensing Data Assimilation), Yaqui'08, HAPEX-Sahel (Hydrology-Atmosphere Pilot Experiment), IHOP (International H2O Project), Monsoon'90 and the Southern Great Plains 1997 (SGP'97) experiment (references are provided in Table 1). Among the 34 sites, 26 are located in agricultural fields (BELon, CHOe2, DEGeb, DEKli, DESeh, DKVou, ESEFE, ESES2, FRAur, FRAvi, FRGri, FRLam, FRRe1, FRRe2, IECa1, ITBCi, ITCas, ITRo4, MOSR1, MOSR2, MEYaq, NIMil, USArm, USIb1, USIHO, USSGP) and the rest over uncropped lands. The natural lands include sand desert (NIHAP), savanna fallow (NISav) and degraded land (NIDeg) in Niger, native grass in Australia (AUStu), grass for silage or hay in the USA (USDk1), short grass following fireforest in the USA (USFwf), and sparse shrub in the USA (USMo1 and USMo7). As indicated in Table 1, the study sites cover a large range of surface conditions, with sand and clay fractions ranging from 0.05 to 0.92 and from 0.02 to 0.56, respectively.

The data collected at the above-selected sites contain sand and clay fractions, and the following surface variables measured at the hourly or subhourly time scale: near-surface soil moisture θ, latent heat flux *LE*, sensible heat flux *H*, net radiation *Rn*, ground conduction *G*, and meteorological forcing composed of air temperature T_a, wind speed u_a, solar radiation R_g and air relative humidity h_a at a reference height (typically 2 m). For the sites where a direct measurement of *LE* is unavailable, latent heat flux is estimated as the residual of the energy balance equation. For the sites where the four flux components (*LE*, *H*, *Rn*, *G*) are available, *H* and *LE* are systematically corrected using the Bowen ratio method [*Twine et al.*, 2000]. Note that θ is generally measured at around 5 cm depth but it is located at a shallower or deeper depth at few sites (see Table 1). The "observed" SEE is derived from the ratio of observed evaporation to the potential evaporation, defined as the evaporation based on equation (2) with no surface resistance ($r_{ss} = 0$) but using other observed variables (Rn,G,T_a,u_a,h_a).

One key aspect in this analysis is the identification of the periods when the sites can be considered as under "bare soil" conditions. In this study, a "bare soil" period is defined as a period of time when the plant transpiration is either negligible or small compared to soil evaporation. Hence the term "bare soil" includes both

actual bare soil conditions, and soils partially covered by mulch, crop residue, or sparse vegetation. Whereas it is difficult to quantitatively assess the relative weight of evaporation and transpiration without any direct measurement of the evaporation/transpiration partitioning [*Wang et al*., 2014], some indirect indicators can be used like the Leaf Area Index (LAI), or in-field knowledge of agricultural practices like sowing, tillage and harvest. The bare soil periods were extracted with as much accuracy as possible.

Several sites (ESEFE, NIHAP, USIHO, USSGP) have been monitored under real bare soil conditions in the frame of short-term intensive field campaigns. Most of the sites though are equipped with long-term flux stations located in agricultural fields for which the sowing, tillage and harvest dates have been recorded across one or several growing seasons. Precise and multiannual field work information are available for 14 sites of the European flux database and 3 sites of the Ameriflux database (including all the other variables required in this analysis). In practice, the soil is assumed to be approximately bare during 20 days after each tillage, sowing or harvest date. In this paper, no distinction is made between the bare soil periods following tillage, sowing and harvest. Such additional information might be used in future studies to help separate the effect of soil roughness (after tillage) and the presence of crop residue (after harvest) on the soil evaporation process. Different strategies have been adopted regarding the uncropped lands. The grassland site AUStu is assumed to be bare when the satellite-derived vegetation index is minimum. The savanna fallow NISav is assumed to be bare from the beginning of the Niger 2006 experiment until grass started growing following the first monsoon rainfall events. The grassland site USDk1 is assumed to be bare during 20 days after the annual or biannual harvest date. The sparsely vegetated grassland USFwf, the degraded land NIDeg, and shrublands USMo1 and USMo7 are assumed to be approximately bare at all time (when flux measurements are available).

3. Three Common Evaporation Models

The evaporation modules of three common land-surface schemes are described below. The soil evaporation module of H-TESSEL was recently updated in *Albergel et al*. [2012]. The soil resistance is expressed as:

$$r_{ss}=\frac{\theta_{fc}-\theta_{res}}{\theta-\theta_{res}}\times r_{ss,min} \text{ for } \theta>\theta_{res} \tag{8}$$

with $r_{ss,min}$ (s m^{-1}) being the minimum soil resistance (set to 50 s m^{-1}) [*ECMWF*, 2014], θ_{fc} (m^3 m^{-3}) the soil moisture at field capacity, and θ_{res} (m^3 m^{-3}) the residual soil moisture.

The soil evaporation module of ISBA is based on the α method [*Noilhan and Planton*, 1989]. It represents the nonlinear behavior of α as:

$$\alpha=\begin{cases}0.5-0.5\cos\left(\pi\theta/\theta_{fc}\right), & \text{if } \theta\le\theta_{fc}\\ 1, & \text{if } \theta>\theta_{fc}\end{cases} \tag{9}$$

Regarding CLM, the soil evaporation module of the former version 3.5 [*Oleson et al*., 2007] was based on both α and r_{ss} methods as in equation (6). The water activity α was obtained using the Kelvin equation [*Philip and de Vries*, 1957]:

$$\alpha=\exp\left[\psi g/(1\times10^3 R_{wv}T)\right] \tag{10}$$

with g (m s^{-2}) being the gravitational constant, R_{wv} (J kg^{-1} K^{-1}) the gas constant for water vapor, and ψ (mm) the soil water matric potential of the surface soil layer computed as:

$$\psi=\psi_{sat}\times\left(\theta/\theta_{sat}\right)^{-b_{CH}} \tag{11}$$

with ψ_{sat} (mm) being the air entry pressure, θ_{sat} (m^3 m^{-3}) the soil moisture at saturation, and b_{CH} the Clapp and Hornberger parameter [*Clapp and Hornberger*, 1978]. r_{ss} was derived from *Passerat de Silans* [1986] and *Sellers et al*. [1992]:

$$r_{ss}=\exp\left(A-B\,\theta/\theta_{fc}\right) \tag{12}$$

with A and B being two best-fit parameters estimated as 8.206 and 4.255 respectively using FIFE'87 measurements in *Sellers et al*. [1992]. Despite its empirical nature, the modeling approach of *Passerat de Silans*

[1986] has been widely used in land-surface models [*Sellers et al*., 1992, 1996; *Kustas et al*., 1998; *Vidale and Stöckli*, 2005; *Gentine et al*., 2007; *Crow et al*., 2008; *Oleson et al*., 2008; *Stöckli et al*., 2008]. The r_{ss} formulation in equation (12) is referred to as S92 in the following.

The soil evaporation module of the last (4.5) CLM version [*Oleson et al*., 2013] combines both α and β methods as in equation (7). α is estimated as in equation (10) and β is expressed as in *Lee and Pielke* [1992]:

$$\beta=\begin{cases}[0.5-0.5\cos(\pi\theta/\theta_{fc})]^2, & \text{if } \theta\le\theta_{fc}\\ 1, & \text{if } \theta>\theta_{fc} \text{ if } \alpha e_{sat}(T)<e_a\end{cases} \tag{13}$$

The pedotransfer functions (PTFs) used to estimate θ_{fc}, θ_{res}, ψ_{sat}, θ_{sat}, and b_{CH} from sand and clay fractions are presented in Appendix A.

4. A Downward Modeling Approach of SEE

The rationale for choosing to model SEE instead of soil evaporation directly is that SEE, as a normalized variable, helps disentangle the two main factors controlling soil evaporation: evaporative demand (or *LEp*) and soil water availability. In particular, the SEE fosters the decoupling between the evaporation cycles associated with (1) the diurnal, seasonal and climatic variations of *LEp* and (2) the variations of soil water availability due to natural (rainfall) and/or man-induced (irrigation) precipitations. Note that the formulation in SEE only partly decouples the effect of soil water availability and *LEp* since the soil moisture profile changes with *LEp* [*Merlin et al*., 2011], and *LE* and *LEp* are generally coupled (e.g., see complementary relationship in *Lintner et al*. 2015]. Moreover, the advective part also contributes to SEE due to the drop in temperature that reduces Cw at the evaporative surface [*Chanzy and Bruckler*, 1993]. Nonetheless, the normalization of actual evaporation by the evaporative demand removes the first order effect of *LEp* on *LE*, and sets SEE to lie between ~0 and 1. The limits are theoretically reached when soil water availability is respectively negligible ($\theta=\theta_{res}$, SEE ~ 0) and maximum ($\theta=\theta_{sat}$, SEE=1) regardless of the atmospheric evaporative demand. Soil evaporation can then be estimated by multiplying the modeled SEE by *LEp*, which is derived from meteorological data solely.

Another significant advantage of the formulation in SEE is the strong link with remote sensing variables available in the thermal and microwave frequencies. In particular, the SEE-based representation of evaporation is fully consistent with both the thermal-derived *T* normalized by wet/dry *T* endmembers [e.g., *Nishida et al*., 2003; *Stefan et al*., 2015], and the θ retrieved from microwave data [e.g., *Prévot et al*., 1984; *Simmonds and Burke*, 1999; *Zribi et al*., 2011].

A new SEE model is developed based on a downward (data-driven) approach. The downward modeling approach aims to minimize the number of model parameters while ensuring a sufficient flexibility of the SEE formulation to cover a large range of soil and atmospheric conditions. In practice, the step-wise procedure below is followed:

1. SEE is expressed based on equation (12), as a function of two parameters noted $r_{ss,ref}=\exp(A)$ and $\theta_{efolding}=\theta_{fc}/B$,
2. $r_{ss,ref}$ and $\theta_{efolding}$ are analytically expressed as a function of meteorological conditions, and of two observational parameters namely the cut-off soil moisture value $\theta_{1/2}$ (m^3m^{-3}) at which SEE=0.5, and the first derivative noted $\Delta\theta_{1/2}^{-1}$ (m^3m^{-3}) of SEE at $\theta_{1/2}$,
3. SEE is assumed to be a unique function of θ, $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$. The variabilities of SEE attributed to factors other than θ (e.g., soil texture) are therefore contained in $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$,
4. a retrieval procedure of $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ is proposed for a given time series of SEE and θ data (the calibration period should include significant variability in θ i.e., at least one drying sequence),
5. variabilities in $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ are interpreted in terms of soil and atmospheric conditions, which can be characterized by the soil texture, soil roughness, presence of stubble or mulch at the soil surface, shrinkage cracks, etc. In this study, a focus is made on a texture-based calibration of $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ because sand and clay fractions are relatively easy to obtain and are generally available at the site level.

The input/output data sets and the main steps of the modeling, calibration and validation approaches are presented in the diagram of Figure 1. An analytical relationship between $(r_{ss,ref};\theta_{efolding})$ and $(\theta_{1/2};\Delta\theta_{1/2}^{-1})$ is first built by running a soil energy balance model for two extreme conditions with $r_{ss}=0$ and $r_{ss}\sim\infty$ using

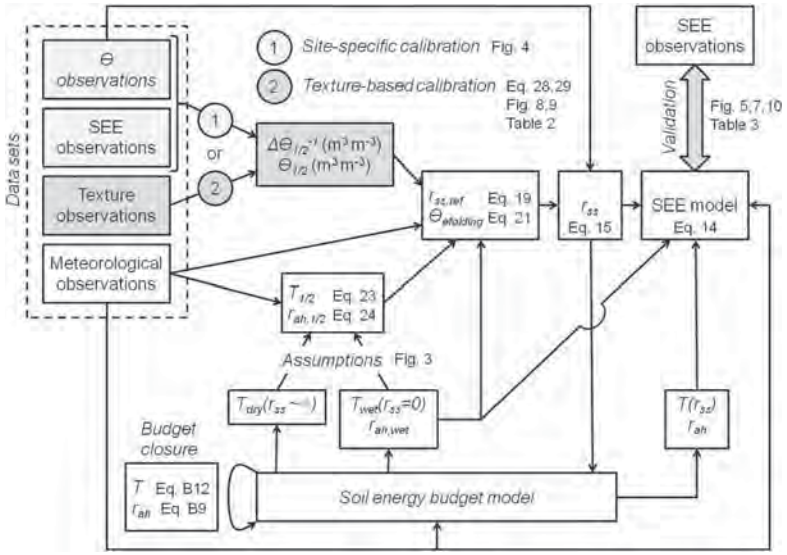


Figure 1. Overview of the modeling, calibration and validation approaches including input/output data sets and the main equations, figures and tables.

meteorological forcing solely, and by approaching the middle point from the two (wet and dry) reference points. Two methods are then investigated to estimate the pair $(\theta_{1/2}; \Delta\theta_{1/2}^{-1})$. The first method (site-specific calibration) is based on the time series of SEE and θ data for a given site, while the second method (texture-based calibration) parameterizes $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ as a function of the clay and/or sand fractions for all sites. The associated equations, figures and tables are also indicated in the diagram for clarity. The model development is described below, along with the underlying assumptions.

4.1. r_{ss}-Based SEE Model

The energy balance of physically based land-surface schemes is generally represented using a resistance network. Therefore, the r_{ss}-based formulation is preferred, as it facilitates the integration of the SEE model in the majority of existing land-surface models. SEE is hence written as:

$$\mathrm{SEE} = \frac{e_{sat}(T) - e_a}{e_{sat}(T_{wet}) - e_a} \times \frac{r_{ah,wet}}{r_{ss} + r_{ah}} \tag{14}$$

with T_{wet} being the temperature of a water-saturated soil (corresponding to $r_{ss} = 0$), and $r_{ah,wet}$ the associated aerodynamic resistance to heat transfer. Note that in the prospect of integrating the above formulation in a given land-surface model, equation (14) can be inverted to express r_{ss} as a function of modeled SEE.

In equation (14), the variability of SEE attributed to soil water availability (via θ and the soil properties including soil texture, structure, and roughness) is assumed to be contained in r_{ss}. In this study, the general form of the S92 r_{ss} formulation is used:

$$r_{ss} = r_{ss,ref} \exp\left(-\theta/\theta_{efolding}\right) \tag{15}$$

with $r_{ss,ref} = \exp(A)$ (s m^{-1}) being the asymptotic value of r_{ss} for $\theta \sim 0$, and $\theta_{efolding} = \theta_{fc}/B$ (m^3 m^{-3}) the soil moisture value at which $r_{ss} = r_{ss,ref}/e$. The exponential form of equation (15) is convenient for analytically expressing the derivatives of r_{ss} and SEE.

4.2. Linear Approximation of SEE at the Mid-Value

Many studies have documented the strongly nonlinear behavior of SEE as a function of θ [e.g., *Chanzy and Bruckler*, 1993; *Komatsu*, 2003; *Merlin et al.*, 2011]. Modeling a nonlinear phenomenon is challenging because small uncertainties in model parameterization may have a large impact on predictions. As an attempt to approximate SEE over its full range $[0-1]$, SEE is approached linearly at the mid-value (0.5). The linear

approximation of SEE(θ) at SEE=0.5 sets two constraints on the model. First, the soil moisture value at which SEE=0.5 is noted $\theta_{1/2}$:

$$\mathrm{SEE}(\theta_{1/2})=0.5 \tag{16}$$

Second, the first derivative of SEE at $\theta_{1/2}$ is set to the slope ($\Delta\theta_{1/2}^{-1}$) of the linear regression between SEE and θ observations:

$$\left(\frac{\partial \mathrm{SEE}}{\partial\theta}\right)(\theta_{1/2})=\Delta\theta_{1/2}^{-1} \tag{17}$$

The combination of the above two equations allows to estimate both $r_{ss,ref}$ and $\theta_{efolding}$ parameters given a time series of SEE and θ observations (described in the following section). As an illustration of the approximation approach, Figure 2 plots the SEE simulated by the model in *Merlin et al.* [2011] as a function of θ for two different sets of parameters. In *Merlin et al.* [2011], SEE was written as:

$$\mathrm{SEE}=\begin{cases}[0.5-0.5\cos(\pi\theta/\theta_{sat})]^P, & \text{if } \theta\le\theta_{sat}\\ 1, & \text{if } \theta>\theta_{sat}\end{cases} \tag{18}$$

with P being a semi-empirical parameter expressed as a function of the soil moisture sensing depth (L) and LEp. The phenomenological expression in equation (18) is based on the observation that both L and LEp have an equivalent impact on SEE, meaning that (1) SEE is controlled by the soil moisture profile within the soil thickness L and (2) the soil moisture profile is affected by both L and LEp. This is consistent with the recent study of *Brutsaert* [2014] who described the daily water flow in the soil profile by considering the soil as an infinite domain during stage 1, and a layer of constant thickness whose lower boundary is a zero-flux plane during stage 2. The decrease of SEE with increasing LEp is generally related to the formation of a dry surface layer above the evaporative front [*Fritton et al.*, 1967; *Yamanaka et al.*, 1998], modifying the soil moisture profile within the soil sensing depth. Figure 2 plots the SEE simulated with $(P_1,\theta_{sat,1})=(1,0.40)$ and $(P_2,\theta_{sat,2})=(4,0.45)$. One observes that $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ are different in both cases. The modeling strategy aims to represent the nonlinear behavior of SEE within the full SEE range from $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ parameters, and the exponential formulation in equation (15).

4.3. Analytical Expressions of $r_{ss,ref}$ and $\theta_{efolding}$

Parameters $r_{ss,ref}$ and $\theta_{efolding}$ in equation (15) are analytically expressed as a function of $\theta_{1/2}$, $\Delta\theta_{1/2}^{-1}$, soil temperature (T_{wet} and $T_{1/2}$) and aerodynamic resistance ($r_{ah,wet}$ and $r_{ah,1/2}$) values corresponding to $r_{ss}=0$ and $\theta=\theta_{1/2}$, respectively. A soil energy balance model [e.g., *Norman et al.*, 1995; *Merlin and Chehbouni*, 2004] is used to estimate both pairs $(T_{wet},r_{ah,wet})$ and $(T_{1/2},r_{ah,1/2})$ for a given meteorological forcing.

Briefly, $r_{ss,ref}$ is derived by inverting equation (15):

$$r_{ss,ref}=r_{ss,1/2}\exp(\theta_{1/2}/\theta_{efolding}) \tag{19}$$

with $r_{ss,1/2}$ being the soil resistance at $\theta_{1/2}$ obtained by combining equations (14) and (16):

$$r_{ss,1/2}=2\frac{e_{sat}(T_{1/2})-e_a}{e_{sat}(T_{wet})-e_a}r_{ah,wet}-r_{ah,1/2} \tag{20}$$

$\theta_{efolding}$ is obtained by applying the first derivative at $\theta=\theta_{1/2}$ to the soil energy balance equation:

$$\theta_{efolding}=\frac{\frac{r_{ss,1/2}}{r_{ss,1/2}+r_{ah,1/2}}(e_{sat}(T_{1/2})-e_a)+f(\theta_{1/2})\dot{e}_{sat}(T_{1/2})}{\frac{r_{ss,1/2}+r_{ah,1/2}}{r_{ah,wet}}(e_{sat}(T_{wet})-e_a)}\times\frac{1}{\Delta\theta_{1/2}^{-1}} \tag{21}$$

with $\dot{e}_{sat}$ being the derivative of saturated vapor pressure with respect to T and $f(\theta_{1/2})$ expressed as:

$$f(\theta_{1/2})=-\frac{\frac{r_{ss,1/2}r_{ah,1/2}}{(r_{ss,1/2}+r_{ah,1/2})^2}(e_{sat}(T_{1/2})-e_a)}{\gamma+\frac{r_{ah,1/2}}{r_{ss,1/2}+r_{ah,1/2}}\dot{e}_{sat}(T_{1/2})+4\frac{\gamma}{\rho C_P}\epsilon\sigma(1-C_G)T_{1/2}^3r_{ah,1/2}} \tag{22}$$

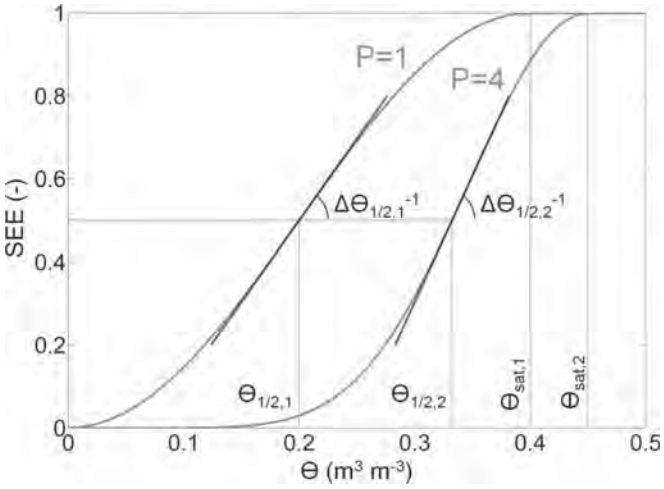


Figure 2. The SEE(θ) relationship is approximated at the midvalue (SEE=0.5) by the tangent defined by the pair $(\theta_{1/2}, \Delta\theta_{1/2}^{-1})$, for two different scenarios 1 and 2.

with C_G being the ratio of the ground conduction to soil net radiation. A presentation of the soil energy balance model is provided in Appendix B and the analytical development of $\theta_{efolding}$ is described in Appendix C.

4.4. Model Assumptions

In equations (20)–(22), a first guess of $T_{1/2}$ is given by:

$$T_{1/2}=(T_{wet}+T_{dry})/2 \qquad (23)$$

with T_{dry} being the T of a fully dry soil (corresponding to $r_{ss} \sim \infty$), and a first guess of $r_{ah,1/2}$ is given by:

$$r_{ah,1/2}=r_{ah}(T_{1/2}) \qquad (24)$$

with $r_{ah}(T_{1/2})$ being the aerodynamic resistance estimated using the Richardson number (see equations (B9) and (B11)) for atmospheric stability corrections at $T=T_{1/2}$. Note that equations (23) and (24) are approximations since an accurate determination of $T_{1/2}$ and $r_{ah,1/2}$ would require an iterative process on $r_{ss,1/2}$ (and $r_{ss,ref}$ and $\theta_{efolding}$), and thus multiple applications of the soil energy balance model. However, such approximations are considered to be valid at first order given that turbulent heat fluxes and surface temperatures are near linearly related [e.g., *Moran et al.*, 1994; *Roerink et al.*, 2000; *Merlin*, 2013].

To assess the validity of the model assumptions, Figures 3b, 3e, and 3h compare the SEE simulated by the S92 and new r_{ss} formulations as a function of observed θ, for NIMil, FRAvi and FRLam data sets, respectively. In general, the scatter in simulated SEE is reduced with the new formulation. This is consistent with the assumed number (3) of degrees of freedom of the SEE model expressed as a function of θ, $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ solely. Moreover, the behavior at around $\theta=\theta_{1/2}$ of the SEE simulated using the new r_{ss} formulation is very close to the mean regression defined by the pair $(\theta_{1/2}, \Delta\theta_{1/2}^{-1})$. The linearity assumption SEE(T) implicitly made in equation (23) can also be verified by investigating the relationship between simulated SEE and the simulated temperature T normalized by T_{dry} and T_{wet}:

$$T_{norm}=\frac{T_{dry}-T}{T_{dry}-T_{wet}} \qquad (25)$$

Figures 3c, 3f, and 3i plot simulated SEE versus simulated T_{norm} for the S92 and new r_{ss} formulations separately, and for NIMil, FRAvi and FRLam data sets, respectively. The physically based soil energy balance model represents a quasi linear relationship between T_{norm} and SEE for all three data sets, regardless the r_{ss} formulation. Note that T_{norm} slightly overestimates simulated SEE, especially at the mid values. This is due to the impact of the dependence of r_{ah} on $T-T_a$ (see equation (B11)) on modeled SEE. However, the mean bias between T_{norm} and simulated SEE is very small in all cases. The above verifications thus indicate that the assumptions in equations (23) and (24) are deemed acceptable to approximate SEE at its mid value.

4.5. Retrieving $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ from SEE and θ Data

An algorithm is proposed to retrieve both $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ from a given time series of SEE and θ observations. The retrieval of $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ is not a trivial task due to 1) the nonlinear behavior of SEE(θ), 2) uncertainties in SEE and θ observations, and 3) as mentioned before the possible impact of variability factors other than θ, such as the sensing depth of θ measurements, soil moisture profile, soil roughness, presence of stubble or mulch at the soil surface, shrinkage cracks, etc., which may significantly affect the observed relationship between SEE and θ. Nonetheless, the procedure described below is designed to provide a robust estimate of $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ for strongly noised and nonlinear SEE(θ) relationships.

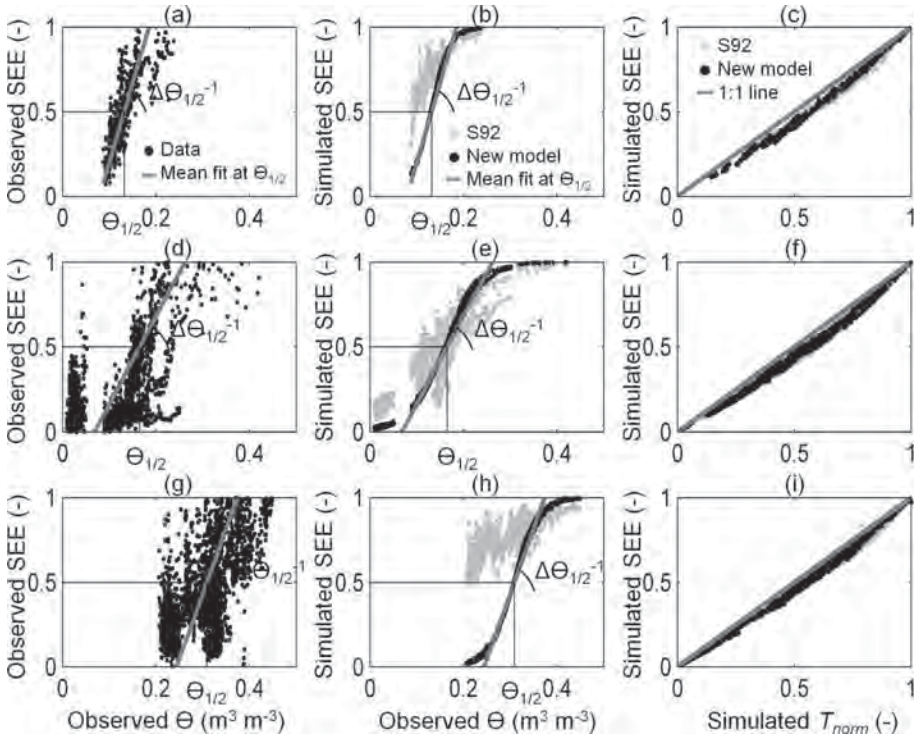


Figure 3. (left) Observed SEE versus observed θ, (middle) simulated SEE versus observed θ, and (right) simulated SEE versus simulated T_{norm}, ranging from sandy to clayey soil conditions: (top) NIMil, (middle) FRAvi, and (bottom) FRLam data set.

The main idea is to consider a regression between SEE and θ around $\theta=\theta_{1/2}$. A schematic representation based on the AUStu data set is provided in Figure 4. First, the full SEE range $[0-1]$ is split into 20 0.05-wide bins, and the SEE and θ values falling into each SEE bin are averaged separately to provide a pair $(\overline{\text{SEE}}_k, \bar{\theta}_k)$ per bin. Then, 10 regression segments are computed by joining the two points $(\overline{\text{SEE}}_k, \bar{\theta}_k)$ and $(\overline{\text{SEE}}_{k+10}, \bar{\theta}_{k+10})$ for $k=1,\ldots,10$. Next, the slope $(\Delta\theta_{1/2}^{-1})$ of the mean regression at around the mid-value (SEE=0.5) is estimated by taking the average of the slope of the 10 distinct regression segments, weighted by the number of data points within each bin pair (i.e., weights are computed as the multiplication of the number of data points within the two bins k and $k+10$). Last, $\theta_{1/2}$ is derived from $\Delta\theta_{1/2}^{-1}$ and the mean observed θ. Note that the average of multiple slopes is more appropriate than using a single slope centered on SEE=0.5 as it allows for a robust application to any data set, including those with observed θ values mostly in the lower or higher soil moisture range.

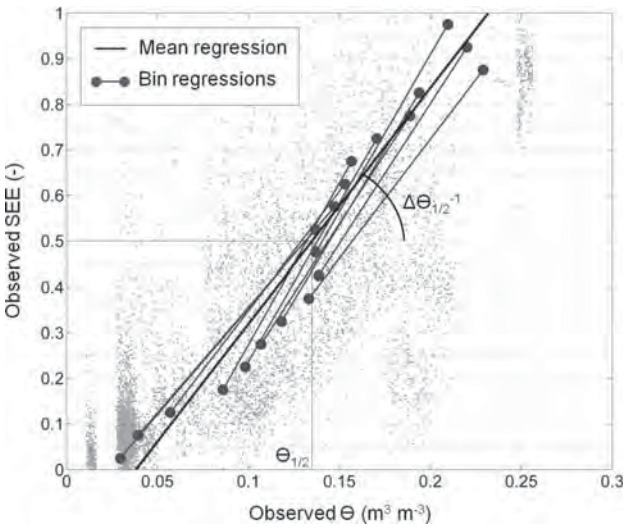


Figure 4. Schematic representation of the retrieval of $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ from the AUStu data set.

As an illustration, Figures 3a, 3d, and 3g plot observed SEE versus observed

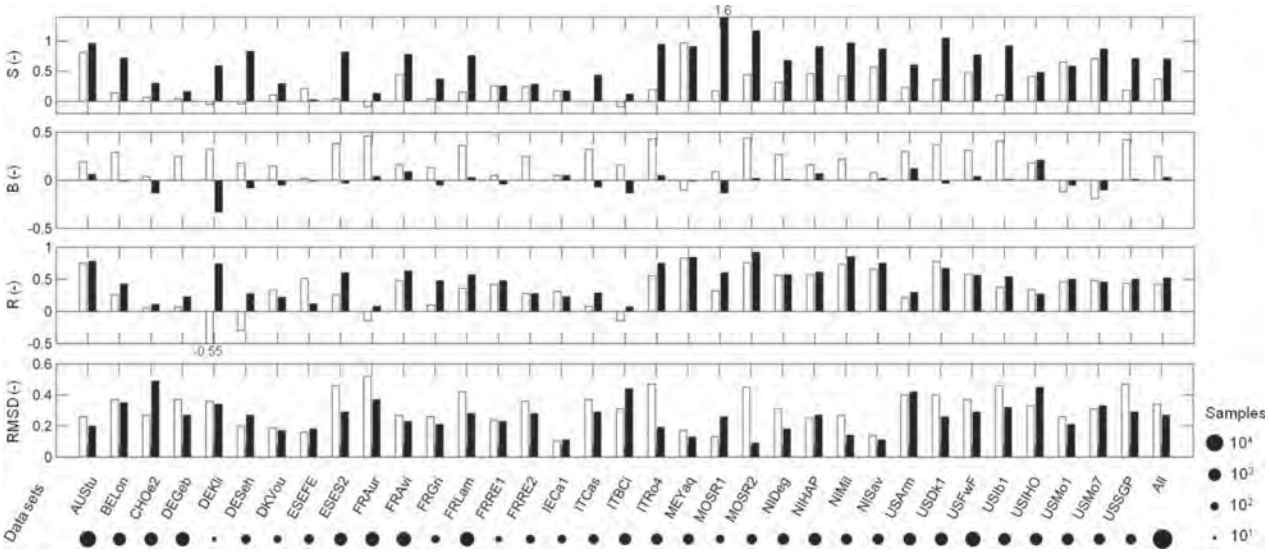


Figure 5. Bar graph of the root mean square difference (RMSD), correlation coefficient (R), mean bias (B), and slope of the linear regression (S) between simulated and observed SEE for the new r_{ss} with site-specific $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ parameters (black) and for the S92 r_{ss} (white). The number of data samples for each site is also illustrated.

θ for NIMil, FRAvi and FRLam data sets respectively, ranging from sandy to clayey soil conditions. The retrieved $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ significantly vary across the three sites.

5. Evaluation and Intercomparison of SEE Models

In this section, the SEE model based on the $r_{ss}(\theta_{1/2}, \Delta\theta_{1/2}^{-1})$ formulation is evaluated using the bare soil data collected at the 34 sites. First, $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ are retrieved for each data set, and the new r_{ss} formulation is assessed using site-specific parameters. Second, a generic parameterization of $(\theta_{1/2}, \Delta\theta_{1/2}^{-1})$ is proposed as a function of soil texture i.e., the clay and sand percentages. Third, the texture-based r_{ss} formulation is compared with the PTFs of four common evaporation models in terms of SEE estimates. Note that only the data with $Rn-G > 100$ W m^{-2} and $LEp > 100$ W m^{-2} are considered in this study to avoid large uncertainties in SEE observations and to avoid energy limited conditions. The C_G coefficient in equation (22) is set to the minimum between 0.315 [*Kustas et al.*, 1991] and the observed G to Rn ratio, and the C_G values below 0.05 are set to 0.05 according to maximum and minimum values found in the literature [*Su*, 2002].

5.1. Site-Specific Calibration

The pair $(\theta_{1/2}, \Delta\theta_{1/2}^{-1})$ is retrieved using the algorithm described in section 4.5 for each site separately. To assess the impact of a site-specific calibration of $r_{ss,ref}$ and $\theta_{efolding}$, SEE simulations are evaluated against multi-site observations. Moreover, results are compared with the SEE simulated by the S92 rss formulation. Figure 5 presents bar graphs of the root mean square difference (RMSD), mean bias (B), correlation coefficient (R) and slope of the linear regression (S) between simulated and observed SEE in both cases, and the number of data samples for each site separately. The mean (weighted by the number of data samples per site) RMSD is 0.27 instead of 0.34, the mean R is 0.52 instead of 0.43, the mean B is 0.03 instead of 0.24, and the mean S is 0.70 instead of 0.37 for the calibrated new and S92 formulations, respectively. Statistics are generally improved with the calibrated rss formulation. Especially, the mean bias is much reduced and the slope of linear fit closer to 1. The strategy of approximating SEE at $(\theta_{1/2}, 0.5)$ thus appears to be effective in improving the representation of SEE over its full range. Note that a sensitivity analysis (not shown) revealed that setting C_G to a constant between 0.25 and 0.40 degrades the modeling results especially in terms of correlation and slope of the linear regression between modeled and observed SEE.

When looking at individual sites in Figure 5, a degradation of RMSD, R and S can be observed. Notably the statistics for CHOe2, ITBCi, and USIHO, and to a lesser extent for ESEFE and IECa1 indicate an increase of the RMSD with the calibrated new (compared to S92) rss formulation. To help interpret those seemingly

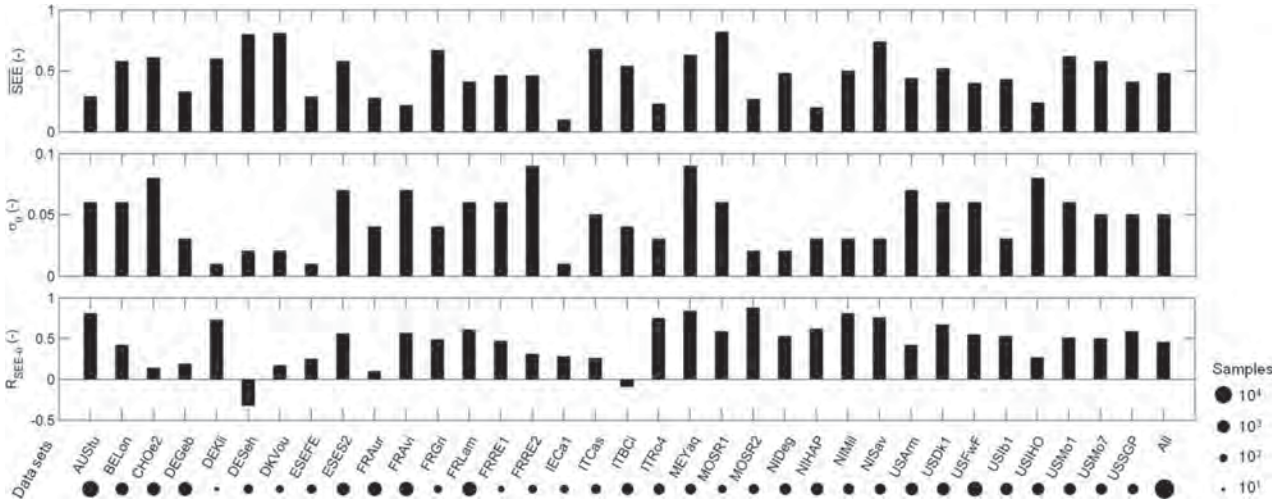


Figure 6. Bar graph of the mean observed SEE ($\overline{\text{SEE}}$), standard deviation of observed θ (σ_θ), and correlation coefficient between observed SEE and observed θ ($R_{\text{SEE}-\theta}$). The number of data samples for each site is also illustrated.

inconsistent results, Figure 6 presents bar graphs of the mean observed SEE ($\overline{\text{SEE}}$), the standard deviation of observed θ (σ_θ), the correlation coefficient between SEE and θ observations ($R_{\text{SEE}-\theta}$) and the number of data samples (n) for each site separately. The correlation coefficient between SEE and θ observations is poor with 0.14, -0.09, and 0.27 for CHOe2, ITBCi, and USIHO respectively, while the mean $R_{\text{SEE}-\theta}$ for all sites is estimated as 0.46. In addition σ_θ for ESEFE and IECa1 is very small (0.01) as compared with the mean σ_θ (0.05) for all sites, and the $\overline{\text{SEE}}$ for ESEFE, IECa1 and USIHO (0.29, 0.10 and 0.24 respectively) is relatively far from the SEE mid value, as compared with the mean $\overline{\text{SEE}}$ (0.48) for all sites. Hence the poorer SEE statistics for CHOe2, ESEFE, IECa1, ITBCi and USIHO are probably attributed to the limited range of soil moisture and atmospheric conditions present in the respective data sets. A lack of variability in the surface conditions encountered in the input data set weakens the robustness of the retrieval approach.

Figure 7 plots the SEE simulated using the site-specific $(\theta_{1/2}, \Delta\theta_{1/2}^{-1})$ as a function of observed SEE for different ranges of clay fractions separately. When comparing the RMSD, R, B and S for each clay fraction range, one observes that the performance of the SEE model is superior for low clay contents ($f_{clay} < 0.20$) than for relatively high clay content ($f_{clay} \geq 0.30$). The effect is especially reflected in R and S, which both are about

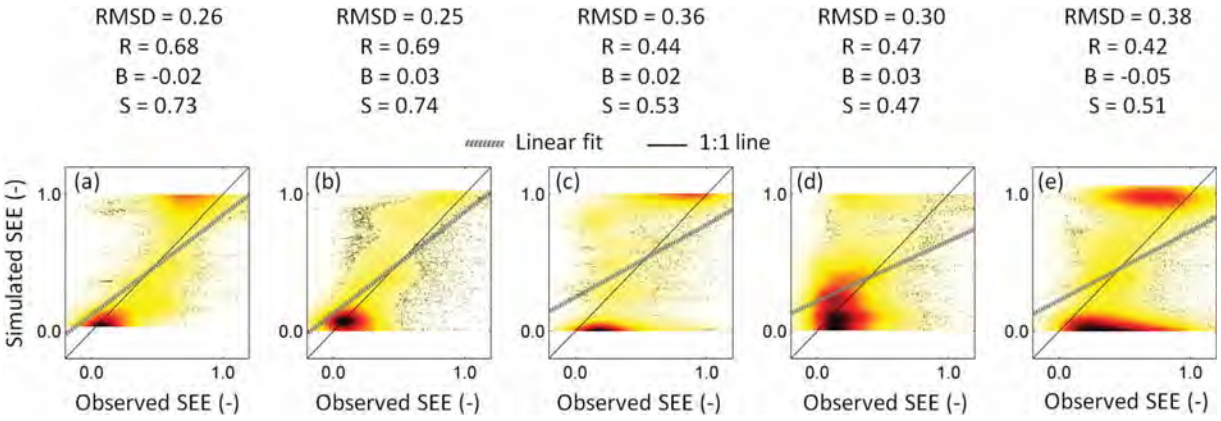


Figure 7. The SEE simulated by the new r_{ss} model with site-specific calibration is plotted as a function of observed SEE for different ranges of clay fraction: (a) $f_{clay} < 0.10$, (b) $0.10 \leq f_{clay} < 0.20$, (c) $0.20 \leq f_{clay} < 0.30$, (d) $0.30 \leq f_{clay} < 0.40$, and (e) $f_{clay} \geq 0.40$. Each graph is a smoothed histogram of the bivariate (modeled versus observed) SEE data. Black shading represents the maximum smoothed density of data points, while the individual points (ouliers) are plotted where the smoothed density is less than 10% of the maximum density. The root mean square difference (RMSD), correlation coefficient (R), mean bias (B), and slope of the linear regression (S) between simulated and observed SEE are also indicated for each case.

Table 2. Slope and Intercept of the Linear Regression, and Correlation Coefficient Between Retrieved $\theta_{1/2}$ and Clay and Sand Fractions, Respectively for Site Data Sets Containing at Least 1, 101, and 501 Samples, Separately

	Clay-Based Linear Fit			Sand-Based Linear Fit		
n	Slope ($m^3\ m^{-3}$)	Intercept ($m^3\ m^{-3}$)	R	Slope ($m^3\ m^{-3}$)	Intercept ($m^3\ m^{-3}$)	R
>0	0.41	0.11	0.62	−0.26	0.29	−0.65
>100	0.41	0.11	0.57	−0.26	0.29	−0.69
>500	0.43	0.10	0.76	−0.27	0.29	−0.76

0.70 and 0.45, for the $f_{clay} < 0.20$ and $f_{clay} \geq 0.30$ case respectively. It is suggested that SEE is more difficult to model from θ data in clayey than in sandy soils, especially because of the "dynamic" formation of a dry surface layer under relatively large evaporative demand conditions [*Fritton et al.*, 1967; *Yamanaka et al.*, 1998]. Nevertheless, the "static" site-specific calibration of r_{ss} via $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ (compared to the default S92 parameters) significantly reduces the bias between simulated and observed SEE for each clay fraction range, and generally improves the R, S and RMSD across the multi-site data set. This is the rationale for developing a texture-based calibration of $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ as a first guess of the variabilities in the parameters $r_{ss,ref}$ and $\theta_{efolding}$.

5.2. Toward a Texture-Based Calibration

Given that $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ are semi-empirical parameters, and that SEE and θ observations may not be available to retrieve the pair $(\theta_{1/2}, \Delta\theta_{1/2}^{-1})$ at all locations, a PTF is proposed. In practice, the parameters $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ retrieved for each site separately are related to the site sand and clay fractions.

A significant correlation is found between retrieved $\theta_{1/2}$ and soil texture with a correlation coefficient in the range of 0.6–0.8 for both sand and clay fractions (see Table 2). Specifically, the R statistics estimated for sites with $n>0$ (34 sites), $n>100$ (30 sites) and $n>500$ (19 sites) is 0.62, 0.57 and 0.76 with clay fraction, and −0.65, −0.69 and −0.76 with sand fraction, respectively. Figure 8 plots retrieved $\theta_{1/2}$ as a function of sand and clay fractions for the 19 sites with $n>500$. The soil moisture $\theta_{1/2}$ at which SEE=0.5 is an increasing function of clay fraction and a decreasing function of sand fraction. This indicates that the hydric potential curves that control evaporation according to θ are shifted as a function of texture. The observed phenomenon is also consistent with Fick's law, predicting that evaporation is inversely proportional to porosity and depth of the vaporization front, which both increase with decreasing size of soil pores and particles [*Or et al.*, 2013]. Note that one site (ESES2) appears to significantly deviate from the linear regression based on either clay or sand fraction (see Figure 8). As a crop rice field, ESES2 is flooded most of the time. When

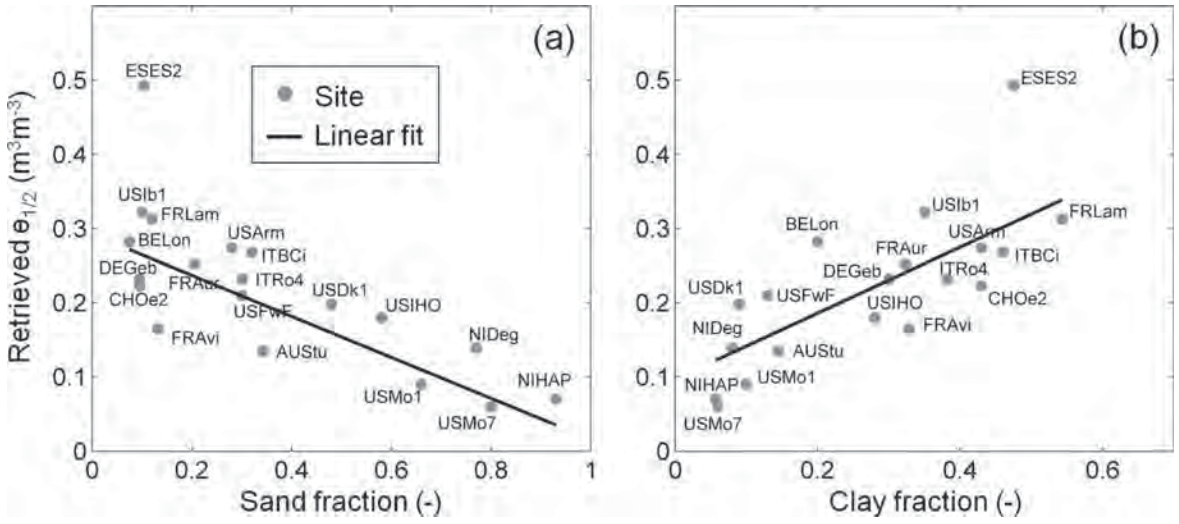


Figure 8. Semi-empirical parameter $\theta_{1/2}$ retrieved for each site separately, as function of (a) sand and (b) clay fraction for data sets containing at least 500 samples ($n>500$), including the respective linear fits.

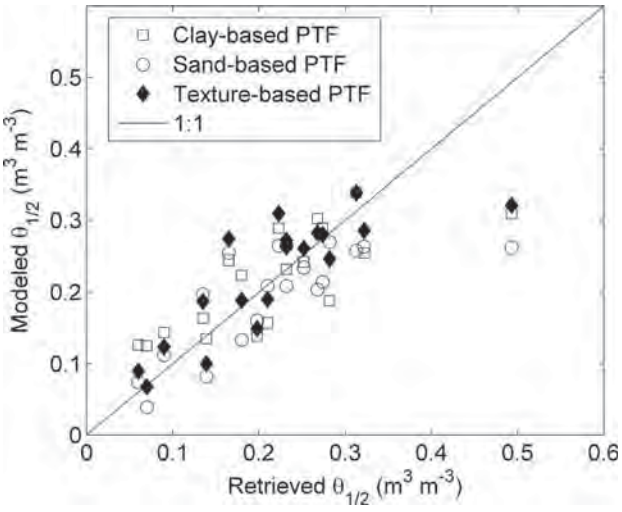


Figure 9. Modeled versus retrieved $\theta_{1/2}$ for clay-based, sand-based and texture-based (sand and clay together) PTF separately.

discarding the specific case of ESES2 from linear regressions, the R between retrieved $\theta_{1/2}$ and clay and sand fraction is 0.77 and −0.83, respectively. To quantify the consequences of site selection decisions, the 6 sparsely vegetated (AUStu, NISav, USDk1, USFwf, USMo1, USMo7) sites were removed from the "bare soil" database. The correlation between retrieved $\theta_{1/2}$ and clay/sand fraction decreased from 0.76 down to 0.69 (absolute value for $n > 500$), indicating that site selection is a tradeoff between total number of points (including a range of clay/sand fractions) and potential quality.

Three PTFs of $\theta_{1/2}$ are tested, using the multi-site data set with $n > 500$. The clay-based $\theta_{1/2}$ is:

$$\theta_{1/2} = 0.10 + 0.43 f_{clay} \tag{26}$$

and the sand-based model is:

$$\theta_{1/2} = 0.29 - 0.27 f_{sand}. \tag{27}$$

An interesting feature with the sand-based linear regression is that the extrapolated value of $\theta_{1/2}$ at $f_{sand} = 1$ is ~0. A third PTF (in the following referred to as "texture-based PTF") is built from the multilinear regression of retrieved $\theta_{1/2}$ with both clay and sand fractions:

$$\theta_{1/2} = 0.20 + 0.28 f_{clay} - 0.16 f_{sand} \tag{28}$$

Figure 9 plots modeled versus retrieved $\theta_{1/2}$ for clay-based, sand-based and texture-based PTFs separately. The multilinear regression of $\theta_{1/2}$ including clay and sand fractions (texture-based PTF) improves the model statistics: the R (and RMSD) between modeled and retrieved $\theta_{1/2}$ is 0.76, 0.76 and 0.81 (and 0.065, 0.068 and 0.058 $m^3\ m^{-3}$) respectively. Although clay and sand fractions are somewhat correlated via the silt fraction ($f_{clay} + f_{sand} = 1 - f_{silt}$ in inorganic soils), it is suggested that both fractions provide complementary information on soil water retention capacity, especially in the case where one of the fraction (f_{clay} or f_{sand}) is small. Consequently, the PTF in equation (28) is used in the following to estimate $\theta_{1/2}$ from site-specific textural information.

Regarding $\Delta\theta_{1/2}^{-1}$, no significant correlation is obtained with either clay or sand fraction. The R estimated for sites with $n > 0$, $n > 100$, and $n > 500$ is 0.06, 0.10 and 0.23 with clay fraction, and −0.04, −0.13 and −0.19 with sand fraction, respectively. Consequently, $\Delta\theta_{1/2}^{-1}$ is set to a constant equal to the mean value for all sites with $n > 500$:

$$\Delta\theta_{1/2}^{-1} = 8\ m^3 m^{-3} \tag{29}$$

Note that the standard deviation of $\Delta\theta_{1/2}^{-1}$ across the 19 sites is about 4, which is relatively large compared to the mean. The variability of $\Delta\theta_{1/2}^{-1}$ can be attributed to a number of factors such as the soil water availability in deeper soil layers, atmospheric conditions (at interannual, seasonal, daily and hourly time scales), surface state (roughness, presence of residus, etc.), and farming practices (e.g., ploughing) for crop sites during the selected bare soil periods. Although the mean $\Delta\theta_{1/2}^{-1}$ may not be representative for all sites, equation (29) is used in the following as a best first guess.

Figure 10 presents bar graphs of the RMSD, R, B and S between simulated and observed SEE for the new r_{ss} with texture-based $\theta_{1/2}$ (equation (28)) and $\Delta\theta_{1/2}^{-1}$ (equation (29)) as well as for the S92 r_{ss} formulation.

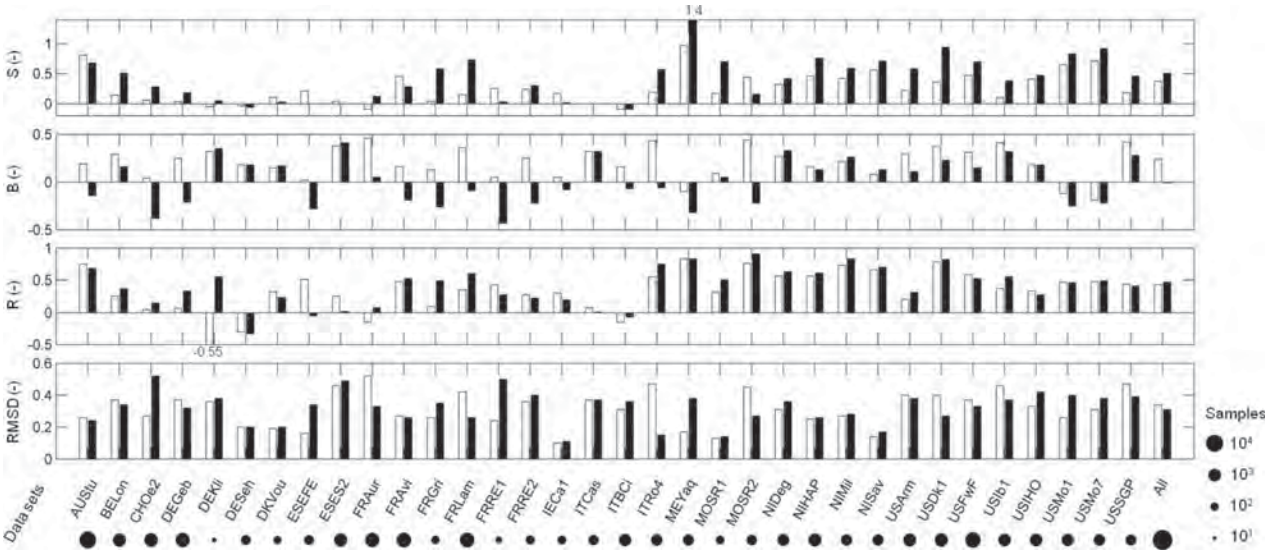


Figure 10. Same as for Figure 5 but for the new r_{ss} with $\theta_{1/2}$ estimated using the texture-based PTF and the mean $\Delta\theta_{1/2}^{-1}$ of all sites with $n > 500$.

Values are presented for each site separately, and additionally, the weighted mean (weighted by the number of data samples per site) is indicated. With respect to the latter, the RMSD is 0.31 instead of 0.34, R is 0.47 instead of 0.43, B is −0.01 instead of 0.24 and S is 0.51 instead of 0.37 for the texture-based new and S92 formulation, respectively. As in case of site-specific $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ parameters (Figure 5), the new r_{ss} formulation outperforms the S92 r_{ss} formulation when using $\theta_{1/2}$ estimated by means of the texture-based PTF together with the mean $\Delta\theta_{1/2}^{-1}$ of all sites with $n > 500$. To assess the information provided by sand fraction and clay fraction separately and by multi-regressional use of the two in terms of SEE, metrics are also computed using the PTFs $\theta_{1/2}(f_{clay})$ in equation (26) and $\theta_{1/2}(f_{sand})$ in equation (27). The (weighted mean) RMSD is 0.32 and 0.34, R is 0.47 and 0.46, B is 0.03 and 0.05, and S is 0.52 and 0.49 for $\theta_{1/2}(f_{clay})$ and $\theta_{1/2}(f_{sand})$, respectively. Consistent with the assessment of the different PTFs per se, SEE estimations using PTFs based on either clay or sand fraction provide relatively similar results, while the PTF based on both $\theta_{1/2}(f_{clay}, f_{sand})$ in equation (28) still provides best SEE estimates.

5.3. Comparison With Common Evaporation Models

The PTFs of equations (28) and (29) are compared in terms of SEE estimates with the PTFs of four common evaporation models. Table 3 reports the RMSD, B, R and S between simulated and observed SEE for ISBA, H-TESSEL, and CLM (version 4.5) evaporation modules, and for S92 and new texture-based (texture-based $\theta_{1/2}$ and mean $\Delta\theta_{1/2}^{-1}$) r_{ss} formulations. In each case, statistics are provided for five different clay fraction ranges: $f_{clay} < 0.10$, $0.10 \le f_{clay} < 0.20$, $0.20 \le f_{clay} < 0.30$, $0.30 \le f_{clay} < 0.40$ and $f_{clay} \ge 0.40$. The RMSD and B are systematically improved by the new r_{ss}. Among the five models, the minimum and maximum

Table 3. Root Mean Square Difference (RMSD), Mean Bias (B), Correlation Coefficient (R) and Slope of the Linear Regression (S) Between Simulated and Observed SEE for the ISBA (ISB), H-TESSEL (HTE), CLM (CLM) Evaporation Modules as Well as the S92 and New (in Bold) Texture-Based r_{ss} Formulations, for Different Ranges of Clay Fraction: $f_{clay} < 0.10$ (1), $0.10 \le f_{clay} < 0.20$ (2), $0.20 \le f_{clay} < 0.30$ (3), $0.30 \le f_{clay} < 0.40$ (4) and $f_{clay} \ge 0.40$ (5)

	RMSD					B					R					S				
f_{clay} range	ISB	HTE	CLM	S92	**New**	ISB	HTE	CLM	S92	**New**	ISB	HTE	CLM	S92	**New**	ISB	HTE	CLM	S92	**New**
1	0.41	0.49	0.39	0.32	**0.30**	0.29	0.09	0.20	0.16	**0.13**	0.41	0.51	0.53	0.57	**0.64**	0.21	0.89	0.64	0.49	**0.65**
2	0.42	0.48	0.42	0.31	**0.30**	0.34	0.18	0.18	0.19	**−0.06**	0.58	0.53	0.60	0.66	**0.62**	0.45	0.92	0.94	0.60	**0.73**
3	0.42	0.51	0.51	0.38	**0.38**	0.30	0.34	0.33	0.26	**0.16**	0.41	0.36	0.39	0.54	**0.50**	0.27	0.43	0.49	0.47	**0.59**
4	0.53	0.57	0.56	0.41	**0.32**	0.41	0.39	0.37	0.28	**−0.08**	0.23	0.28	0.33	0.34	**0.43**	0.19	0.39	0.48	0.27	**0.44**
5	0.42	0.48	0.48	0.38	**0.42**	0.28	0.32	0.30	0.20	**−0.09**	0.05	0.11	0.14	0.20	**0.36**	0.02	0.07	0.12	0.12	**0.46**
All	0.44	0.51	0.47	0.36	**0.34**	0.32	0.26	0.28	0.22	**0.01**	0.34	0.36	0.40	0.46	**0.51**	0.23	0.54	0.53	0.39	**0.58**

overall RMSD is estimated as 0.34 and 0.51 for the new r_{ss} and H-TESSEL, and the minimum and maximum overall B is estimated as 0.01 and 0.32 for the new r_{ss} and ISBA, respectively. Regarding the correlation with SEE measurements, the new r_{ss} outperforms the other four models with an overall R estimated as 0.51 compared to 0.34, 0.36, 0.40 and 0.46 for ISBA, H-TESSEL, CLM and S92 model, respectively. Note that S92 has a slightly larger R (0.54–0.66 versus 0.50–0.62) than the new r_{ss} for $0.1 \leq f_{clay} < 0.3$. This is probably due to a slight increase in the uncertainty in $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ attributed to the PTFs of equations (28) and (29) applied to a limited range of soil texture ($0.1 \leq f_{clay} < 0.3$), over which the S92 r_{ss} formulation is deemed acceptable. The overall S between simulated and observed SEE is 0.23, 0.55, 0.53, 0.39 and 0.58, for ISBA, H-TESSEL, CLM, S92 and the new r_{ss}, respectively. S is systematically closer to 1 with the new r_{ss} than with S92 model. The relatively good overall perfomance of H-TESSEL and CLM is attributed to a S close to 1 for low clay content ($f_{clay} < 0.2$), while the S for both H-TESSEL and CLM decreases strongly for larger clay fractions down to $\sim$0.1 for $f_{clay} > 0.4$. For the entire texture range considered, the PTFs of equations (28) and (29) are more robust in terms of SEE estimates than the PTFs of the other four evaporation models.

The evaluation of ISBA, H-TESSEL, and CLM evaporation modules and S92 r_{ss} formulation highlights a significant bias in simulated SEE, especially for soils with a $f_{clay} > 0.2$. These four models were not derived from the data set used in the paper to derive the PTFs of equations (28) and (29), which most likely contributes to the better results of the new r_{ss} formulation. The point is that the parameters of the ISBA, H-TESSEL and CLM evaporation modules have pre-set values and, to date, there is no PTF for the *A* and *B* parameters in S92. Systematic biases in modeled SEE can also result from differing depth of the top soil layer used to compute evaporation compared to the observation depth. The depth of the top soil layer is 1 cm, 1.75 cm, 5 cm and 7 cm in ISBA [*Parrens et al*., 2014], CLM [*Tang and Riley*, 2013a], S92 [*Sellers et al*., 1992] and H-TESSEL [*Albergel et al*., 2012], respectively. Several studies have addressed the inconsistency of the sensing depth of soil moisture observations (about 0–5 cm in this study) with the top soil layer of land-surface models [e.g., *Parrens et al*., 2014]. For instance, the soil layer used to calibrate the S92 r_{ss} formulation is 0–5 cm in *Sellers et al*. [1992] and 0–1 cm in *Van de Griend and Owe* [1994], resulting in quite distinct values of *A* and *B*. In the same vein, *Merlin et al*. [2011] investigated the effect of the top soil layer thickness on the exponent *P* of the SEE formulation derived from *Lee and Pielke* [1992]. They found that *P* is an increasing (and quasi linear) function of the top soil layer thickness, so that for a given θ value, SEE is a decreasing function of the soil moisture sensing depth. The shallow depth (1 cm and 1.75 cm) of the top soil layer in ISBA and CLM may thus be (partly) responsible for the models overestimation. Given that in situ measurements are usually available in the 0–5 cm soil layer or deeper, the models that use a soil layer shallower than the 5 cm depth are difficult to evaluate, even though their validity over a wide range of soil types needs to be checked.

6. Summary and Perspectives

A meta-analysis data-driven approach is developed to represent SEE over a large range of soil and atmospheric conditions. SEE is modeled using a soil resistance (r_{ss}) formulation based on surface soil moisture (θ) and two resistance parameters $r_{ss,ref}$ and $\theta_{efolding}$. The data-driven approach aims to express both parameters as a function of observable data including meteorological forcing, cut-off soil moisture value $\theta_{1/2}$ at which SEE=0.5, and first derivative of SEE at $\theta_{1/2}$, named $\Delta\theta_{1/2}^{-1}$. An analytical relationship between $(r_{ss,ref}; \theta_{efolding})$ and $(\theta_{1/2}; \Delta\theta_{1/2}^{-1})$ is first built by running a soil energy balance model for two extreme conditions with $r_{ss} = 0$ and $r_{ss} \sim \infty$ from meteorological data solely, and by approaching the middle point from the two (wet and dry) reference points. Two different methods are then investigated to estimate the pair $(\theta_{1/2}; \Delta\theta_{1/2}^{-1})$ either from the time series of SEE and θ observations for a given site, or using the soil texture information for all sites.

The new model is tested across a bare soil database composed of more than 30 sites around the world, a clay fraction range of 0.02–0.56, a sand fraction range of 0.05–0.92, and about 30,000 acquisition times between 8 am and 6 pm local time. In an effort to test the regionalization capabilities of the model using readily available data, a parameterization of $\theta_{1/2}$ is proposed as a PTF of clay and sand percentages separately as well as using both in multi-regressional fashion, and $\Delta\theta_{1/2}^{-1}$ is set to a constant mean value for all sites with $n > 500$. The correlation coefficient between modeled and retrieved $\theta_{1/2}$ is 0.76 (absolute value) for both clay-based and sand-based PTFs, while the multilinear regression of $\theta_{1/2}$ with both clay and sand fractions (texture-based PTF) improves the correlation coefficient (0.81).

The new PTF-based r_{ss} model is compared in terms of SEE estimates with the PTFs of the evaporation modules of the ISBA, H-TESSEL, CLM surface schemes as well as the S92 r_{ss} formulation. All models are forced by the same input data set including meteorological data, texture information, and the near-surface (mostly 0–5 cm depth) soil moisture observations. The SEE simulated by ISBA, H-TESSEL, CLM and S92 models generally overestimates observations, especially for soils with a clay fraction larger than 0.2. In this texture range, the overestimation (about 0.30–0.40) is larger for ISBA, H-TESSEL and CLM, while the S92 r_{ss} formulation tends to reduce the mean bias (about 0.20–0.30) between modeled and observed SEE. The new texture-based r_{ss} formulation reduces the mean bias (0.0 in average) for all clay fraction classes. Moreover, the nonlinearities of the SEE(θ) relationship are relatively well represented by the new texture-based r_{ss} across the entire texture range. The shallow depth (1 cm and 1.75 cm) of the top soil layer in the ISBA and CLM models compared to the observation depth may be (partly) responsible for the models overestimation. Nonetheless, the ad hoc nature of the evaporation formulations in ISBA, H-TESSEL and CLM does not guarantee (in the absence of consistent validation) their validity over a wide range of soil types.

While the r_{ss} formulation developed in this paper is mostly semi-empirical, the strength of the approach relies on the capability to calibrate its parameters ($\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$) from observable variables (SEE, θ, and meterological data). Specifically, four main benefits can be identified for future researches and applications:

1. the soil evaporation formulation as a function of r_{ss} has clear physical meaning, and thus, enables the implementation of the new evaporation model in a range of physically based land-surface models [*Pitman*, 2003]. Moreover, the SEE formulation of soil evaporation is fully consistent with the evaporation modules of operational models like the FAO-56 dual crop approach [*Allen*, 2000; *Lhomme et al*., 2015].
2. the proposed modeling framework is generic. It can be applied to characterize the variability of $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ as a function of soil texture as it done in this paper. It can also be used to represent other variability factors such as the presence of stubble or mulch [*Sakaguchi and Zeng*, 2009], soil heterogeneity [*Or et al*., 2013], soil roughness, and shrinkage cracks in clayey soils. Further research is needed to account for the impact of the (seasonal, daily, instantaneous) variability of evaporative demand on $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ through the time varying moisture profile in the top soil layer [*Merlin et al*., 2011].
3. such a meta-analysis data-driven approach is complementary to the upward modeling approaches based on fine physical knowledge and discretization of the soil layer. In particular, a key issue would be to interpret the variability of semi-empirical (but observed) $\theta_{1/2}$ and $\Delta\theta_{1/2}^{-1}$ in terms of the physical (but poorly observed in real field conditions) SHPs. Physically based soil water diffusion models [e.g., *Tang and Riley*, 2013b] will be very helpful in that direction.
4. given that a significant correlation exists between $\theta_{1/2}$ and sand and clay fractions, one could imagine a remote sensing approach for estimating surface soil texture from multi-sensor/multi-spectral remote sensing. In practice, several issues will need to be addressed beforehand such as the estimation of SEE from thermal infrared data [*Chanzy et al*., 1995; *Stefan et al*., 2015], the downscaling of microwave-derived θ [e.g., *Merlin et al*., 2013], and the partitioning between soil evaporation and plant transpiration from available remote sensing data [e.g., *Merlin et al*., 2014].

Appendix A: PTF-Derived SHPs

Soil moisture at field capacity is estimated as in *Noilhan and Mahfouf* [1996]:

$$\theta_{fc}=0.089\times(100f_{clay})^{0.3496} \tag{A1}$$

with f_{clay} being the clay fraction.

The residual soil moisture is estimated as in *Brisson and Perrier* [1991]:

$$\theta_{res}=0.15f_{clay} \tag{A2}$$

The soil moisture at saturation is estimated as in *Cosby et al*. [1984]:

$$\theta_{sat}=0.489-0.126f_{sand} \tag{A3}$$

with f_{sand} being the sand fraction.

Parameterized air entry pressure (in mm of water) at θ_{sat} is estimated as in *Cosby et al.* [1984]:

$$\psi_{sat} = -10 \times \exp(1.88 - 1.31 f_{sand}) \tag{A4}$$

The Clapp and Hornberger parameter is estimated as in *Cosby et al.* [1984]:

$$b_{CH} = 2.91 + 15.9 f_{clay} \tag{A5}$$

Appendix B: Soil Energy Balance Model

The evaporation model solves the classical energy budget equation over bare soil:

$$LE = Rn - G - H \tag{B1}$$

with LE (W m^{-2}) being the soil latent heat flux, H (W m^{-2}) the soil sensible heat flux, Rn (W m^{-2}) the soil net radiation and G (W m^{-2}) the ground conduction at 5 cm depth. Soil net radiation is expressed as:

$$Rn = (1-a)R_g + \epsilon(R_a - \sigma T^4) \tag{B2}$$

with a being the soil albedo (set to 0.20), R_g (W m^{-2}) the incoming solar radiation, ϵ the soil emissivity (set to 0.97), R_a (W m^{-2}) the atmospheric longwave radiation, σ (Wm^{-2}K^{-4}) the Stephan-Boltzmann constant and T (K) the soil skin temperature. Downward atmospheric radiation at ground level is expressed as:

$$R_a = \epsilon_a \sigma T_a^4 \tag{B3}$$

with ϵ_a being the effective atmospheric emissivity, and T_a (K) the air temperature. The emissivity of clear skies is estimated as in *Brutsaert* [1975]:

$$\epsilon_a = 0.553 (e_a/100)^{1/7} \tag{B4}$$

with:

$$e_a = e_{sat}(T_a)(h_a/100) \tag{B5}$$

with h_a (%) being the air relative humidity and:

$$e_{sat}(T_a) = 611 \exp[17.27 (T_a - 273.2)/(T_a - 35.9)] \tag{B6}$$

with T_a in K.

Ground conduction is estimated as a fraction of soil net radiation [*Choudhury et al.*, 1987; *Kustas and Daughtry*, 1990]:

$$G = C_G Rn \tag{B7}$$

with C_G a coefficient. Sensible heat flux is expressed as:

$$H = \rho C_P \frac{T - T_a}{r_{ah}} \tag{B8}$$

with the aerodynamic resistance being estimated as in *Choudhury et al.* [1986]:

$$r_{ah} = \frac{r_{ah0}}{(1+Ri)^{\eta}} \tag{B9}$$

with r_{ah0} (s m^{-1}) being the neutral aerodynamic resistance, and Ri the Richardson number which represents the importance of free versus forced convection, and η a coefficient set to 0.75 in unstable conditions ($T > T_a$) and to 2 in stable conditions ($T < T_a$). The neutral r_{ah0} is computed as:

$$r_{ah0} = \frac{1}{k^2 u_a} \left[\ln\left(\frac{Z}{z_{0m}}\right) \right]^2 \tag{B10}$$

with k being the von Karman constant, u_a (m s^{-1}) the wind speed measured at the reference height Z (m) and z_{0m} (m) the momentum soil roughness. At all sites, z_{0m} is set to 0.001 m [*Yang et al.*, 2008; *Stefan et al.*, 2015]. The Richardson number is computed as:

$$Ri = \frac{5gZ(T - T_a)}{T_a u_a^2} \tag{B11}$$

with g (m s^{-2}) being the gravitational constant.

The energy balance equation (B1) is solved by initializing the surface soil temperature $T = T_a$, and by looking for the value of T which minimizes the cost function $F(T)$:

$$F(T) = (LE + H - Rn + G)^2 \tag{B12}$$

with LE being expressed as in equations (2)–(4) for the r_{ss}, α and β formulation, respectively.

Appendix C: Derivation of $\theta_{efolding}$

$\theta_{efolding}$ is derived by applying the constraint $\partial \mathrm{SEE}/\partial\theta(\theta_{1/2}) = \Delta\theta_{1/2}^{-1}$ (equation (17)). The first derivative of SEE is:

$$\frac{\partial \mathrm{SEE}}{\partial \theta} = \frac{r_{ah,wet}}{e_{sat}(T_{wet}) - e_a} \times \left[\frac{\dot{e}_{sat}(T)}{r_{ss} + r_{ah}} \times \frac{\partial T}{\partial \theta} + \frac{(e_{sat}(T) - e_a) r_{ss}}{(r_{ss} + r_{ah})^2} \times \frac{1}{\theta_{efolding}} \right] \tag{C1}$$

with $\dot{e}_{sat}(T)$ being the derivative of saturated vapor pressure with respect to T and $\partial T/\partial\theta$ the derivative of T with respect to θ. As $\partial T/\partial\theta$ is unknown, additional information is needed via the soil energy balance model expanded from equation (B1):

$$\frac{\rho C_P}{\gamma} \frac{e_{sat}(T) - e_a}{r_{ss} + r_{ah}} + \rho C_P \frac{T - T_a}{r_{ah}} = (1 - C_G)[(1 - a)R_g + \epsilon(R_a - \sigma T^4)] \tag{C2}$$

By applying the first derivative to equation (C2), it comes:

$$\frac{\rho C_P \dot{e}_{sat}(T)}{\gamma(r_{ss} + r_{ah})} \frac{\partial T}{\partial \theta} + \frac{\rho C_P (e_{sat}(T) - e_a) r_{ss}}{\gamma (r_{ss} + r_{ah})^2} \times \frac{1}{\theta_{efolding}} + \frac{\rho C_P}{r_{ah}} \frac{\partial T}{\partial \theta} = -4(1 - C_G)\epsilon\sigma T^3 \frac{\partial T}{\partial \theta} \tag{C3}$$

and then:

$$\frac{\partial T}{\partial \theta} = \frac{f(\theta)}{\theta_{efolding}} \tag{C4}$$

with $f(\theta)$ being defined in equation (22). Finally, an expression of $\theta_{efolding}$ is obtained in equation (21) by inserting the above expression of $\partial T/\partial\theta$ in equation (C1).

Acknowledgments

This study was supported by the French Agence Nationale de la Recherche (MIXMOD-E project, ANR-13-JS06-0003-01) and the European Commission Horizon 2020 Programme for Research and Innovation (H2020) in the context of the Marie Skłodowska-Curie Research and Innovation Staff Exchange (RISE) action (REC projet, grant agreement no: 645642). We acknowledge the following sites of the European Flux Database (BELon, CHOe2, DEGeb, DEKli, DESeh, ESES2, FRAur, FRAvi, FRGri, FRLam, IECa1, ITBCi, ITCas, ITRo4), the following AmeriFlux sites (USArm, USDk1, USFwf, USIb1), the following OzFlux site (AUStu), the following long term observatories (AMMA, HOBE and SudMed), and the following short term intensive field campaigns (EFEDA, ReSeDa, Yaqui'08, HAPEX-Sahel, IHOP, Monsoon'90 and SGP'97) for their data records. This work utilized flux tower data collected by grants funded by the Australian Research Council (DP0772981, DP130101566, DP0344744 and LP100100073). Beringer is funded under an ARC Future Fellowship (FT110100602). Some support for OzFlux is provided through the Australia Terrestrial Ecosystem Research Network (TERN) (http://www.tern.org.au). The DMI (Danish Meteorological Institute) and Rasmus Jensen from HOBE are gratefully acknowledged. Funding for AmeriFlux data resources was provided by the U.S. Department of Energy Office of Science, Biological and Environmental Research and Atmospheric Radiation Measurement Program.

References

Alaoui, A., and B. Goetz (2008), Dye tracer and infiltration experiments to investigate macropore flow, *Geoderma*, *144*, 279–286, doi: 10.1016/j.geoderma.2007.11.020.

Albergel, C., G. Balsamo, P. de Rosnay, J. Muñoz-Sabater, and S. Boussetta (2012), A bare ground evaporation revision in the ECMWF land-surface scheme: Evaluation of its impact using ground soil moisture and satellite microwave data, *Hydrol. Earth Syst. Sci.*, *16*, 3607–3620, doi:10.5194/hess-16-3607-2012.

Allen, R. G. (2000), Using the FAO-56 dual crop coefficient method over an irrigated region as part of an evapotranspiration intercomparison study, *J. Hydrol.*, *229*(1-2), 27–41, doi:10.1016/S0022-1694(99)00194-8.

Beringer, J., et al. (2011), SPECIAL-–Savanna Patterns of Energy and Carbon Integrated Across the Landscape, *Bull. Am. Meteorol. Soc.*, *92*, 1467–1485.

Béziat, P., E. Ceschia, and G. Dedieu (2009), Carbon balance of a three crop succession over two cropland sites in South West France, *Agric. For. Meteorol.*, *149*(10), 1628–1645, doi:10.1016/j.agrformet.2009.05.004.

Bircher, S., N. Skou, K. H. Jensen, J. P. Walker, and L. Rasmussen (2012), A soil moisture and temperature network for SMOS validation in Western Denmark, *Hydrol. Earth Syst. Sci.*, *16*, 1445–1463, doi:10.5194/hess-16-1445-2012.

Bittelli, M., F. Ventura, G. S. Campbell, R. L. Snyder, F. Gallegati, and P. R. Pisa (2008), Coupling of heat, water vapor, and liquid water fluxes to compute evaporation in bare soils, *J. Hydrol.*, *362*, 191–205.

Braud, I., J. Noilhan, P. Mascart, P. Bessemoulin, R. Haverkamp, and M. Vauclin (1993), Bare ground surface heat and water exchanges under dry conditions: Observations and parameterization, *Boundary Layer Meteorol.*, *66*, 173–200.

Brisson, N., and A. Perrier (1991), A semiempirical model of bare soil evaporation for crop simulation models, *Water Resour. Res.*, *27*(5), 719–727.

Brutsaert, W. (1975), On a derivable formula for long-wave radiation from clear skies, *Water Resour. Res.*, *11*(5), 742–744.

Brutsaert, W. (2014), Daily evaporation from drying soil: Universal parameterization with similarity, *Water Resour. Res.*, *50*, 3206–3215, doi: 10.1002/2013WR014872.

Cahill, A. T., M. B. Parlange, T. J. Jackson, P. O'Neill, and T. J. Schmugge (1999), Evaporation from nonvegetated surfaces: Surface aridity methods and passive microwave remote sensing, *J. Appl. Meteorol.*, *38*, 1346–1351.

Chanzy, A., and L. Bruckler (1993), Significance of soil surface moisture with respect to daily bare soil evaporation, *Water Resour. Res.*, *29*(4), 1113–1125.
Chanzy, A., L. Bruckler, and A. Perrier (1995), Soil evaporation monitoring: A possible synergism of microwave and infrared remote sensing, *J. Hydrol.*, *165*, 235–259.
Chanzy, A., M. Mumen, and G. Richard (2008), Accuracy of top soil moisture simulation using a mechanistic model with limited soil characterization, *Water Resour. Res.*, *44*, W03432, doi:10.1029/2006WR005765.
Chirouze, J., et al. (2014), Inter-comparison of four remote sensing based surface energy balance methods to retrieve surface evapotranspiration and water stress of irrigated fields in semi-arid climate, *Hydrol. Earth Syst. Sci.*, *18*, 1165–1188, doi:10.5194/hess-18-1165-2014.
Choudhury, B., R. J. Reginato, and S. B. Idso (1986), An analysis of infrared temperature observations over wheat and calculation of latent heat flux, *Agric. For. Meteorol.*, *37*, 75–88.
Choudhury, B., S. B. Idso, and R. J. Reginato (1987), Analysis of an empirical model for soil heat flux under a growing wheat crop for estimating evaporation by an infrared-temperature based energy balance equation, *Agric. For. Meteorol.*, *39*(4), 283–297.
Clapp, R. B., and G. M. Hornberger (1978), Empirical equations for some soil hydraulic properties, *Water Resour. Res*, *14*(4), 601–604.
Cosby, B., G. Hornberger, R. Clapp, and T. Ginn (1984), A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils, *Water Resour. Res*, *20*(6), 682–690.
Crow, W. T., W. P. Kustas, and J. H. Prueger (2008), Monitoring root-zone soil moisture through the assimilation of a thermal remote sensing-based soil moisture proxy into a water balance model, *Remote Sens. Environ.*, *112*, 1268–1281.
Dekic, L. J., D. T. Mihailovic, and B. Rajkovic (1995), A study of the sensitivity of bare soil evaporation schemes to soil surface wetness, using the coupled soil moisture and surface temperature prediction model, BARESOIL, *Meteorol. Atmos. Phys.*, *55*, 101–112.
Denef, K., I. D. Galdo, A. Venturi, and M. F. Cotrufo (2013), Assessment of soil C and N stocks and fractions across 11 European soils under varying land uses, *Open J. Soil Sci.*, *3*, 297–313.
Desborough, C. E., A. J. Pitman, and P. Irannejad (1996), Analysis of the relationship between bare soil evaporation and soil moisture simulated by 13 land surface schemes for a simple non-vegetated site, *Global Planet. Change*, *13*, 47–56.
Dickinson, R. E., A. Henderson-Sellers, P. J. Kennedy, and M. F. Wilson (1986), Biosphere-Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model, Technical Note NCAR/TN-275-+STR, National Center for Atmospheric Research, Boulder, Colo., doi:10.5065/D6668B58.
Dore, S., M. Montes-Helu, S. Hart, B. Hungate, G. Koch, J. Moon, A. Finkral, and T. Kolb (2012), Recovery of southwestern ponderosa pine ecosystem carbon and water fluxes from thinning and stand replacing fire, *Global Change Biol.*, *18*(10), 3171–3185.
ECMWF (2014), *IFS Documentation, Part IV: Physical Processes*, vol. Cy40r1, Eur. Cent. for Medium-Range Weather Forecasts, Reading, U. K.
Er-Raki, S., A. Chehbouni, N. Guemouria, B. Duchemin, J. Ezzahar, and R. Hadria (2007), Combining FAO-56 model and ground-based remote sensing to estimate water consumptions of wheat crops in a semi-arid region, *Agric. Water Manage.*, *87*, 41–54.
Er-Raki, S., A. Chehbouni, G. Boulet, and D. Williams (2010), Using the dual approach of FAO-56 for partitioning ET into soil and plant components for olive orchards in a semi-arid region, *Agric. Water Manage.*, *97*(11), 1769–1778.
Feddes, R. A., et al. (2001), Modeling root water uptake in hydrological and climate models, *Bull. Am. Meteorol. Soc.*, *82*(12), 2797–2809.
Fischer, M. L., D. P. Billesbach, J. A. Berry, W. J. Riley, and M. S. Torn (2007), Spatiotemporal variations in growing season exchanges of co2, h2o, and sensible heat in agricultural fields of the southern great plains, *Earth Interact.*, *11*(17), 1–21.
Fritton, D. D., D. Kirkham, and R. H. Shaw (1967), Soil water and chloride redistribution under various evaporation potentials, *Soil Sci. Soc. Am. Proc.*, *31*, 599–603.
Garrigues, S., et al. (2015), Evaluation of land surface model simulations of evapotranspiration over a 12-year crop succession: Impact of soil hydraulic and vegetation properties, *Hydrol. Earth Syst. Sci.*, *19*(7), 3109–3131.
Gentine, P., D. Entekhabi, A. Chehbouni, G. Boulet, and B. Duchemin (2007), Analysis of evaporative fraction diurnal behaviour, *Agric. For. Meteorol.*, *143*, 13–29.
Good, S. P., D. Noone, and G. Bowen (2015), Hydrologic connectivity constrains partitioning of global terrestrial water fluxes, *Science*, *349*(6244), 175–177.
Heitman, J. L., X. Xiao, R. Horton, and T. J. Sauer (2008), Sensible heat measurements indicating depth and magnitude of subsurface soil water evaporation, *Water Resour. Res.*, *44*, W00D05, doi:10.1029/2008WR006961.
Jarlan, L., et al. (2015), Remote Sensing of Water Resources in the semi-arid Mediterranean areas: The Joint International Laboratory TREMA, *Int. J. Remote Sens.*, *36*(19–20), 4879–4917.
Kindler, R., et al. (2011), Dissolved carbon leaching from soil is a crucial component of the net ecosystem carbon balance, *Global Change Biol.*, *17*(2), 1167–1185, doi:10.1111/j.1365-2486.2010.02282.x.
Komatsu, T. S. (2003), Towards a robust phenomenological expression of evaporation efficiency for unsaturated soil surfaces, *J. Appl. Meteorol.*, *42*, 1330–1334.
Kool, D., N. Agam, N. Lazarovitch, J. L. Heitman, T. J. Sauer, and A. Ben-Gal (2014), A review of approaches for evapotranspiration partitioning, *Agric. For. Meteorol.*, *184*, 56–70.
Kustas, W. P., and C. S. T. Daughtry (1990), Estimation of the soil heat flux/net radiation ratio from spectral data, *Agric. For. Meteorol.*, *49*, 205–223.
Kustas, W. P., et al. (1991), An interdisciplinary field study of the energy and water fluxes in the atmosphere-biosphere system over semi-arid rangelands: Description of some preliminary results, *Bull. Am. Meteorol. Soc.*, *72*, 1683–1705.
Kustas, W. P., X. Zhan, and T. J. Schmugge (1998), Combining optical and microwave remote sensing for mapping energy fluxes in a semi-arid watershed, *Remote Sens. Environ.*, *64*, 116–131.
Kutsch, W. L., et al. (2010), The net biome production of full crop rotations in Europe, *Agric. Ecosyst. Environ.*, *139*, 336–345.
Lawrence, D. M., P. E. Thornton, K. W. Oleson, and G. B. Bonan (2007), The partitioning of evapotranspiration into transpiration, soil evaporation, and canopy evaporation in a GCM: Impacts on land-atmosphere interaction, *J. Hydrometeorol.*, *8*, 862–880.
Lee, T. J., and R. A. Pielke (1992), Estimating the soil surface specific humidity, *J. Appl. Meteorol.*, *31*, 480–484.
Lemone, M. A., F. Chen, J. G. Alfieri, R. H. Cuenca, Y. Hagimoto, P. Blanken, D. Niyogi, S. Kang, K. Davis, and R. L. Grossman (2007), NCAR/CU surface, soil, and vegetation observations during the international H2O Project 2002 field campaign, *Bull. Am. Meteorol. Soc.*, *88*, 65–81, doi:10.1175/BAMS-88-1-65.
Lhomme, J., N. Boudhina, M. Masmoudi, and A. Chehbouni (2015), Estimation of crop water requirements: Extending the one-step approach to dual crop coefficients, *Hydrol. Earth Syst. Sci.*, *19*, 3287–3299, doi:10.5194/hess-19-3287-2015.
Lintner, B., P. Gentine, K. Findell, and G. Salvucci (2015), The budyko and complementary relationships in an idealized model of large-scale land–atmosphere coupling, *Hydrol. Earth Syst. Sci.*, *19*(5), 2119–2131.
Mahfouf, J. F., and J. Noilhan (1991), Comparative study of various formulations of evaporation from bare soil using in situ data, *J. Appl. Meteorol.*, *30*, 1354–1365.

Mahrt, L., and H. Pan (1984), A two-layer model of soil hydrology, *Boundary Layer Meteorol.*, *29*, 1–20.
Manabe, S. (1969), Climate and the ocean circulation: I. The atmospheric circulation and the hydrology of the earth's surface, *Mon. Weather Rev.*, *97*(11), 739–774.
Marchesini, L. B., R. Casa, D. Papale, F. Pieruccetti, P. Stefani, N. Arriga, F. Mazzenga, and R. Valentini (2008), Evaluation of the greenhouse-gas reduction efficacy of a rapeseed (Brassica napa L.) biofuel crop and impact of agronomic techniques, in *EGU General Assembly 2008, Geophys. Res. Abstr.*, 10, EGU2008-A-11269.
Merlin, O. (2013), An original interpretation of the wet edge of the surface temperature-albedo space to estimate crop evapotranspiration (SEB-1S), and its validation over an irrigated area in northwestern Mexico, *Hydrol. Earth Syst. Sci.*, *17*, 3623–3637.
Merlin, O., and G. Chehbouni (2004), Different approaches in estimating heat flux using dual angle observations of radiative surface temperature, *Int. J. Remote Sens.*, *25*(1), 275–289.
Merlin, O., A. Al Bitar, V. Rivalland, P. Béziat, E. Ceschia, and G. Dedieu (2011), An analytical model of evaporation efficiency for unsaturated soil surfaces with an arbitrary thickness, *J. Appl. Meteorol. Climatol.*, *50*(2), 457–471, doi:10.1175/2010JAMC2418.1.
Merlin, O., M. J. Escorihuela, M. A. Mayoral, O. Hagolle, A. A. Bitar, and Y. Kerr (2013), Self-calibrated evaporation-based disaggregation of SMOS soil moisture: An evaluation study at 3 km and 100 m resolution in Catalunya, Spain, *Remote Sens. Environ.*, *130*, 25–38.
Merlin, O., J. Chirouze, A. Olioso, L. Jarlan, G. Chehbouni, and G. Boulet (2014), An image-based four-source surface energy balance model to estimate crop evapotranspiration from solar reflectance/thermal emission data (SEB-4S), *Agric. For. Meteorol.*, *184*, 188–203.
Mihailovic, D. T., B. Rajkovic, L. Dekic, R. A. Pielke, T. J. Lee, and Z. Ye (1995), The validation of various schemes for parameterizing evaporation from bare soil for use in meteorological models: A numerical study using in situ data, *Boundary Layer Meteorol.*, *76*, 259–289.
Milly, P. (1984), A simulation analysis of thermal effects on evaporation from soil, *Water Resour. Res.*, *20*(8), 1087–1098.
Moran, M. S., T. R. Clarke, Y. Inoue, and A. Vidal (1994), Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index, *Remote Sens. Environ.*, *49*, 246–263.
Nishida, K., R. R. Nemani, J. M. Glassy, and S. W. Running (2003), Development of an evapotranspiration index from Aqua/MODIS for monitoring surface moisture status, *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 493–501.
Noilhan, J., and J.-F. Mahfouf (1996), The ISBA land surface parameterisation scheme, *Global Planet. Change*, *13*, 145–159.
Noilhan, J., and S. Planton (1989), A simple parameterization of land surface processes for meteorological models, *Mon. Weather Rev.*, *117*, 536–549.
Norman, J. M., W. P. Kustas, and K. S. Humes (1995), Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature, *Agric. For. Meteorol.*, *77*, 263–293.
Novick, K., P. Stoy, G. Katul, D. Ellsworth, M. Siqueira, J. Juang, and R. Oren (2004), Carbon dioxide and water vapor exchange in a warm temperate grassland, *Oecologia*, *138*(2), 259–274.
Oleson, K. W., G.-Y. Niu, Z.-L. Yang, D. M. Lawrence, P. E. Thornton, P. J. Lawrence, R. Stockli, R. E. Dickinson, G. B. Bonan, and S. Levis (2007), *CLM3.5 Documentation*, Natl. Cent. for Atmos. Res., Boulder, Colo.
Oleson, K. W., et al. (2008), Improvements to the Community Land Model and their impact on the hydrological cycle, *J. Geophys. Res.*, *113*, G01021, doi:10.1029/2007JG000563.
Oleson, K. W., et al. (2013), Technical description of version 4.5 of the Community Land Model (CLM), *TN-503+STR*, Natl. Cent. for Atmos. Res., Boulder, Colo.
Olioso, A., et al. (2002), Monitoring energy and mass transfers during the Alpilles-ReSeDA experiment, *Agron. Sustainable Dev.*, *22*(6), 597–610, doi:10.1051/agro:2002051.
Or, D., P. Lehmann, E. Shahraeeni, and N. Shokri (2013), Advances in Soil Evaporation Physics - A Review, *Vadose Zone J.*, *12*(4), 1–16, doi:10.2136/vzj2012.0163.
Papale, D., et al. (2006), Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: Algorithms and uncertainty estimation, *Biogeosciences*, *3*, 571–583.
Parrens, M., J.-F. Mahfouf, A. L. Barbu, and J.-C. Calvet (2014), Assimilation of surface soil moisture into a multilayer soil model: Design and evaluation at local scale, *Hydrol. Earth Syst. Sci.*, *18*, 673–689, doi:10.5194/hess-18-673-2014.
Passerat de Silans, A. (1986), Transferts de masse et de chaleur dans un sol stratifié soumis une excitation atmosphèrique naturelle. Comparaison modèle expérience, 205 pp., PhD thesis, Inst. Natl. Polytechnique de Grenoble, Grenoble, France.
Pellarin, T., J. Laurent, B. Cappelaere, B. Decharme, L. Descroix, and D. Ramier (2009), Hydrological modelling and associated microwave emission of a semi-arid region in South-western Niger, *J. Hydrol.*, *375*, 262–272.
Philip, J. R., and D. A. de Vries (1957), Moisture movement in porous materials under temperature gradients, *Eos Trans. AGU*, *38*, 222–232.
Pitman, A. J. (2003), The evolution of, and revolution in, land surface schemes designed for climate models, *Int. J. Climatol.*, *23*(5), 479–510.
Porporato, A., F. Laio, L. Ridolfi, and I. Rodriguez-Iturbe (2001), Plants in water-controlled ecosystems: Active role in hydrologic processes and response to water stress: III. Vegetation water stress, *Adv. Water Resour.*, *24*(7), 725–744.
Prévot, L., R. Bernard, O. Taconet, and D. V. Madjar (1984), Evaporation from a bare soil evaluated using a bare soil water transfer model and remotely sensed surface soil moisture data, *Water Resour. Res.*, *20*(2), 311–316.
Robock, A., K. Y. Vinnikov, C. A. Schlosser, N. A. Speranskaya, and Y. Xue (1995), Use of midlatitude soil moisture and meteorological observations to validate soil moisture simulations with Biosphere and Bucket models, *J. Clim.*, *8*, 15–35.
Roerink, G. J., Z. Su, and M. Menenti (2000), S-SEBI: A simple remote sensing algorithm to estimate the surface energy balance, *Phys. Chem. Earth, Part B*, *25*(2), 147–157.
Sakaguchi, K., and X. Zeng (2009), Effects of soil wetness, plant litter, and under-canopy atmospheric stability on ground evaporation in the Community Land Model (CLM3.5), *J. Geophys. Res.*, *114*, D01107, doi:10.1029/2008JD010834.
Santanello, J. A., C. D. Peters-Lidard, M. E. Garcia, D. M. Mocko, M. A. Tischler, M. S. Moran, and D. Thoma (2007), Using remotely-sensed estimates of soil moisture to infer soil texture and hydraulic properties across a semi-arid watershed, *Remote Sens. Environ.*, *110*, 79–97.
Schlesinger, W. H., and S. Jasechko (2014), Transpiration in the global water cycle, *Agric. For. Meteorol.*, *189*, 115–117.
Sellers, P. J., M. D. Heiser, and F. G. Hall (1992), Relations between surface conductance and spectral vegetation indices at intermediate (100 m^2 to 15 km^2) length scales, *J. Geophys. Res.*, *97*(D17), 19,033–19,059.
Sellers, P. J., D. A. Randall, G. J. Collatz, J. A. Berry, C. B. Field, D. A. Dazlich, C. Zhang, G. D. Collelo, and L. Bounoua (1996), A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation, *J. Clim.*, *9*, 676–705.
Shahraeeni, E., P. Lehmann, and D. Or (2012), Coupling of evaporative fluxes from drying porous surfaces with air boundary layer: Characteristics of evaporation from discrete pores, *Water Resour. Res.*, *48*, W09525, doi:10.1029/2012WR011857.
Shuttleworth, W. J., and J. S. Wallace (1985), Evaporation from sparse canopies-an energy combination theory, *Q. J. R. Meteorol. Soc.*, *111*, 839–855.
Simmonds, L. P., and E. J. Burke (1999), Application of a coupled microwave, energy and water transfer model to relate passive microwave emission from bare soils to near-surface water content and evaporation, *Hydrol. Earth Syst. Sci.*, *3*(1), 31–38.

Sivapalan, M., G. Bloschl, L. Zhang, and R. Vertessy (2003), Downward approach to hydrological prediction, *Hydrol. Processes*, *17*, 2101–2111.
Smits, K. M., V. V. Ngo, A. Cihan, T. Sakaki, and T. H. Illangasekare (2012), An evaluation of models of bare soil evaporation formulated with different land surface boundary conditions and assumptions, *Water Resour. Res.*, *48*, W12526, doi:10.1029/2012WR012113.
Soarès, J., R. Bernard, O. Taconet, D. Vidal-Madjar, and A. Weill. (1988), Estimation of bare soil evaporation from airborne measurements, *J. Hydrol.*, *99*, 281–296.
Stefan, V. G., O. Merlin, S. Er-Raki, M.-J. Escorihuela, and S. Khabba (2015), Consistency between in situ, model-derived and high-resolution-image-based soil temperature endmembers: Towards a robust data-based model for multi-resolution monitoring of crop evapotranspiration, *Remote Sens.*, *7*(8), 10,444–10,479.
Stöckli, R., D. M. Lawrence, G.-Y. Niu, K. W. Oleson, P. E. Thornton, Z.-L. Yang, G. B. Bonan, A. S. Denning, and S. W. Running (2008), Use of FLUXNET in the Community Land Model development, *J. Geophys. Res.*, *113*, G01025, doi:10.1029/2007JG000562.
Su, Z. (2002), The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, *Hydrol. Earth Syst. Sci.*, *6*(1), 85–99.
Sun, S. F. (1982), Moisture and heat transport in a soil layer forced by atmospheric conditions, MS thesis, 72 pp., Univ. of Conn., Storrs.
Tang, J., and W. J. Riley (2013a), Impacts of a new bare-soil evaporation formulation on site, regional, and global surface energy and water budgets in CLM4, *J. Adv. Model. Earth Syst.*, *5*, 558–571, doi:10.1002/jame.20034.
Tang, J., and W. J. Riley (2013b), A new top boundary condition for modeling surface diffusive exchange of a generic volatile tracer: Theoretical analysis and application to soil evaporation, *Hydrol. Earth Syst. Sci.*, *17*, 873–893, doi:10.5194/hess-17-873-2013.
Timmermans, W. J., W. P. Kustas, M. C. Anderson, and A. N. French (2007), An intercomparison of the Surface Energy Balance Algorithm for Land (SEBAL) and the Two-Source Energy Balance (TSEB) modeling schemes, *Remote Sens. Environ.*, *108*, 369–384.
Twine, T. E., W. Kustas, J. Norman, D. Cook, P. Houser, T. Meyers, J. Prueger, P. Starks, and M. Wesely (2000), Correcting eddy-covariance flux underestimates over a grassland, *Agric. For. Meteorol.*, *103*(3), 279–300.
Van de Griend, A. A., and M. Owe (1994), Bare soil surface resistance to evaporation by vapor diffusion under semiarid conditions, *Water Resour. Res.*, *30*(2), 181–188.
Van den Hoof, C., E. Hanert, and P. L. Vidale (2011), Simulating dynamic crop growth with an adapted land surface model-–JULES-SUCROS: Model development and validation, *Agric. For. Meteorol.*, *151*, 137–153.
Van den Hurk, B. J. J. M., P. Viterbo, A. C. M. Beljaars, and A. K. Betts (2000), Offline validation of the ERA40 surface scheme, *European Centre for Medium-Range Weather Forecasts, Tech. Memo. 295*, Reading, U. K.
Vidale, L., and R. Stöckli (2005), Prognostic canopy air space solutions for land surface exchanges, *Theor. Appl. Climatol.*, *80*(2-4), 245–257.
Viterbo, P., and A. C. M. Beljaars (1995), An improved land surface parametrization scheme in the ECMWF model and its validation, *European Centre for Medium-Range Weather Forecasts, Tech. Rep. 75*, Res. Dep., ECMWF, Reading, U. K.
Wallace, J. S., S. J. Allen, J. H. C. Gash, C. J. Holwill, and C. R. Lloyd (1993), Components of the energy and water balance at the HAPEX-Sahel southern super-site, in *Proceedings of the Yokohama Symposium*, vol. 212, International Association of Hydrological Sciences Publ., pp. 365–371, Yokohama, Japan.
Walmsley, D. C., J. Siemens, R. Kindler, L. Kirwan, K. Kaiser, M. Saunders, M. Kaupenjohann, and B. A. Osborne (2011), Dissolved carbon leaching from an Irish cropland soil is increased by reduced tillage and cover cropping, *Agric. Ecosyst. Environ.*, *142*, 393–402.
Wang, L., S. P. Good, and K. K. Caylor (2014), Global synthesis of vegetation control on evapotranspiration partitioning, *Geophys. Res. Lett.*, *41*, 6753–6757, doi:10.1002/2014GL061439.
Wetzel, P. J., and J.-T. Chang (1988), Evapotranspiration from nonuniform surfaces: A first approach for short-term numerical weather prediction, *Mon. Weather Rev.*, *116*, 600–621.
Williams, D. G., et al. (2004), Evapotranspiration components determined by stable isotope, sap flow and eddy covariance technique, *Agric. For. Meteorol.*, *125*, 241–258.
Wood, E. F., D. P. Lettenmaier, and V. G. Zartarian (1992), A land-surface hydrology parameterization with subgrid variability for general circulation models, *J. Geophys. Res.*, *97*(D3), 2717–2728.
Wu, C., et al. (2012), An underestimated role of precipitation frequency in regulating summer soil moisture, *Environ. Res. Lett.*, *7*(2), 024011.
Yamanaka, T., A. Takeda, and J. Shimada (1998), Evaporation beneath the soil surface: Some observational evidence and numerical experiments, *Hydrol. Processes*, *12*, 2193–2203.
Yang, K., T. Koike, H. Ishikawa, J. Kim, X. Li, H. Liu, S. Liu, Y. Ma, and J. Wang (2008), Turbulent flux transfer over bare-soil surfaces: Characteristics and parameterization, *J. Appl. Meteorol. Climatol.*, *47*(1), 276–290.
Yang, Z.-L., R. E. Dickinson, W. J. Shuttleworth, and M. Shaikh (1998), Treatment of soil, vegetation and snow in land surface models: A test of the biosphere-atmosphere transfer scheme with the hapex-mobilhy, abracos and russian data, *J. Hydrol.*, *212-213*, 109–127.
Ye, Z., and R. A. Pielke (1993), Atmospheric parameterization of evaporation from non-plant-covered surfaces, *J. Appl. Meteorol.*, *32*, 1248–1258.
Zribi, M., A. Chahbi, M. Shabou, Z. Lili-Chabaane, B. Duchemin, N. Baghdadi, R. Amri, and A. Chehbouni (2011), Soil surface moisture estimation over a semi-arid region using ENVISAT ASAR radar data for soil evaporation evaluation, *Hydrol. Earth Syst. Sci.*, *15*, 345–358.

Remote Sensing of Environment 180 (2016) 361–376

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

SMOS disaggregated soil moisture product at 1 km resolution: Processor overview and first validation results

B. Molero [a,*], O. Merlin [a,b], Y. Malbéteau [a], A. Al Bitar [a], F. Cabot [a], V. Stefan [a], Y. Kerr [a], S. Bacon [a], M.H. Cosh [c], R. Bindlish [c], T.J. Jackson [c]

[a] Centre d'Etudes Spatiales de la Biosphère, Toulouse, France
[b] Université Cadi Ayyad, Marrakech, Morocco
[c] USDA ARS Hydrology and Remote Sensing Laboratory, USA

ARTICLE INFO

Article history:
Received 30 July 2015
Received in revised form 12 February 2016
Accepted 19 February 2016
Available online 23 March 2016

Keywords:
Disaggregation
Soil moisture
Processor
SMOS
CATDS
Level 4

ABSTRACT

The SMOS (Soil Moisture and Ocean Salinity) mission provides surface soil moisture (SM) maps at a mean resolution of ~50 km. However, agricultural applications (irrigation, crop monitoring) and some hydrological applications (floods and modeling of small basins) require higher resolution SM information. In order to overcome this spatial mismatch, a disaggregation algorithm called Disaggregation based on Physical And Theoretical scale Change (DISPATCH) combines higher-resolution data from optical/thermal sensors with the SM retrieved from microwave sensors like SMOS, producing higher-resolution SM as the output. A DISPATCH-based processor has been implemented for the whole globe (emerged lands) in the Centre Aval de Traitement des Données SMOS (CATDS), the French data processing center for SMOS Level 3 products. This new CATDS Level-4 Disaggregation processor (C4DIS) generates SM maps at 1 km resolution. This paper provides an overview of the C4DIS architecture, algorithms and output products. Differences with the original DISPATCH prototype are explained and major processing parameters are presented. The C4DIS SM product is compared against L3 and *in situ* SM data during a one year period over the Murrumbidgee catchment and the Yanco area (Australia), and during a four and a half year period over the Little Washita and the Walnut Gulch watersheds (USA). The four validation areas represent highly contrasting climate regions with different landscape properties. According to this analysis, the C4DIS SM product improves the spatio-temporal correlation with *in situ* measurements in the semi-arid regions with substantial SM spatial variability mainly driven by precipitation and irrigation. In sub-humid regions like the Little Washita watershed, the performance of the algorithm is poor except for summer, as result of the weak moisture-evaporation coupling. Disaggregated products do not succeed to have and additional benefit in the Walnut Gulch watershed, which is also semi-arid but with well-drained soils that are likely to cancel the spatial contrast needed by DISPATCH. Although further validation studies are still needed to better assess the performance of DISPATCH in a range of surface and atmospheric conditions, the new C4DIS product is expected to provide satisfying results over regions having medium to high SM spatial variability.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Soil moisture (SM) is an essential component of the water cycle that impacts infiltration, runoff and evaporation processes. In addition, it modulates the energy exchange as well as the carbon exchange at the land surface (Daly & Porporato, 2005). SM has influence over a range of spatial scales: the climatic (Douville, 2004; Laio, Porporato, Ridolfi, & Rodríguez-Fernández, 2002), the meteorological (Dirmeyer, 2000; Drusch, 2007), the hydrological (Chen, Crow, Starks, & Moriasi, 2011; Draper, Reichle, De Lannoy, & Liu, 2012), the parcel and the local scale (Guérif & Duke, 2000).

Current satellite missions provide surface SM observations at large scales on a global basis. Passive microwave L-band observations are widely used for surface SM retrievals, but in practice they constrain the resolution of the retrievals to 30–60 km (Kerr & Njoku, 1990; Njoku & Entekhabi, 1996; Schmugge, 1998) with current technology. The Soil Moisture Ocean Salinity (SMOS) mission, launched in November 2009, incorporates an interferometric radiometer at L-band (1.4 GHz) and provides SM with a resolution of 30–55 km and a sensing depth of 3–5 cm (Kerr et al., 2001, 2010). SMOS Level 2 (L2) and Level 3 (L3) SM products have been validated extensively on a regular basis since the beginning of the mission (Al Bitar et al., 2012; Delwart et al., 2008) and they have been assessed as suitable for hydro-climate applications (Lievens et al., 2015; Wanders, Bierkens, de Jong, de Roo, & Karssenberg, 2014). However, most hydro-agricultural applications need SM measurements of sub-kilometer spatial resolution with a still

* Corresponding author.
E-mail address: beatriz.molero@cesbio.cnes.fr (B. Molero).

http://dx.doi.org/10.1016/j.rse.2016.02.045
0034-4257/© 2016 Elsevier Inc. All rights reserved.

representative temporal coverage (Walker & Houser, 2004). We should strive to provide a high resolution (HR) SM product that would enhance the knowledge of the hydrological processes at local scale.

Different satellite-based approaches have been proposed to retrieve SM. One of the most popular is the use of active sensors like the synthetic aperture radars (SAR) (ERS, ALOS, Sentinel 1) or scatterometers (ASCAT). These instruments provide observations with a variety of spatial and time resolutions but they are influenced to a great extent by the scattering produced by vegetation structure and surface roughness, among other factors. Unlike active sensors, passive instruments are much less sensitive to scattering but provide surface SM estimations at coarse resolutions (>40 km). C- and X-band radiometers like AMSR-E and WindSat have shown good results (Mladenova et al., 2011), but because of the frequency used, their sensing depth is shallow (~1 cm) and vegetation becomes rapidly opaque. In contrast, L-band radiometer acquisitions from SMOS provide SM estimations for a much wider range of vegetation conditions, with a sensing depth of around 5 cm and a revisit time of ~3 days. However, the spatial resolution provided is also coarse (35–55 km) as mentioned previously. The main strategies to work around this issue while maintaining the benefits of L-band consist of merging the L-band acquisitions with HR ancillary data, namely radar and optical observations.

Over the past decade, various methods have been proposed to combine active and passive sensors to produce HR SM (Das, Entekhabi, & Njoku, 2011; Narayan, Lakshmi, & Jackson, 2006; Zhan, Houser, Walker, & Crow, 2006). The NASA Soil Moisture Active Passive (SMAP) mission, launched in 2015, intended to combine L-band brightness temperatures (TB) and HR L-band radar backscatter data (Entekhabi, Njoku, O'Neill, Kellogg, Crow, Edelstein, et al., 2010). Despite the radar failure in July 2015, related previous studies showed that SM could have been delivered at 9 km and even 3 km resolution (Das et al., 2014).

Optical sensors (visible/near-infrared/thermal-infrared) can achieve finer spatial resolutions. However, the quality of their observations is critically compromised by the presence of clouds. Examples of optical sensors include the Landsat instruments and the Advanced Spaceborne Thermal Emission and Reflection radiometer (ASTER), with data at ~100 m resolution, and the MODerate resolution Imaging Spectroradiometer (MODIS), with data at ~1 km resolution. Such data include soil temperature and vegetation cover information, which are variables linked to soil water content (Fang et al., 2013). The relationship between land surface temperature (LST) and normalized difference vegetation index (NDVI) was first formalized in the 90s with the triangle (Carlson, 2007; Carlson, Gillies, & Perry, 1994) and the trapezoid (Moran, Clarke, Inoue, & Vidal, 1994) approaches.

Most of the methods for deriving HR SM from the synergy between optical and microwave observations are based on the triangle/trapezoid approaches. Chauhan, Miller, and Ardanuy (2003) stated that the relationship between LST, NDVI and SM can be formulated as a regression formula specific to the region and climatic conditions. Later, Piles et al. (2011) included SMOS TBs in the equation, which reduced the bias but slightly degraded the spatio-temporal correlation between the obtained HR SM and the *in situ* measurements. These empirical methods need local calibration of the regression coefficients at low resolution (LR) before applying them to the HR ancillary data. On the contrary, semi-physical methods replace the polynomial function by physically-based models that use evaporation as a proxy variable for SM variability. Merlin, Walker, Chehbouni, and Kerr (2008) linked the SM to the soil evaporative efficiency (SEE), defined as the ratio of actual to potential soil evaporation. Kim and Hogue (2012) established a linear relationship between the soil evaporative fraction of Jiang and Islam (2003) and SM. Both approaches improved the satellite SM spatial variability and showed better correspondence with ground measurements in the area of study (SMEX04).

The semi-physical methods have three important advantages with respect to the purely empirical methods: (i) the mean SM is preserved across the merging process (which justifies calling it 'disaggregation' or 'downscaling'), (ii) a physical link is established for HR between SM and the evaporation/evapotranspiration rate and (iii) no local calibration or fit is needed. These are key factors in developing a robust and global operational algorithm for HR SM.

Recent studies by Merlin et al. (2012); Merlin et al. (2013) have improved the evaporation rate calculation and the evaporation-SM link of Merlin et al. (2008). The DISaggregation based on Physical And Theoretical scale Change (DISPATCH) algorithm estimates SEE at high-resolution from soil temperature and vegetation data for modeling the spatial variations inside the microwave SM observation. In Merlin et al. (2012), DISPATCH included corrections for the microwave sensor weighting function and grid oversampling and provided an estimate of the uncertainty in the output disaggregated data. Later, Merlin et al. (2013) demonstrated that the linear approximation of the SEE—SM link model is suitable for kilometer scales and included soil temperature corrections for elevation effects. Both studies were conducted under semi-arid conditions, in a 500 × 100 km study area within the Murrumbidgee river catchment, in southeastern Australia, and in a 60 × 60 km study area east of Lleida in Catalunya, Spain. They showed that DISPATCH improves the spatio-temporal correlation with *in situ* measurements, but that the accuracy of disaggregated products is highly dependant on the SM-evaporation coupling. The downscaled resolution of 1 km (Merlin, Al Bitar, Walker, & Kerr, 2009; Merlin et al., 2013) and the combination of satellite data from different time stamps in DISPATCH (Malbéteau, Merlin, Molero, Rüdiger, & Bacon, 2016; Merlin et al., 2012) have been considered as a good trade-off between spatial representativeness and overall accuracy, given the current status of the algorithm.

Recently, a new Level-4 (L4) processor (C4DIS) based on DISPATCH has been implemented in the Centre Aval de Traitement des Données SMOS (CATDS), the French ground segment for SMOS Level-3 and Level-4 data. The aim is to disaggregate the SMOS CATDS Level-3 (L3) 1-day SM maps to produce maps of SM at 1 km resolution for any part of the globe on an operational basis. The ancillary temperature and vegetation data are retrieved from the MODIS mission.

This paper seeks (i) to provide an overview of the C4DIS architecture, processing algorithms, output products, strengths and weaknesses and (ii) to derive the first conclusions on the performance of the C4DIS product depending on the climatic and landscape conditions. To do so, we evaluate the C4DIS product against *in situ* data from the Murrumbidgee catchment and two additional contrasting networks. Former versions of DISPATCH have so far been evaluated mostly in semi-arid conditions (Malbéteau et al., 2016; Merlin et al., 2012, 2013). The Murrumbidgee network belongs to these previous studies, and it is included here to serve as a reference for the current version of DISPATCH and the C4DIS processor and for the other validation areas. The two other *in situ* networks considered in this study are located in the Little Washita watershed in Oklahoma, USA, which exhibits sub-humid conditions, and the Walnut Gulch watershed in Arizona, USA, which exhibits semi-arid to arid conditions. Their relief, soil properties and land use differ from the Murrumbidgee's. The L4 disaggregated SM product is evaluated using *in situ* 0–5 cm and *in situ* 0–8 cm measurements taken at the same time as SMOS overpasses (around 6 am, 6 pm) during the period 01/06/2010 to 31/05/2011 for the Australian network and 01/06/2010 to 31/12/2014 for the USA networks. These networks have been providing ground SM data in a continuous basis and have contributed to the validation of different satellite missions, SMOS among them (Cosh, Jackson, Bindlish, & Prueger, 2004; Jackson et al., 2010, 2012; Leroux et al., 2013; Peischl et al., 2012).

It is important to note that the DISPATCH algorithm will continue to evolve. Validation activities on the Level-4 processor C4DIS will provide valuable information for the improvement of the algorithm and processing chain. This current study is conducted on the products of the first version of the C4DIS processor.

2. Input data collection

2.1. In situ measurements

Three validation networks were selected for this work, the Murrumbidgee Soil Moisture Monitoring Network (MB) in Australia (Smith et al., 2012) and two different USDA (United Stated Department of Agriculture) networks: Little Washita (LW) in Oklahoma (Cosh, Jackson, Starks, & Heathman, 2006) and Walnut Gulch (WG) in Arizona (Cosh, Jackson, Moran, & Bindlish, 2008). They exhibit contrasted types of climate, soil properties, land use and spatial extension.

The MB network covers a large extension (82,000 km^2) in southern New South Wales. Its climate ranges from semi-arid in the west (average annual precipitation of 300 mm), to humid in the east (annual precipitation of 1900 mm at the Snowy Mountains). The MB has been studied in previous DISPATCH campaigns (Malbéteau et al., 2016; Merlin et al., 2012). It is included here for different reasons: it permits to confront results with previous versions of the algorithm, it contains within the Yanco area, which gathers the nominal landscape and climatic conditions for DISPATCH (flat, semi-arid with low vegetation), and it shows a variety of climate, soil and land use cases that can reveal the usefulness of disaggregation.

The MB consists in 38 validation stations: 18 of them provide SM integrated over the first 8 cm of soil (Campbell Scientific water content reflectometers) and the rest provide SM integrated over the first 5 cm of soil (Stevens Hydra Probe). The stations are situated in four areas: 7 stations in the limits of the catchment near to regional centers; 5 stations in Adelong Creek (145 km^2), a grazing area with steep slopes; 13 stations in Kyeamba creek (600 km^2), a catchment with gentle slopes and grazing and dairy land use; and finally, 13 stations in the Yanco region (3000 km^2).

Yanco soils are mainly silty-loam. The climate is semi-arid with an average annual rainfall of about 400 mm, with most of the precipitation occurring in winter and spring. The land use is divided into irrigation and dry land cropping and pastures. This area has been extensively monitored since 2001 (Smith et al., 2012) and has been used in a variety of satellite validation campaigns (Mladenova et al., 2011; Panciera et al., 2014; Peischl et al., 2012).

The USDA networks have been operating since 2002 and they have been used in the validation of Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) products (Jackson et al., 2010), Aquarius (Bindlish, 2015), ASCAT (Leroux et al., 2013) and SMOS products (Jackson et al., 2012). The probes are installed at a depth of 5 cm, with an effective measurement depth between 3 and 7 cm (Stevens Hydra Probe).

LW is located in southwest Oklahoma and covers an area of about 610 km^2. The climate is sub-humid with an average annual rainfall of 750 mm. Summers are hot and relatively dry while winters are short and temperate. Autumn and spring are when most of the precipitation occurs (Allen & Naney, 1991). The land use is mainly rangeland and crops that include winter wheat and some corn and grasses. Soils include a wide range of textures, with large regions of sands, loams and clays. The topography is moderately rolling with few hills.

WG occupies an area of 148 km^2 in southeastern Arizona. The climate is semi-arid, with an average annual rainfall of 324 mm, lower than in the Yanco region. Most of the rains occur in the form of small scale high-intensity thunderstorms during the summer months as part of the North American Monsoon System (Cosh et al., 2008). Soils are mainly sands and gravel with good drainage. Desert shrubs and short grasses dominate the landscape. The topography is considered as rolling with significant rock cover. Although the climate class of WG is defined semi-arid as the Yanco area, the contrasting landscape and precipitation conditions make WG an interesting validation area (Table 1).

It is important to outline that the area extent covered by the networks is different so it may have an impact on the validation process: the MB comprises multiple SMOS pixels through sparse stations and more dense localized sites, the Yanco region covers approximately one SMOS pixel, and the LW and WG cover around 1/4 and 1/16 of the surface of one SMOS pixel. This does not affect the C4DIS processor, which handles input larger surfaces, but it may affect the validation process since the smaller networks may not be representative of the ~40 km surface.

2.2. SMOS soil moisture data

The SMOS satellite was launched in November 2009. SMOS has global coverage with a revisit period of 3 days at the equator, with all together in the same line, if possible overpass at 6:00 am and descending (D) overpass 6:00 pm local solar time. The SMOS instrument is a passive 2D interferometer operating at L band (1.4 GHz) (Kerr et al., 2001, 2010). The spatial resolution ranges from 35 to 55 km, depending on the incident angle. The goal is to retrieve SM (first 5 cm) with a target accuracy of 0.04 m^3/m^3 (Kerr et al., 2012).

The C4DIS processor disaggregates the SM provided by the SMOS Level-3 1-day global SM product (MIR CLF31A/D). In this paper, the version 2.72 (in 220 reprocessing mode RE02) product is used. Level-3 (L3) products are presented in NetCDF format on the EASE (Equal Area Scalable Earth) grid, with a grid spacing of ~25 × 25 km.

The L3 SM products are directly computed from the SMOS Level-1 products at the CATDS. The core of the algorithm for retrieving SM from brightness temperatures is derived from the L2 retrieval algorithm (Kerr et al., 2012; Wigneron et al., 2007). In both processing chains, SM is derived from the combination of multiangular observations. While the L2 chain considers only the multiangular observations of the same day and orbit (ascending/descending), the L3 chain uses several overpasses (3 at most) over a 7-day window. This results in more coverage and robustness for the L3 products (Al-Yaari et al., 2014). Details on the L3 processing algorithm can be found in the Algorithm Theoretical Baseline Document (Kerr et al., 2013) and in the L3 data product description (Kerr et al., 2014).

2.3. MODIS temperature and vegetation data

The C4DIS processor uses three ancillary products at 1 km resolution. Two of them are derived from MODIS acquisitions: LST and NDVI. These are necessary elements for the SEE calculation inside DISPATCH.

Table 1
Main characteristics of validation areas.

	Murrumbidgee	Yanco	LW	WG
Extension	82,000 km^2	3000 km^2	610 km^2	148 km^2
Climate	Semi-arid (west) to humid (east)	Semi-arid	Sub-humid	Semi-arid to arid
Annual precipitation	300–1900 mm	400 mm	750 mm	324 mm
Main precipitation periods	Relatively constant at the basin scale	Winter, spring	Autumn, spring	Summer (intense, localized)
Soils	Clayey (west) to sandy (east)	Silty-loam	Sands, loams and clays	Sands and gravel
Topography	Diverse, mountains in the east	Flat	Moderate rolling	Rolling

The LST datasets are extracted from in the same line, if possible MODIS/Terra LST and emissivity daily L3 global 1-km grid products (MOD11A1) and version-5 MODIS/Aqua LST and emissivity daily L3 global 1-km grid products (MYD11A1). The NDVI dataset is extracted from the version-5 MODIS/Terra vegetation indices 16-day Level-3 global 1-km grid product (MOD13A2).

The MODIS products are retrieved from the NASA Land Processes Distributed Active Archive Center (LP DAAC). They are presented in sinusoidal projection at 1 km resolution (Solano, Didan, Jacobson, & Huete, 2010; Wan, 1999, 2006). The disaggregation approach requires the NDVI dataset acquired within the last 15 days and the LST datasets of the day before, the same day and the day after. The MODIS products are available between 1 and 9 days after the acquisition day.

2.4. Digital elevation model

The C4DIS processor requires elevation information, which is extracted from the GTOPO30 Digital Elevation Model (DEM) product available in the WGS84 sphere at 30-arc sec resolution. The GTOPO30 product is distributed by the U.S. Geological Survey's EROS Data Center (USGS, https://lta.cr.usgs.gov/GTOPO30).

3. The CATDS level-4 disaggregation (C4DIS) processor

The CATDS Level-4 (L4) Disaggregation (C4DIS) processor is the first operational version of the DISPATCH algorithm. The C4DIS processor selects the best algorithm and parameter configuration according to past DISPATCH studies and the latest research (Merlin, Al Bitar, Walker, & Kerr, 2010; Merlin, Chehbouni, Boulet, & Kerr, 2006; Merlin et al., 2009; Merlin, Rüdiger, Richaume, Al Bitar, Mialon, Walker and Kerr, 2010; Merlin et al., 2012; Merlin et al., 2013). It also makes possible to obtain disaggregated SM on a global and daily basis (under the assumption of no cloud-covered scenes and availability of input data). The C4DIS products have been marked as 'scientific' products because the algorithm is still evolving: their access will be granted on demand for specific areas of the world. In this and the following sections, we describe both the DISPATCH prototype and the C4DIS processor.

3.1. DISPATCH algorithm

DISPATCH relies on a SEE term to model the spatial variability over the low-resolution (LR) SMOS pixel. The first step is to account for the SEE term at HR, described as a linear function of soil temperature:

$$SEE_{HR} = (T_{s,max} - T_{s,HR})/(T_{s,max} - T_{s,min}) \quad (1)$$

Soil ($T_{s,HR}$) and vegetation ($T_{v,HR}$) temperatures are derived from LST and NDVI datasets as in Merlin et al. (2012), where the surface temperature is partitioned into its soil and vegetation components according to the trapezoid method of Moran et al. (1994). Soil temperature is calculated as follows:

$$T_{s,HR} = (T_{MODIS} - f_{v,HR}T_{v,HR})/(1 - f_{v,HR}) \quad (2)$$

with T_{MODIS} being the MODIS LST and f_v the MODIS-derived fractional vegetation cover. Here, the fractional vegetation cover is calculated as:

$$f_{v,HR} = (NDVI_{MODIS} - NDVI_s)/(NDVI_v - NDVI_s) \quad (3)$$

with $NDVI_{MODIS}$ being the MODIS NDVI, $NDVI_s$ the NDVI for bare soil (set to 0.15), and $NDVI_v$ the NDVI for full-cover vegetation (set to 0.90).

The vegetation temperature $T_{v,HR}$ is calculated according to the "hourglass" approach (Moran et al., 1994), as a function of the position of the HR pixel in the LST-f_v space, and the soil ($T_{s,min}$, $T_{s,max}$) and vegetation ($T_{v,min}$, $T_{v,max}$) temperature end-members (Merlin et al., 2012).

Given the minimum and maximum LST values of the scene $T_{MODIS,min}$ and $T_{MODIS,max}$, and the f_v values associated to the same pixels, $f_{v,Tmin}$ and $f_{v,Tmax}$, the following approximations hold (Merlin et al., 2013):

I). $T_{v,min} = T_{MODIS,min}$
II). When the vegetation portion is low at $T_{MODIS,min}$ ($f_{v,Tmin} < 0.5$), then $T_{s,min} = T_{MODIS,min}$
III). When the vegetation portion is considerable at $T_{MODIS,min}$ ($f_{v,Tmin} >= 0.5$), then $T_{s,min}$ is set to the minimum value of the $T_{s,HR}$ derived from Eq. (2), with $T_{v,HR} = T_{v,min}$ and $f_{v,HR} < 0.5$
IV). When the vegetation portion is low at $T_{MODIS,max}$ ($f_{v,Tmax} < 0.5$), then $T_{s,max} = T_{MODIS,max}$ and $T_{v,max}$ is set to the maximum value of the $T_{v,HR}$ derived from Eq. (2), with $T_{s,HR} = T_{s,max}$ and $f_{v,HR} \geq 0.5$
V). When the vegetation portion is considerable at $T_{MODIS,max}$ ($f_{v,Tmax} >= 0.5$), then $T_{v,max} = T_{MODIS,max}$ and $T_{s,max}$ is set to the maximum value of the $T_{s,HR}$ derived from Eq. (2), with $T_{v,HR} = T_{v,max}$ and $f_{v,HR} < 0.5$

Note that LST has been preliminary corrected for elevation effects (decrease of air temperature with altitude) by using the DEM information at HR (Merlin et al., 2013):

$$T_{MODIS} = T_{MODIS-ori} + \gamma(H_{HR} - H_{LR}) \quad (4)$$

with T_{MODIS} being the topography-corrected LST used in the previous equations, $T_{MODIS-ori}$ the original MODIS LST, γ (°C m^{-1}) the mean lapse rate (set to 0.006 °C m^{-1}), H_{HR} the altitude of the MODIS pixel and H_{LR} the mean altitude within the LR pixel.

In a second step, the semi-empirical linear model of Budyko (1956) and Manabe (1969) is used to link the surface SM (0–5 cm) and the SEE terms. According to Merlin et al. (2013), the linear model is a good approximation for kilometer scales so the SEE for each HR pixel can be written as:

$$SEE_{HR} = SM_{HR}/SM_p \quad (5)$$

where SM_p is a parameter estimated at LR at each execution from daily SM and SEE observations as follows:

$$SM_p = SM_{LR}/SEE_{LR} \quad (6)$$

with SM_{LR} the radiometer-sensed SM and SEE_{LR} the average of the SEE_{HR} values inside the LR pixel.

The disaggregation is finished by applying a first order Taylor series to the SM-SEE model at each HR pixel (downscaling relationship). The corresponding disaggregated SM is:

$$SM_{HR} = SM_{LR} + SM'(SEE_{LR}) \times (SEE_{HR} - SEE_{LR}) \quad (7)$$

with $SM'(SEE_{LR})$ the partial derivative of SM relative to SEE at LR (SM_p).

3.2. DISPATCH operational implementation

Following the methodology introduced in Merlin et al. (2012), C4DIS executes DISPATCH on a set of possible combinations of input datasets, producing multiple HR outputs that are averaged together into a single final disaggregated SM field (SM_HR). The rationale behind this is to account for the uncertainty of the approach and to reduce independent random errors (Malbéteau et al., 2016; Merlin et al., 2012). The input ensemble is formed by 4 downsampled instances of the original L3 SM dataset and up to 6 LST datasets corresponding to 3 consecutive days of MODIS acquisitions (Aqua and Terra overpasses). This means that each SM_HR output comes from the composition of up to 24 DISPATCH outputs (up to 24 input SM-LST possible pairs).

SMOS original datasets are downsampled in order to work at the radiometer resolution. SMOS L3 products are provided on a 25 km grid, which can be up to half of the original SMOS resolution (35–50 km).

The four SM datasets are derived from the original SM map by sampling the data at 50 km and are assumed to be independent. This is not totally true, since grid cells depend on the surrounding cells from a radiometric perspective, but helps to potentially reduce (and provide an estimate of) random errors in the SM_HR data. Regarding the selection of 6 MODIS LST datasets from 3 consecutive days, it is assumed that SM fields are spatially stable for periods of at least 1 day around the SMOS overpass time. This 3-day derived product with daily estimated SM_p is referred as the 'sm1k3d' product in Malbéteau et al. (2016) and is the one built by the C4DIS processor. The 3-day product has much better temporal coverage than its 1-day counterpart ('sm1k1d'), but the uncertainty associated to the methodology is expected to be higher since the temporal stability assumption can be often violated by precipitation and irrigation events.

There is no dedicated dataset in the C4DIS product that specifies explicitly whether the 3-days stability condition is respected or not. In the future, this will be achievable with the use of ancillary precipitation information, for example. Meanwhile, in addition to the SM_HR dataset, two more datasets are produced as indicators of the aggregation of the DISPATCH ensemble: the STD dataset, which is the standard deviation of the up to 24 disaggregated SM fields, and the COUNT dataset, which is the size of the ensemble. The aggregation is conducted if at least 3 SM fields are generated, so the COUNT values range from 3 to 24. In this paper, we study the STD and the COUNT datasets as potential sources of information for a future quality control flag (Section 5.5).

Finally, the current version of DISPATCH filters out any LST pixel values that have associated QC flags different from 0 and 17, which correspond to maximum LST quality (error < 1 K) and maximum emissivity error of 0.01 and 0.02 respectively (Solano et al., 2010; Wan, 2006). Areas with more than 1/3 of their surface covered by clouds are also discarded. Differences between the operational and the prototype versions of DISPATCH are summarized in Table 2.

3.3. Pre-processor

The C4DIS pre-processor prepares the input ensemble that is required by DISPATCH. The pre-processor uses the MODIS sinusoidal tiling system as the execution reference, meaning that the processor is executed for the SMOS and ancillary data contained within each MODIS tile bounds. More information about the grid can be found in http://modis-land.gsfc.nasa.gov/MODLAND_grid.html. The SMOS and ancillary data inside the tile bounds are selected and re-projected to an equal-spaced lat-lon WGS84 grid. Considering that ancillary products are presented in different datums and grids, the choice of the WGS84 projection minimizes the total number of resampling operations.

The pre-processor is divided into modules for file format transformation, dataset extraction, re-projection and re-gridding. As explained in the previous section, DISPATCH requires 4 subsampled instances of SMOS data and up to 6 LST datasets. As a consequence, the re-projection and re-gridding are sensible operations that deserve being explained in detail.

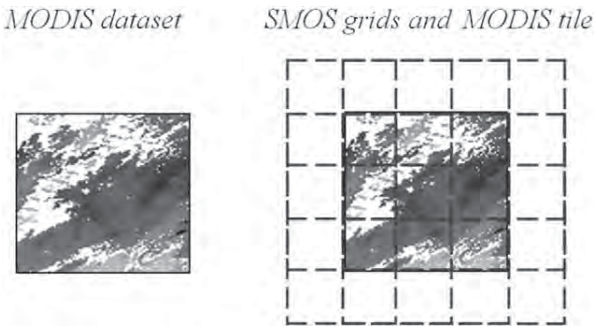


Fig. 1. Simplistic representation of the relation between the SMOS subsampled grids (at 0.4°) and the re-projected ancillary data at 0.01°. The extent of the re-projected ancillary image (LST, NDVI, etc.) matches the intersection of the four SMOS grids. The disaggregation is only applied in this overlapping zone.

The pre-processor outputs are re-projected to the same WGS84 projection, but resampled to different resolutions: SMOS subsampled rasters are provided on 0.4° grids while ancillary raster data are provided on a 0.01° grid. The SMOS 0.4° grids are derived from an original global grid at 0.2° by sliding a 0.4° window over it, so that the pixel centers are coincident. Based on this, the SM values become representative of the double of the original grid resolution 0.2°, which approximately matches the average SMOS resolution. The disaggregation is only performed in the intersection area between the 4 SMOS grids and the ancillary data grid (Fig. 1).

3.4. Post-processor

The C4DIS post-processor transforms the DISPATCH outputs into the CATDS format. It includes two significant transformations that impact the disaggregated data. First, in the case that DISPATCH generates negative SM values (which is mathematically possible), the post-processor clips them to 0 to respect physical meaning. Second, since the outputs of DISPATCH are presented in local time and day, the post-processor assigns to them the corresponding UTC time and day to keep consistency with other SMOS products.

3.5. Assumptions and applicability domains of the algorithm

The application requirements of the C4DIS processor are directly inherited from DISPATCH. The following considerations must be taken into account:

Table 2
Main differences between the DISPATCH operational implementation in the C4DIS processor and the previous prototype versions.

	C4DIS processor	Merlin et al. (2013)	Merlin et al. (2012)
SEE model	Linear (Budyko, 1956; Manabe, 1969)	Linear (Budyko, 1956; Manabe, 1969)	Non-linear (Noilhan & Planton, 1989)
Calculation of T_v	"Hourglass" approach (Moran et al., 1994)	"Hourglass" approach (Moran et al., 1994)	"Hourglass" approach (Moran et al., 1994)
Calculation of temperature end-members ($T_{s,min}$, $T_{s,max}$, $T_{v,min}$, $T_{v,max}$)	Estimated by a simpler approach based on the combination of LST and f_v	Estimated by a simpler approach based on the combination of LST and f_v	Estimated by plotting MODIS LST against MODIS albedo and NDVI within the LR pixel (Merlin, Duchemin, et al., 2010)
Input SM data	SMOS L3 SM	SMOS L2 SM	SMOS L2 SM
Input LST data	"sm1k3d" mode (3 × 2 input LST datasets)	"sm1k1d" mode (1 × 2 input LST datasets)	"sm1k3d" mode (3 × 2 input LST datasets)
Input DEM data	GTOPO30	GTOPO30	Not implemented
LST filtering	Yes, QC flags 0 and 17	Yes, QC flags 0 and 17	No
Cloud-free threshold	0.67	0.90	0.90
Sea-free threshold	0.90	0.90	Not implemented

- *Cloud free conditions*: soil temperature can only been retrieved from optical sensors if clouds are not present. C4DIS products show data gaps associated with clouds.
- *Low vegetation cover*: The LST-NDVI trapezoid describes a zone of values where no useful disaggregated data can be produced since LST is mainly controlled by vegetation transpiration, with no sensitivity to surface SM (Merlin et al., 2013). Sites with partial fractional vegetation cover at the 1 km resolution are desired.
- *Moisture-driven evapotranspiration*: the disaggregation relies on the dependence established between LST, evapotranspiration and SM. Some climates exhibit low dependency between those variables. Typically, climates characterized as energy-limited, like humid climates, exhibit a weaker moisture-evaporation coupling.
- *Medium to high spatial variability*: the MODIS-derived SEE is computed with a polygon method that relies on LST and reflectance end-members (Moran et al., 1994). In the current version, DISPATCH is contextual and thus heterogeneous scenes with meaningful dry-wet contrast are needed in order to ensure good end-members accuracy (Merlin, Al Bitar, Walker and Kerr, 2010). Note that LST end-members could be estimated using available meteorological data (Moran et al., 1994) independently from the surface (wet/dry) conditions observed at the 1 km resolution within the LR pixel (Stefan, Merlin, Er-Raki, Escorihuela, & Khabba, 2015).
- *Accuracy of the SM_p parameter*: the SM_p parameter is calculated at LR scale by using a linear relationship that has been studied as suitable for kilometer scales (Merlin et al., 2013). It is based on the assumption that the sub-pixel variability of SM_p at HR is negligible. Soil characteristics (texture, porosity, etc.) may impact the relationship between SEE and SM and thus SM_p. Hence, the current versions of C4DIS and DISPATCH should perform better in areas with homogeneous soil characteristics where the intra-pixel spatial SM variability is mainly due to forcing agents, namely precipitation and irrigation.
- *Mismatch of overpass times*: the C4DIS processor uses MODIS LST datasets at 6 different timestamps. This is based on the assumption that the SM pattern is maintained over a period of 3 days, with no rain events occurring in between.
- *Mismatch of sensing depths*: SMOS L-band SM estimations are representative of the soil first 5 cm content, while MODIS temperature acquisitions are representative of the soil skin layer. DISPATCH assumes that the soil skin temperature is correlated with the soil evaporation process occurring in the 0–5 cm of soil (Merlin, Al Bitar, Walker and Kerr, 2010).

3.6. Global product description

- *Coverage, grid and resolution.* C4DIS products are presented in a regular lat-lon grid at 0.01° resolution. The projection is divided in a tiled grid that follows the MODIS sinusoidal tiling system, meaning that the C4DIS tiles are centered at MODIS tiles and follow the same name convention in (h,v) coordinates. Due to reprojection, the tiles present different size. C4DIS products can be generated for all emerged lands (tiles with more than 50% of land), but since they are tagged as 'scientific' products, the tiles of interest have to be delivered on demand. For this study, the following tiles have been produced: (29,12) and (30,12) for the validation over the MB, (09,05) for LW and (08,05) for WG. Fig. 2 and Fig. 3 show annual averages of C4DIS products for the selected tiles. The extension and border of the tiles are easily distinguishable.
- *Availability and timeliness.* The delivering of C4DIS products is determined by the availability and timeliness of the input datasets. The limiting dataset is the MODIS MOD13A2 product (NDVI), which is valid for a period of 15 days starting at its date of acquisition (DoA) but can be delivered some days later. In consequence, C4DIS products for dates DoA to DoA + 15 are produced at date DoA + 25. In other words, each 16 days the C4DIS products for acquisition dates between 25 to 10 days before are delivered.
- *Datasets and quality control.* We cannot provide a full-proof quality flag given the current status of the processor and the algorithm. Nevertheless, the output COUNT and STD datasets can help to assess the quality of the SM_HR dataset. Combining these datasets with additional ancillary data like precipitation or MODIS/SMOS quality flags, may help to build a quality control dataset in the future.

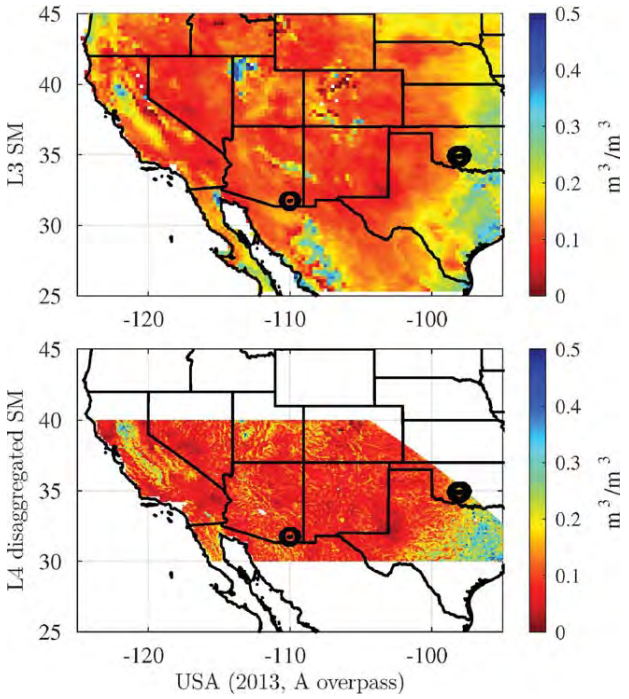


Fig. 2. Year averages of SMOS L3 and L4 disaggregated products (ascending orbit) for part of the USA and for the period 06/2013 to 05/2014. The L4 figure includes only the tiles (08,05) and (09,05), joined together. The black circles correspond, from left to right, to the location of Walnut Gulch and Little Washita validation networks.

As introduced in Section 3.2, the COUNT field determines the number of SM–LST combinations used by DISPATCH to produce one output. Low COUNT values indicate missing input data as result of diverse reasons: SMOS RFI contamination, MODIS cloudy scenes, failures in the SMOS/MODIS acquisitions delivering, etc. SM_HR fields generated when low COUNT values are present do not profit from the reduction in independent random errors as result of averaging. The STD field contains the per-pixel standard deviation of the up to 24 disaggregated datasets with respect to the averaged output SM_HR. Low values of STD are desirable since they reveal temporal persistency of both temperature and moisture variables. High values may indicate external forcing agents (precipitation and irrigation) within the 3-days window.

4. Analysis methodology

Our analysis involves two main approaches: qualitative assessment of disaggregated SM maps and statistical evaluation. The statistical evaluation consists on comparing the L3 SMOS product (LR) and the L4 product (HR) against the *in situ* SM by using standard statistical metrics (e.g. correlation, bias, etc.). This can be accomplished in the spatial or in the time domain. We base the statistical evaluation on the assumption that the 1 km pixel is more representative of the *in situ* measurement than the whole LR pixel.

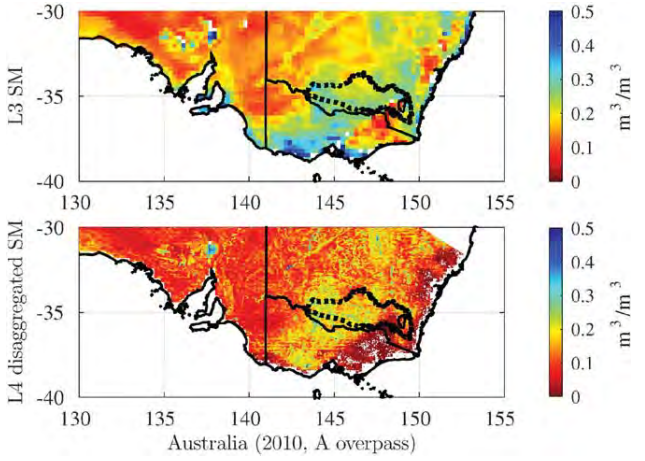


Fig. 3. Year averages of SMOS L3 and L4 disaggregated products (ascending orbit) for part of Australia and for the period 06/2010 to 05/2011. The L4 figure includes part of the tiles (29,12) and (30,12), joined together. The dotted line depicts the boundary of the Murrumbidgee catchment. The presence of clouds affects the L4 and not the L3 products, preventing the first one to show lower average values of SM.

In order to assess the relative spatial performance of both L3 and L4 products at HR, we directly compare the station measurements to the satellite retrievals, without aggregating them at LR. In the subsequent sections, MB refers to the whole Murrumbidgee network, including Yanco area. Yanco only refers to the 12 stations contained in this region.

4.1. Data preparation

We filter L3 and L4 SM time series for radio frequency interference (RFI) by removing pixels having more than 10% RFI probability. The RFI information is extracted from the same CLF31A/D product and accounts for the percentage of brightness temperatures acquisitions affected by RFI presence (Kerr et al., 2013; Oliva et al., 2012). In addition, regarding the *in situ* data, we only keep the SM values at the SMOS overpass times. Finally, we filter the three SM time series (*in situ*, L3 and L4) for common dates with valid SM values (>0.0 m^3/m^3).

4.2. Analysis of the temporal and spatial variability of the in situ SM

As expected for any data disaggregation approach, the application of DISPATCH is relevant when the SM spatial variability at the downscaled resolution is larger than the output uncertainty. Since the current version of DISPATCH relies on the spatial contrast of LST and SM of the scene, a preliminary study on the spatial SM variability of the validation areas is desired. In homogeneous SM landscapes, the output uncertainty is likely to be greater than the spatial gain provided at HR by disaggregation.

Similarly, it is desirable that the evaluation include *in situ* time series spanning the full range of SM conditions and seasonal changes. In other words, the temporal standard deviation (σ) should be large enough so that all the states of the SM variable are represented and no selection bias is present. Additionally, stations exhibiting very different temporal σ may suggest landscape spatial heterogeneity: soil characteristics like texture, vegetation and topography affect the dry-down process, generating different extreme values in time.

Based upon the considerations discussed above, the evaluation of the performance of the C4DIS products should include a preliminary assessment of the spatial and temporal SM variability of the validation networks. The performance of DISPATCH outputs over MB and Yanco has been identified as rather satisfactory in recent studies (Malbéteau et al., 2016; Merlin et al., 2012), which makes them good references for spatial and temporal σ.

4.3. Classical metrics

Given the spatial mismatch between *in situ* and satellite estimations and the spatial scarcity of ground stations, most classical satellite validation campaigns only evaluate the temporal dimension, by means of metrics like correlation (*R*), root mean square error (RMSE) and bias (B) (Albergel et al., 2012; Albergel, Brocca, Wagner, de Rosnay, & Calvet, 2013; Entekhabi, Reichle, Koster, & Crow, 2010; Al Bitar et al., 2012). In this study, we use similar temporal analysis but we also include an evaluation in the spatial domain since disaggregation techniques aim at producing better spatial representation. The spatial statistical analysis consists of computing the metrics between the satellite and *in situ* values for each day, then, deriving the average of each metric for the whole period. We deliberately establish a minimum of 5 points per day to compute the metrics.

Herein, instead of the RMSE, we use as error metric the standard deviation of the error (Eq. (8)) (Mood, Graybill, & Boes, 1974; Salkind, 2010), which is a non-biased estimation of the error and so it is not compromised by the bias in the mean and amplitude of the time series that affects the RMSE. The relationship between both metrics is written in Eqs. (9) and (10). Since we already use multiple terms to refer to different standard deviation measures and datasets in this paper (σ, STD), we will refer to this metric as unbiased-RMSE or ubRMSE (Entekhabi, Reichle, et al., 2010). Given that the 1 km pixels are in general heterogeneous and that the ground data also present measurement uncertainties, the term 'error' has been replaced by 'difference' in these metrics, i.e. RMSD and ubRMSD.

$$\mathrm{ubRMSD} = \sqrt{\mathrm{E}\left\{\left[\left(\mathrm{SM_{satellite}} - \mathrm{E}\{\mathrm{SM_{satellite}}\}\right) - \left(\mathrm{SM_{insitu}} - \mathrm{E}\{\mathrm{SM_{insitu}}\}\right)\right]^2\right\}} \quad (8)$$

$$\mathrm{RMSD} = \sqrt{\mathrm{E}\left\{\left(\mathrm{SM_{satellite}} - \mathrm{SM_{insitu}}\right)^2\right\}} \quad (9)$$

$$\mathrm{ubRMSD} = \sqrt{\left(\mathrm{RMSD}^2 - \mathrm{B}^2\right)} \quad (10)$$

where $\mathrm{E}\{\cdot\}$ is the expectation operator, $\mathrm{SM_{satellite}}$ and $\mathrm{SM_{insitu}}$ the satellite and the *in situ* SM time series.

We include one additional metric to assess the efficiency gained in spatial representativeness: the slope (*S*) of the regression line between *in situ* and satellite estimates:

$$\mathrm{S} = \mathrm{R} \cdot \sigma_{\mathrm{satellite}} / \sigma_{\mathrm{insitu}} \quad (11)$$

with $\sigma_{\mathrm{satellite}}$ and σ_{insitu} being the standard deviations of satellite and *in situ* SM, respectively. The S metric can help to understand how much better the SM redistribution is represented after the disaggregation process. Whereas aggregation systematically decreases the $\sigma_{\mathrm{satellite}}$, disaggregation specifically aims to improve the spatial representation of satellite SM by increasing the $\sigma_{\mathrm{satellite}}$ at the level of σ_{insitu}, while keeping a significant R. Mathematically speaking, R is the slope of the standardized regression line, and S is scaled by the σ values of both data ensembles (Rodgers & Nicewander, 1988). Since the σ_{insitu} is fixed, S is more sensitive than R to changes in $\sigma_{\mathrm{satellite}}$. In summary, an increase in random uncertainties (larger ubRMSD, smaller R) in disaggregated SM might be acceptable if S is closer to 1. Note that the random uncertainties in satellite SM can be significantly reduced via the techniques of data assimilation in land surface models, but the systematic errors associated with the mismatch between data resolution and model application scale are more difficult to take into account at HR (Merlin et al., 2006).

Finally, the metrics here (S, R, ubRMSD, B) assume that a linear relationship exists between the two datasets compared. This means that they cannot replace the visual assessment of the data. In the general case, both SMOS L3 and disaggregated SM may exhibit non-linear behavior with respect to *in situ* SM.

Table 3
List of performance metrics used in this study, from (Merlin et al., 2015).

Gain(S) …….....	G_{EFFI}
Gain(R) …….....	G_{ACCU}
Gain(B) …….....	G_{ROBU}
Gain(ubRMSD)....	G_{ubRMSD}

4.4. Relative performance metrics

Comparing the improvement/degradation in statistics for different cases of study (networks, filtering, time period, etc.) may be difficult: we propose as solution to calculate their relative gains as introduced in Merlin et al. (2015). Briefly, the gain is a measure of the improvement in the statistics obtained for the L4-*in situ* pair with respect to the L3-*in situ* pair. The gain can range from −1 to 1, where positive values indicate disaggregated data having better correspondence with *in situ* than LR data. In this study, we keep the nomenclature of Merlin et al. (2015) and we add a new gain term for the ubRMSD (see Table 3). The gains are calculated as in Eq. (12) for in S and R metrics, and as in Eq. (13) for B and ubRMSD.

$$G_X = -(|1-X_{L4}|-|1-X_{L3}|)/(|1-X_{L4}|+|1-X_{L3}|) \quad (12)$$

$$G_X = -(|X_{L4}|-|X_{L3}|)/(|X_{L4}|+|X_{L3}|) \quad (13)$$

where X designates the metric (S, R, B, ubRMSD), X_{L4} the value of the metric when disaggregated SM is compared against *in situ*, and X_{L3} the value of the metric when L3 SM is compared against *in situ*.

5. Results and discussion

This study seeks to provide a first assessment on the applicability of the DISPATCH-based processor under different climatic and landscape conditions. It also attempts to provide statistical guidelines on the *a priori* suitability of a geographical area for the production of meaningful C4DIS fields. The analysis spans the 01/06/2010 to 31/05/2011 period for the MB network and Yanco area and the 01/06/2010 to 31/12/2014 period for the LW and WG networks. The SMOS data collected during the commissioning phase (until 31/05/2010) is discarded.

5.1. Preliminary analysis

In order to predict the performance of the processor, we conduct a statistical analysis on the *in situ* SM data. We derive conclusions about their temporal and spatial variabilities by looking at their distribution of SM values and their distribution of 'spatial σ' and 'temporal σ'. The 'spatial σ' (upper row in Fig. 4) is the standard deviation of the SM distribution on each day. The 'temporal σ' (middle row) is the standard deviation of the SM series of each station.

As stated in Section 4.2, we consider the *in situ* SM distribution characteristics of MB and Yanco networks as reference in the present study. The spatial σ plot shows narrower distributions for LW and WG, and the mean value is much lower for the latter (0.03 m^3/m^3). This means that the spatial variability at LW and WG seen at the satellite overpass times is lower than in the reference cases, so we expect poorer performances in the spatial domain.

In the temporal domain (middle row of Fig. 4), the mean variability of LW and WG networks is lower than that of the Australian cases.

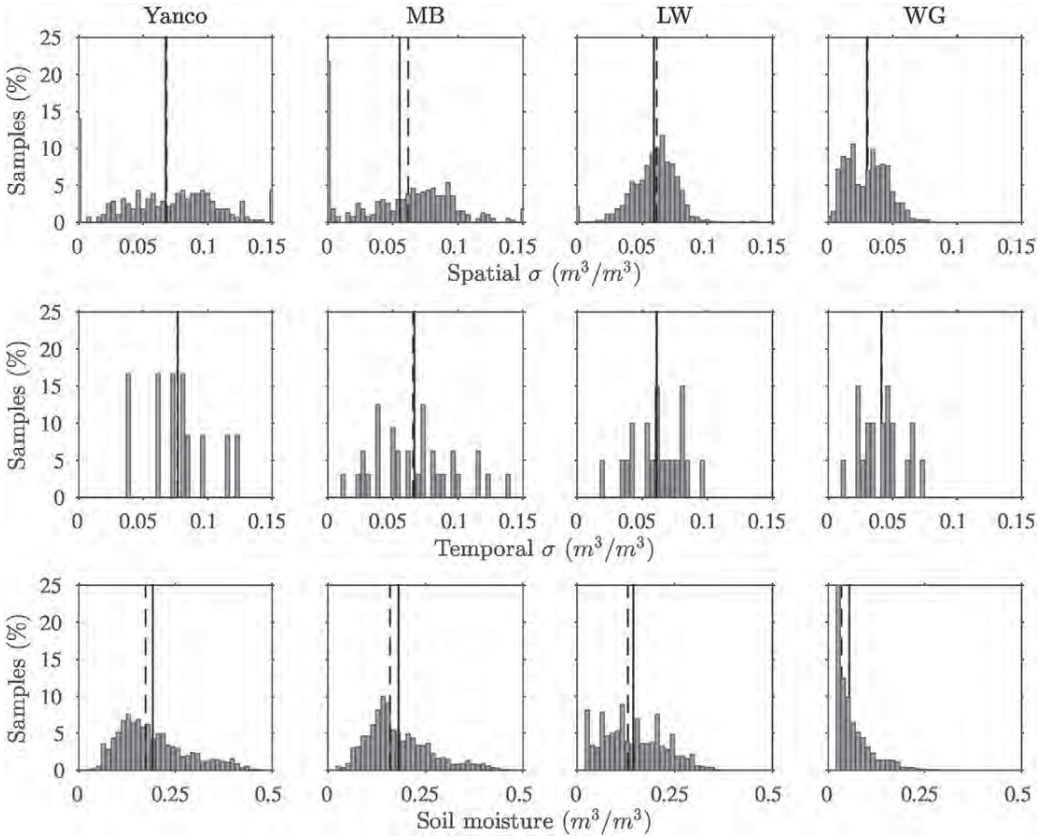


Fig. 4. Distribution of spatial and temporal standard deviations and SM values for the *in situ* samples of Yanco, MB, LW and WG (1st to 4th columns respectively) at the SMOS overpass times. Number of bins of the histograms is 40. The median of the distributions is depicted in dashed line and the mean in solid line. The WG soil moisture maximum percentage is not shown (right-down graph) for readability and it reaches 47% of the samples.

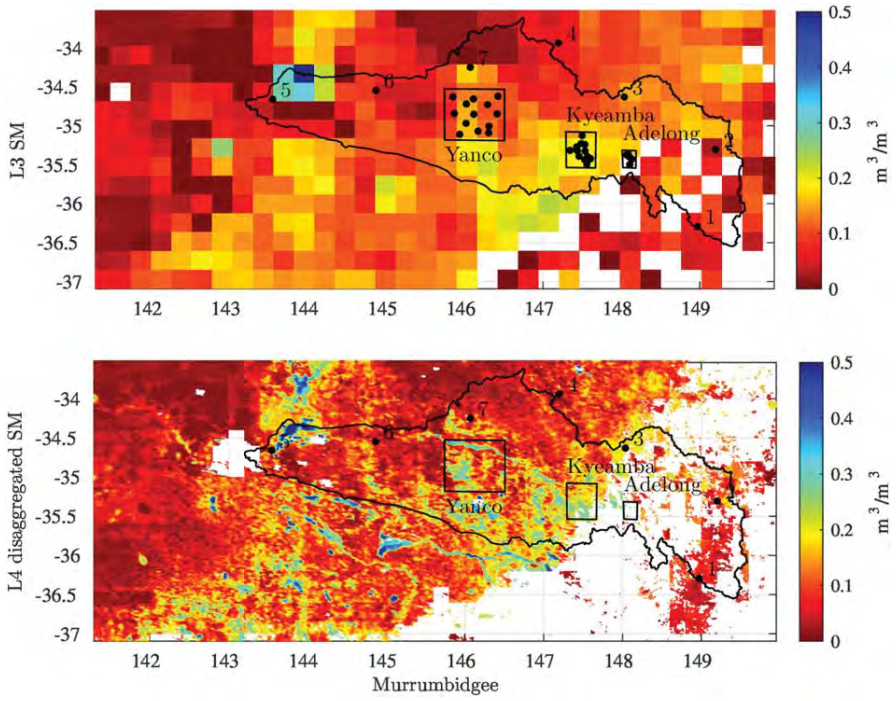


Fig. 5. Maps of L3 SM (CLF31D) and L4 disaggregated SM for MB watershed on 22/11/2010 for the SMOS descending overpass.

The SM distribution of WG (lower-right figure) shows a very strong peak near zero that accounts for almost the half of the samples. Under these conditions, we expect WG to be the network with worst temporal performance of C4DIS products, while LW should behave similarly to MB and Yanco. It is important to mention that LW and WG only represent a portion of a SMOS pixel and the *in situ* samples only concern some HR pixels in space, so the distributions depicted here serve only as approximation.

5.2. Qualitative examples

The qualitative inspection of disaggregated SM maps for MB, Yanco, LW and WG, shows that the L4 product is able to reveal spatial entities like small and sparse water bodies.

Figs. 5 and 6 contain sample outputs of the C4DIS processor on cloud-free days for the four areas. In the MB picture (Fig. 5), the Murrumbidgee river is revealed thanks to disaggregation, while the south-eastern region is empty due to clouds and the SMOS non-retrieved pixels over the mountains. In Fig. 6, disaggregation does not help reveal the Little Washita river course but it does with the surrounding lakes. The processor fails to display any spatial pattern inside the WG watershed. These maps are in agreement with the evaluation in the previous section.

Yanco maps are a good example of the usefulness and relevance of the C4DIS products when the algorithm assumptions are met. Fig. 7 shows the Yanco area with the limits of the Coleambally Irrigation Area (CIA) units superimposed. At a first glance, the L4 SM map reveals the farms that are actually irrigated, while original SM map do not.

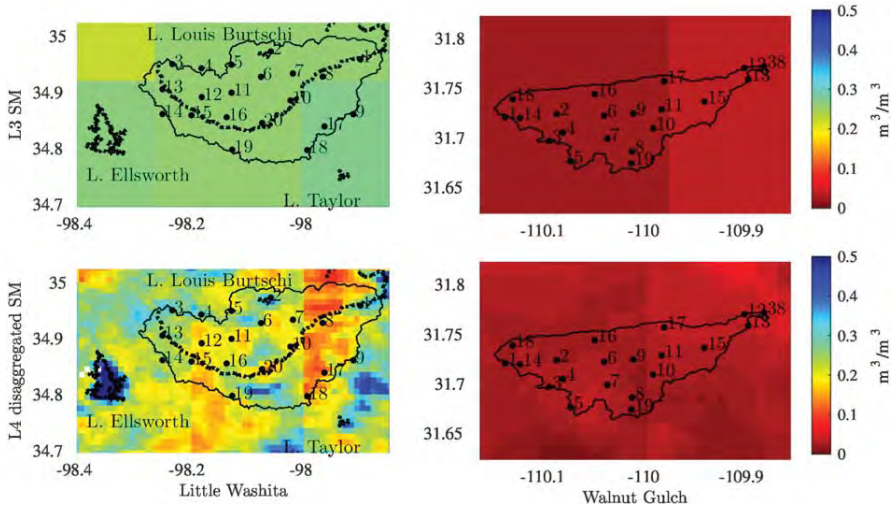


Fig. 6. Maps of L3 SM (CLF31A) and L4 disaggregated SM for LW (left column) and WG (right column) watersheds on 02/05/2011 and 01/05/2011 respectively. Solid black contours correspond to watershed boundaries. In the left column, the bold dotted line in the middle of the watershed correspond to the Little Washita river and the bold dotted contours to surrounding lakes.

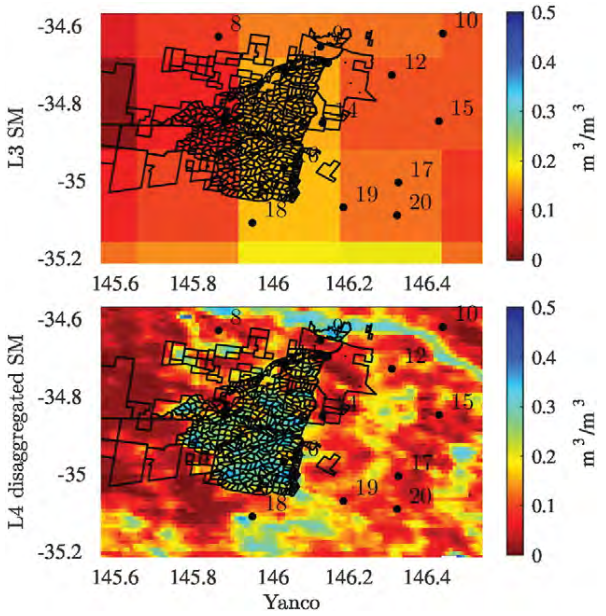


Fig. 7. Maps of L3 SM (CLF31D) and L4 disaggregated SM for Yanco area on 22/11/2010. Black lines represent the contours of Coleambally irrigated farms.

Finally, we show in Fig. 8 a series of C4DIS disaggregated outputs between the 4th and the 18th day of 2011. We can identify in detail the areas affected by the floods that affected the states of New South Wales and Victoria on those days. Likewise, we see how the dry-down process is faster in some small areas than in others (west of Yanco).

5.3. Spatial evaluation

In this section, the L4 and L3 SM products are compared at HR on a daily basis against the *in situ* measurements.

Table 4 shows daily statistics averaged over the periods of analysis. When comparing the statistics obtained for L3 and L4 products in MB and Yanco networks, it is noted an important enhancement of the S and the R values, ranging between 0.24–0.32 and 0.09–0.17, respectively. Results are consistent with the conditions of the area, especially those of Yanco (semi-arid climate with SM spatial heterogeneity dominated by irrigation). Spatial B is maintained while ubRMSD increases (around 0.02 m^3/m^3) which can be explained by the added uncertainty when combining data from different sources.

LW and WG statistics are much poorer than MB ones: R and S never exceed 0.11. The reasons for that can be found in both the algorithm and the conditions of the validation area. First, the L3 statistics (R and S) are much worse in the American than in the Australian networks, which may entail uncertainty present in the LR product that is propagated to the L4 product. Second, according to the preliminary statistical analysis (Section 5.1), the spatial σ distribution of WG is narrower and span over lower values than those of the Australian networks. The spatial variability cannot explain however the poor statistics of LW, since here the mean spatial σ is similar to the Australian ones (0.07 m^3/m^3 for Yanco, 0.06 m^3/m^3 for MB and LW). Another important aspect to take into consideration is the mismatch between the validation extent and the SMOS resolution. LW and WG cover only part of the surface of one SMOS pixel (~1/4 and ~1/12 of its equivalent surface, respectively), so the distribution of spatial σ may not be representative of the surface perceived by DISPATCH. All this suggests that a qualitative analysis of the area is strongly recommended.

The LW watershed has rolling relief and a variety of soil textures and vegetation types, which are not considered in the soil temperature equations of DISPATCH. Moreover, its extension is around 4 times smaller than the Yanco area: we can think that a higher heterogeneity within the 1 km pixel would hamper R and S statistics as

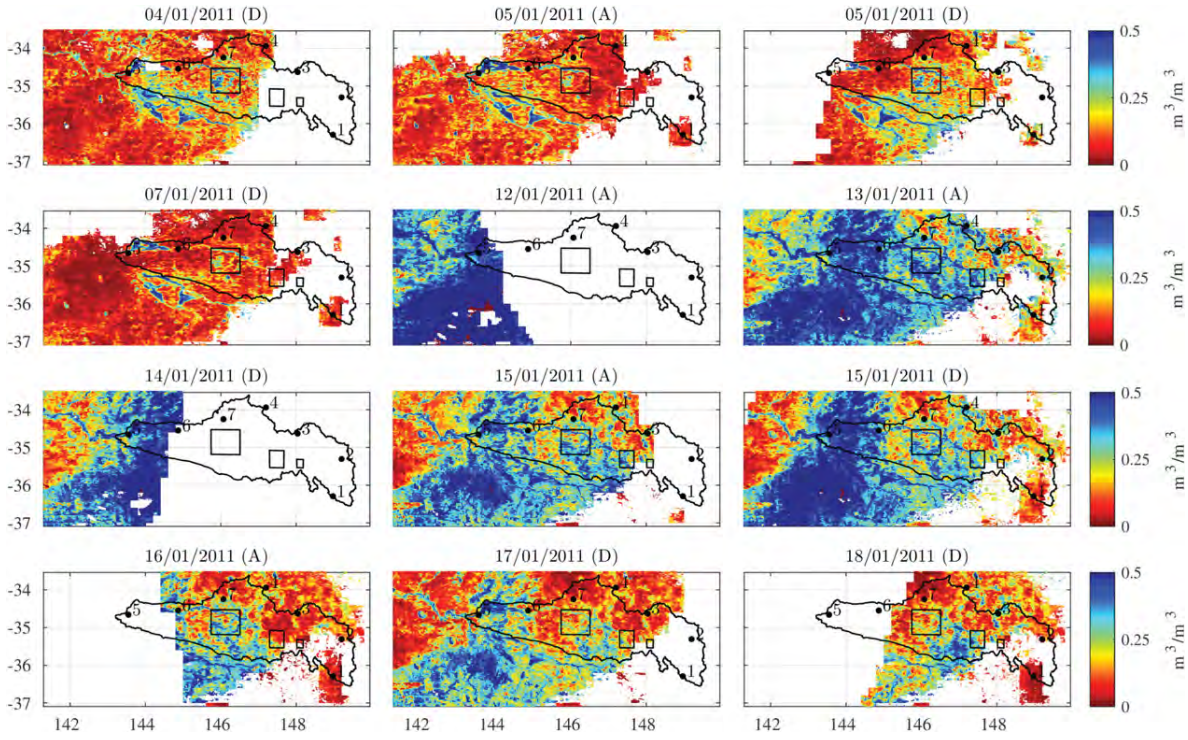


Fig. 8. Maps of L4 disaggregated SM for MB on the first days of January 2011, showing the progression of floods that affected New South Wales and Victoria states. The title of each image contains the date and the SMOS overpass ('A' for ascending, 'D' for descending).

Table 4
Spatial statistics of Yanco and MB for the period 01/06/2010 to 31/05/2011 and of LW and WG for the period 01/06/2010 to 31/12/2014. 'L3' refers to the comparison between L3 SM and *in situ* SM and 'L4' refers to the comparison of L4 disaggregated SM and *in situ* SM. 'A' stands for ascending orbit and 'D' for descending orbit. All the values are expressed in m^3/m^3, except for R and Number of days, which are unitless.

		Yanco		MB		LW		WG	
		L3	L4	L3	L4	L3	L4	L3	L4
S	A	0.064	0.309	0.086	0.403	0.003	0.047	0.004	0.110
	D	0.080	0.378	0.195	0.430	0.031	0.046	0.017	0.111
R	A	0.201	0.316	0.156	0.288	0.030	0.064	0.015	0.102
	D	0.194	0.363	0.251	0.335	0.115	0.057	0.042	0.111
B	A	0.018	0.021	0.031	0.035	0.023	0.016	0.031	0.026
	D	0.006	0.011	0.016	0.020	0.023	0.012	0.029	0.026
ubRMSD	A	0.072	0.094	0.082	0.103	0.063	0.076	0.030	0.037
	D	0.077	0.091	0.080	0.100	0.062	0.076	0.033	0.040
Nb	A	74		100		573		552	
Days	D	66		95		557		545	

well. Most importantly, LW climate is defined as sub-humid, so we can expect the link moisture-evaporation to be weak. Concerning WG, the soils are of fast infiltration (sands and gravels), which reduces the apparent SM spatial contrast at the satellite overpass times, a necessary condition for an accurate computation of the DISPATCH SMp parameter.

The comparison of the results here with previous versions of the algorithm can shed light on the pertinence of the choices made in the algorithm since Merlin et al., 2012. Regarding the most recent study, similar spatial statistics for MB and Yanco can be found in Malbéteau et al., 2016, which proves that the performance of the processor is coherent with that of the prototype algorithm. The remaining differences are originated by two factors. First, in our aim to assess the qualities of the entire C4DIS processor, we use as LR SM reference the original SMOS CLF31A/D product, while Malbéteau et al., 2016 employed a reprojected form of the same product used by DISPATCH, which was a reasonable choice from the algorithm point of view. Second, the C4DIS post-processor clips to zero the negative values produced by DISPATCH, a module that was not still implemented at the time of Malbéteau et al., 2016.

Another two former validation campaigns of DISPATCH showed better correspondence with *in situ* measurements, but they were accomplished for specific areas with known high-evaporative demand and for no more than a dozen of dates. For the Murrumbidgee catchment and AACES-I campaign (Merlin et al., 2012), and the Catalunya campaign (Merlin et al., 2013), summer 2010 and 2011 respectively, the correlation values were close to the double of those obtained for MB in this study. However, the AACES-based study also reported negative values for those dates with very dry homogeneous SM scenes. This confirms our hypothesis for WG, were the large number of 'flat' SM scenes is probably behind the unsatisfactory statistics. In the same article of 2012, the AACES-II results (winter), allowed to presum that the weak evaporation-SM coupling was behind negative R values. Our statistics for LW seem to confirm this point, but since the mean R is higher, it suggests that the algorithm might be useful for some periods of the year.

5.4. Temporal evaluation

For the temporal analysis, we consider the same period and datasets as in Section 5.3. We compute statistics on the concatenation of all the SM series within a network. Table 5 displays temporal statistics for the four validation networks. Regarding Yanco and MB, the S metric is better for the HR SM product (between 0.12 and 0.18 higher), which is consistent with the spatial evaluation results. R is slightly degraded in Yanco while maintained in MB. This, and the increase in ubRMSD, can be explained by the temporal uncertainty induced by the processor when considering as inputs observations acquired in different days and times. These results are consistent with previous validation studies of DISPATCH: Merlin et al., 2013 showed that the temporal S could increase between 0.15 to 0.25 after disaggregation, while R being maintained or increased and ubRMSE increased.

In the case of LW, the disaggregated SM (L4) has a slightly better S when compared to *in situ* SM than does L3 SM for both orbits (improvement of +0.06 for A orbit and of +0.03 for D orbit). The same evaluation holds for WG (improvement of +0.05 and of +0.08 for A and D orbits respectively). Like in the Yanco case, disaggregation slightly degrades R and ubRMSD for both SMOS orbits, showing again the increase of random uncertainties attributed to the models and data used by DISPATCH.

Table 5
Temporal statistics of Yanco and for the period 01/06/2010 to 31/05/2011, and of LW and WG for the period 01/06/2010 to 31/12/2014. 'L3' refers to the comparison between L3 SM and *in situ* SM and 'L4' refers to the comparison of L4 disaggregated SM and *in situ* SM. In the second column, 'A' stands for ascending orbit and 'D' for descending orbit. All the values are expressed in m^3/m^3, except for R and Number of points, which are unitless, and RFI percentage, which is in %.

		Yanco		MB		LW		WG	
		L3	L4	L3	L4	L3	L4	L3	L4
S	A	0.368	0.489	0.363	0.538	0.406	0.463	0.490	0.544
	D	0.333	0.465	0.383	0.542	0.415	0.441	0.381	0.458
R	A	0.432	0.370	0.321	0.377	0.468	0.434	0.468	0.436
	D	0.369	0.356	0.361	0.368	0.460	0.410	0.352	0.366
B	A	0.019	0.023	0.033	0.027	0.023	0.017	0.031	0.026
	D	0.004	0.014	0.020	0.019	0.025	0.014	0.030	0.026
ubRMSD	A	0.090	0.120	0.105	0.118	0.078	0.088	0.044	0.051
	D	0.095	0.118	0.095	0.118	0.077	0.088	0.052	0.056
RFI perc.	A	0.000	–	0.248	–	1.893	–	1.958	–
	D	0.000	–	0.000	–	1.893	–	1.562	–
Nb points	A	754		754		1429		9027	
	D	723		723		1409		9337	

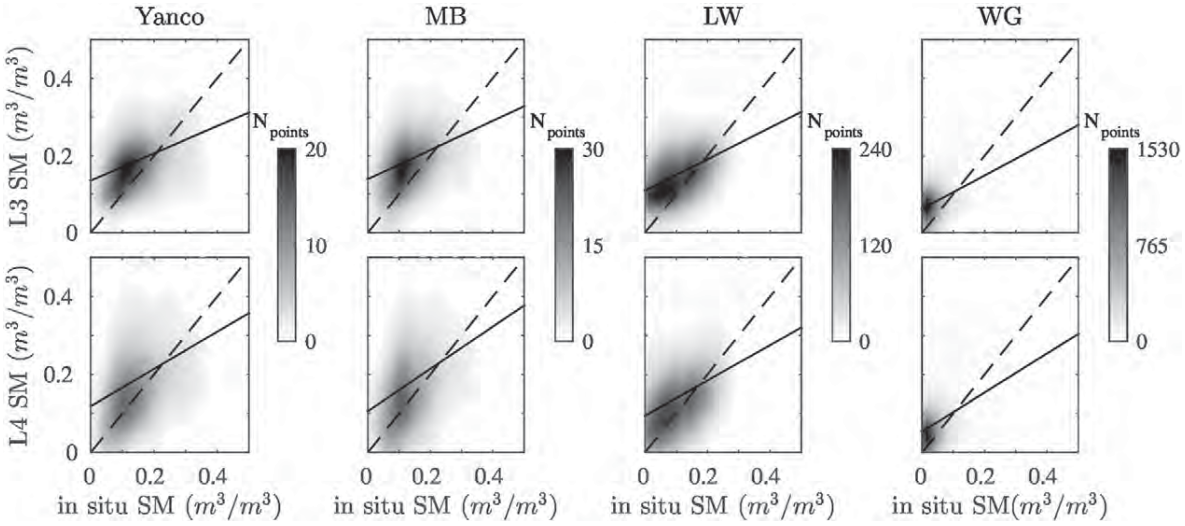


Fig. 9. Scatterplots of original L3 SM (1st row) and L4 disaggregated SM (2nd row) versus *in situ* measurements for both A and D orbits. The samples here correspond to the periods 06/2010 to 05/2011 for MB and Yanco, and 06/2010 to 12/2014 for LW and WG. Dashed line represents the 1:1 slope and the solid line corresponds to the linear regression line (S statistic).

According to our preliminary analysis on *in situ* temporal σ and SM samples, WG should at least behave differently with respect to the other networks (much narrower distribution of SM values, skewed to the dry section of the range and lower σ variability). However, no significant differences are found in the temporal statistics.

Differences can be appreciated more easily through qualitative inspection of scatter plots (Fig. 9). In Yanco and MB plots, the increase in ubRMSD is observed in the more dispersed cloud of points, although the distribution appears slightly closer and more symmetric around the 1:1 line. In the case of LW, we can see that for drier SM conditions (<0.15 m^3/m^3), disaggregated values are closer to *in situ* values and become equally distributed around the 1:1 line. Since LW climate is subhumid, evapotranspiration processes are mainly energy-driven; however, we can expect them to be moisture-driven during periods with lower water availability and higher temperatures like summer. This is confirmed in Fig. 10, which shows the scatter plot for LW summers. Regarding WG, the scatter plots show no major differences between L3 and L4 data. This is consistent with the very low spatial and temporal *in situ* σ: DISPATCH is operating at the limit of its nominal range at 1 km resolution and the amount of information obtained is not more important than the uncertainty introduced. It outlines also the importance of qualitative assessments: although LW and WG show similar global spatial and temporal statistics, C4DIS disaggregated fields, which are not of interest in WG, are valuable in the case of LW summers.

5.5. Analysis of the STD and COUNT datasets

As introduced in Section 3.6, the STD and COUNT datasets can help derive conclusions on the quality of the SM_HR values. In this section, we evaluate spatial and temporal statistics on SM samples with different corresponding STD and COUNT values. We first select the samples with values falling inside a given STD or COUNT range of values; then, we compute statistics on the *in situ*, L3 and L4 values for those samples. This analysis is conducted on MB and Yanco networks as USDA networks still show low statistics after filtering for STD and COUNT values. Herein, we use the gain metrics introduced in Section 4.4, which will simplify the task of comparison between bins of STD and COUNT.

Table 6 shows spatial statistics for MB and Yanco divided in 3 ranges of STD (<0.03 m^3/m^3, 0.03–0.07 m^3/m^3, >0.07 m^3/m^3). Note that the total number of days analyzed drops drastically when STD or COUNT filtering is applied to spatial metrics. This is as expected since for a given time stamp, the samples have STD and COUNT values that belong to different bins and we need at least 5 samples in the same bin to compute statistics. C4DIS SM dataset exhibits the lowest correlation (S and R) and the highest error (ubRMSD) with *in situ* when most of the pixels have high STD (>0.07 m^3/m^3). This seems plausible since large ubRMSD values can be produced by forcing events (rain, irrigation) in the 3-days window of DISPATCH, so the final SM_HR values would contain high uncertainty. We cannot generalize any behavior in performances for the medium and lower STD ranges (<0.07 m^3/m^3) since MB and Yanco show different trends. If we consider only Yanco, which is a much more homogeneous area in terms of climate and landscape properties, we can conclude that, regardless of the bias, the rest of spatial metrics are better as STD gets lower. Whether this is applicable to other homogeneous areas or not need to be the subject of additional studies.

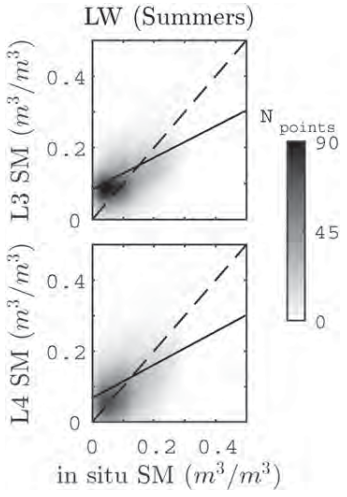


Fig. 10. Scatterplot of L3 SM (1st row) and L4 disaggregated SM (2nd row) against *in situ* SM samples for LW network for summer periods (June, July and August months of years 2010 to 2014). Dashed line represents the 1:1 slope and the solid line corresponds to the linear regression line (S statistic).

Table 6
Spatial statistics as a function of the values of the STD dataset for MB and Yanco areas from 01/06/2010 to 31/05/2011. Best statistics are outlined and in italics. Last line of Yanco table (in bold) should not be considered because it refers to only one day of statistics.

STD	Yanco					MB				
	G_{EFFI}	G_{ACCU}	G_{ROBU}	G_{ubRMSD}	N_{days}	G_{EFFI}	G_{ACCU}	G_{ROBU}	G_{ubRMSD}	N_{days}
<0.03	*0.27*	*0.24*	−0.22	*0.05*	11	0.15	*0.11*	−0.12	*−0.04*	45
0.03–0.07	0.13	0.06	*−0.11*	−0.10	39	*0.17*	0.05	−0.03	−0.07	108
>0.7	**−0.47**	**−0.12**	**−0.42**	**−0.57**	**1**	−0.02	−0.09	*0.05*	−0.28	16

Table 7
Spatial statistics as a function of the COUNT dataset for MB and Yanco areas from 01/06/2010 to 31/05/2011. Best statistics are outlined and in italics.

COUNT	Yanco					MB				
	G_{EFFI}	G_{ACCU}	G_{ROBU}	G_{ubRMSD}	N_{days}	G_{EFFI}	G_{ACCU}	G_{ROBU}	G_{ubRMSD}	N_{days}
1–8	0.16	0.08	−0.16	−0.16	69	0.16	0.07	*−0.05*	−0.11	143
9–16	0.12	0.16	−0.15	−0.07	22	0.14	−0.01	−0.12	−0.15	51
17–24	*0.44*	*0.29*	*−0.08*	*0.06*	11	*0.35*	*0.15*	−0.24	*−0.04*	13

Spatial statistics are also filtered for COUNT values (Table 7). In this respect, statistics are better for large values of COUNT (17–24 datasets). However, the number of days used in this computation is low in the same line, if possible so the results may not be accurate.

Regarding the temporal domain, Yanco shows a deterioration of the metrics as STD increases (Table 8), which is consistent with the preliminary *in situ* spatial analysis and would be mainly due to the uncertainty added when precipitation or irrigation take place in the 3-days window of DISPATCH. Such trend is not revealed in the MB data (same table), and conclusions are difficult to be derived given the high heterogeneity within the network.

Concerning the COUNT dataset, Table 9 clearly shows that temporal statistics improve as COUNT increases. This seems to confirm that the methodology of averaging of the disaggregated ensemble helps to reduce random uncertainties in the temporal domain.

6. Conclusions

The C4DIS processor is the new SMOS L4 processor of the French ground segment CATDS, which provides global maps of disaggregated SM at 1 km resolution. The C4DIS processor is the operational version of the DISPATCH prototype (Merlin et al., 2012, 2013). DISPATCH disaggregates LR SM observations using HR soil temperature data. It models the physical link between soil temperature, evaporation and moisture with a semi-empirical SEE model and a first-order Taylor series expansion around the SM observation. The soil temperature is derived from the combination of LST, NDVI and elevation information. The C4DIS processor uses the SM dataset of the SMOS 1-day L3 CLF31A/D product from CATDS, the LST dataset of the MODIS MOD11A1 and MYD11A1 products from LP DAAC services, the NDVI dataset from the MOD13A2 product from LP DAAC services, and the elevation dataset from the GTOPO30 product from the USGS Eros Data Center.

In this study, the C4DIS products were evaluated for four different geographical areas: the Murrumbidgee validation network and the Yanco area for the period 06/2010 to 05/2011, and the Little Washita and Walnut Gulch networks for the period 01/2010 to 12/2014. The objective was to provide a first assessment of the processor under different climatic and land conditions. The performance was assessed by comparing the disaggregated (L4) and non-disaggregated (L3) SM datasets against the *in situ* measurements in both the spatial and temporal domains. The *in situ* SM data was statistically analyzed beforehand in

Table 8
Temporal statistics as a function of the STD dataset for MB and Yanco areas from 01/06/2010 to 31/05/2011. Best statistics are outlined and in italics.

STD	Yanco					MB				
	G_{EFFI}	G_{ACCU}	G_{ROBU}	G_{ubRMSD}	$N_{samples}$	G_{EFFI}	G_{ACCU}	G_{ROBU}	G_{ubRMSD}	$N_{samples}$
<0.025	*0.18*	*0.04*	−0.30	*−0.06*	472	*0.16*	0.03	−0.81	−0.06	904
0.025–0.040	0.04	−0.06	*−0.14*	−0.11	813	0.11	0.01	*0.03*	*−0.10*	1459
0.040–0.055	0.03	−0.04	−0.41	−0.18	192	0.13	*0.06*	−0.12	−0.03	475

Table 9
Temporal statistics as a function of the COUNT dataset for MB and Yanco areas from 01/06/2010 to 31/05/2011. Best statistics are outlined and in italics.

COUNT	Yanco					MB				
	G_{EFFI}	G_{ACCU}	G_{ROBU}	G_{ubRMSD}	$N_{samples}$	G_{EFFI}	G_{ACCU}	G_{ROBU}	G_{ubRMSD}	$N_{samples}$
1–8	0.08	−0.06	−0.21	−0.15	965	0.14	0.02	0.08	−0.08	1910
9–16	0.17	−0.02	−0.18	−0.12	386	0.19	0.02	0.04	−0.09	737
17–24	*0.22*	*0.19*	*0.35*	*0.01*	126	*0.21*	*0.15*	*0.44*	*−0.03*	191

order to predict the suitability of the C4DIS processor for each area. We also evaluated the output COUNT and STD datasets as potential sources of information for quality assessment.

The evaluation of the disaggregated SM dataset in Murrumbidgee and Yanco brought results in coherence with previous versions of DISPATCH (Malbéteau et al., 2016; Merlin et al., 2012), and presented improvements on the spatial correlation in the range 0.09–0.17. Similar enhancements were present in the temporal domain. Additionally, C4DIS SM maps succeeded to reveal spatial heterogeneities (rivers, irrigation areas, floods).

Little Washita and Walnut Gulch showed very low spatial metric values for both non-disaggregated and disaggregated SM fields, though disaggregation slightly improved the statistics. For the Little Washita, the scatter plots revealed that the performances were better in the dry section of the SM range (<0.15 m^3/m^3) and during summers, meaning that the improvement in spatial representation was possible under moisture-driven evaporation periods. Visual assessment of C4DIS SM maps showed that the disaggregated product was capable of revealing the presence of water bodies in the surrounding areas namely lakes.

For the Walnut Gulch network, the poor spatial correspondence with *in situ* was easily explained by the preliminary statistical analysis that we conducted on *in situ* SM data: this revealed very low spatial variability (mean spatial σ was equal to 0.03 m^3/m^3), which is one of the essential conditions for a good performance of the algorithm. The evaluation of this network brought to view that the algorithm needs to be improved to adapt to all types of soil. Although Walnut Gulch watershed also has a moisture-controlled evaporative profile (semi-arid to arid climate) like the Australian areas, the soil is mainly sandy with high infiltration rates, which obstructs the detection of surface SM variations by the algorithm.

When evaluating the temporal behavior of the (non-disaggregated and C4DIS) satellite SM series, we found an improvement of the slope of the regression line between C4DIS and the *in situ* data. The correlation was slightly hampered, especially in LW and WG, and the standard deviation of the differences also increased. This was likely to be caused by the increase in uncertainty associated with the use of multi-satellite data.

With the aim of making the C4DIS products useful in a global perspective, we evaluated how the other two output datasets, COUNT and STD, could help in the future definition of a quality flag. We showed that for a homogeneous area like Yanco, spatial and temporal metrics were better as STD decreased. Consistently, large COUNT values helped to decrease the random uncertainties and they improved temporal statistics. In this area, heterogeneity is mainly driven by precipitation and irrigation, and STD was directly linked to such events. On the contrary, STD and COUNT could not give sufficient information for quality control in more heterogeneous areas (like the entire Murrumbidgee), so we concluded that output C4DIS datasets must be combined with ancillary information like precipitation or other heterogeneity-related data sources to implement a good quality flag field.

In conclusion, the C4DIS processor performs well in regions with SM spatial variability mainly produced by external forcing agents (precipitation or irrigation). Additionally, the degree of variability must be enough so the application of a disaggregation technique is advisable. These two characteristics are mainly conditioned by the climate (semi-arid), soil properties (with moderate drainage), and land properties (low topography, quasi-homogeneous land cover). The proper performance of the processor can be predicted by looking at the *in situ* SM variability and assessing qualitatively the enounced characteristics. The C4DIS SM products can be evaluated by applying ordinary spatial and temporal statistics, visual inspection of maps as well as using the STD and COUNT datasets on homogeneous areas. In the future, including meteorological forcing (solar radiation, air temperature, wind speed and air humidity at 2 m; Stefan et al., 2015), precipitation (Djamai et al., submitted for publication), soil texture (Merlin et al., submitted for publication) and solar exposure (Malbéteau et al., submitted for publication) as ancillary data will help improve DISPATCH and elaborate a quality control dataset that will enlarge the applicability areas of the processor.

Acknowledgements

Initial setup and maintenance of the Murrumbidgee monitoring network used in this study was funded by the Australian Research Council (DP0343778, DP0557543) and by the CRC for Catchment Hydrology.

The MODIS products were retrieved from the online server http://e4ftl01.cr.usgs.gov/, courtesy of the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota.

The SMOS products were obtained from the Centre Aval de Traitement des Données SMOS (CATDS), operated for the "Centre National d'Etudes Spatiales" (CNES, France) by IFREMER (Brest, France).

This study was supported by the CNES "Terre, Océan, Surfaces Continentales, Atmosphère" program and by the French "Agence Nationale de la Recherche" MIXMOD-E project (ANR-13-JS06-0003-01).

References

Al Bitar, A., Leroux, D. J., Kerr, Y. H., Merlin, O., Richaume, P., Sahoo, A., & Wood, E. F. (2012). Evaluation of SMOS soil moisture products Over Continental U.S. using the SCAN/SNOTEL network. *IEEE Transactions on Geoscience and Remote Sensing, 50*(5), 1572–1586.

Albergel, C., Brocca, L., Wagner, W., de Rosnay, P., & Calvet, J. C. (2013). Selection of performance metrics for global soil moisture products: The case of ascat soil moisture product. *Remote sensing of energy fluxes and soil moisture content* (pp. 431–448).

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

Allen, P. B., & Naney, J. W. (1991). *Hydrology of the Little Washita River Watershed, Oklahoma: Data and analyses*. United States Department of Agriculture, Agricultural Research Service (ARS-90).

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

Bindlish, R. (2015). Global soil moisture from the Aquarius/SAC-D satellite: description and initial assessment. *IEEE Geoscience and Remote Sensing Letters, 12*(5), 923–927. http://dx.doi.org/10.1109/LGRS.2014.2364151.

Budyko, M. I. (1956). Heat balance of the Earth's surface. *Leningrad*.

Carlson, T. N. (2007). An overview of the "Triangle method" for estimating surface evapotranspiration and soil moisture from satellite imagery. *Sensors, 7*(8), 1612–1629. http://dx.doi.org/10.3390/s7081612.

Carlson, T. N., Gillies, R. R., & Perry, E. M. (1994). A method to make use of thermal infrared temperature and NDVI measurements to infer surface soil water content and fractional vegetation cover. *Remote Sensing Reviews*. http://dx.doi.org/10.1080/02757259409532220.

Chauhan, N. S., Miller, S., & Ardanuy, P. (2003). Spaceborne soil moisture estimation at high resolution: a microwave-optical/IR synergistic approach. *International Journal of Remote Sensing, 24*(22), 4599–4622. http://dx.doi.org/10.1080/0143116031000156837.

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

Cosh, M. H., Jackson, T. J., Bindlish, R., & Prueger, J. H. (2004). Watershed scale temporal and spatial stability of soil moisture and its role in validating satellite estimates. *Remote Sensing of Environment, 92*, 427–435. http://dx.doi.org/10.1016/j.rse.2004.02.016.

Cosh, M. H., Jackson, T. J., Moran, S., & Bindlish, R. (2008). Temporal persistence and stability of surface soil moisture in a semi-arid watershed. *Remote Sensing of Environment, 112*(2), 304–313. http://dx.doi.org/10.1016/j.rse.2007.07.001.

Cosh, M. H., Jackson, T. J., Starks, P. J., & Heathman, G. (2006). Temporal stability of surface soil moisture in the Little Washita River watershed and its applications in satellite soil moisture product validation. *Journal of Hydrology, 323*(1–4), 168–177. http://dx.doi.org/10.1016/j.jhydrol.2005.08.020.

Daly, E., & Porporato, A. (2005). A review of soil moisture dynamics: From rainfall infiltration to ecosystem response. *Environmental Engineering Science*. http://dx.doi.org/10.1089/ees.2005.22.9.

Das, N. N., Entekhabi, D., & Njoku, E. G. (2011). An algorithm for merging SMAP radiometer and radar data for high-resolution soil-moisture retrieval. *IEEE Transactions on Geoscience and Remote Sensing, 49*(5), 1504–1512. http://dx.doi.org/10.1109/TGRS.2010.2089526.

Das, N. N., Entekhabi, D., Njoku, E. G., Shi, J. J. C., Johnson, J. T., & Colliander, A. (2014). Tests of the SMAP combined radar and radiometer algorithm using airborne field campaign observations and simulated data. *IEEE Transactions on Geoscience and Remote Sensing, 52*(4), 2018–2028. http://dx.doi.org/10.1109/TGRS.2013.2257605.

Delwart, S., Bouzinac, C., Wursteisen, P., Berger, M., Drinkwater, M., Martín-Neira, M., & Kerr, Y. H. (2008). SMOS validation and the COSMOS campaigns. *IEEE Transactions on Geoscience and Remote Sensing*, *46*(63), 695–703. http://dx.doi.org/10.1109/TGRS.2007.914811.

Dirmeyer, P. A. (2000). Using a global soil wetness dataset to improve seasonal climate simulation. *Journal of Climate*, *13*, 2900–2922.

Djamai, N., Magagi, R., Goïta, K., Merlin, O., Kerr, Y. H., & Roy, A. (2015). Downscaling satellite-based soil moisture for cloudy days using the DISPATCH algorithm and CLASS land surface scheme. (submitted for publication) *Remote Sensing of Environment*.

Douville, H. (2004). Relevance of soil moisture for seasonal atmospheric predictions: Is it an initial value problem? *Climate Dynamics*, *22*(4), 429–446. http://dx.doi.org/10.1007/s00382-003-0386-5.

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

Drusch, M. (2007). Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's integrated forecast system and the TMI soil moisture data set. *Journal of Geophysical Research-Atmospheres*, *112*(3). http://dx.doi.org/10.1029/2006JD007478.

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

Entekhabi, D., Reichle, R. H., Koster, R. D., & Crow, W. T. (2010b). Performance metrics for soil moisture retrievals and application requirements. *Journal of Hydrometeorology*, *11*, 832–840. http://dx.doi.org/10.1175/2010JHM1223.1.

Fang, B., Lakshmi, V., Bindlish, R., Jackson, T. J., Cosh, M. H., & Basara, J. (2013). Passive microwave soil moisture downscaling using vegetation index and skin surface temperature. *Vadose Zone Journal*, *12*(3). http://dx.doi.org/10.2136/vzj2013.05.0089.

Guérif, M., & Duke, C. (2000). Adjustment procedures of a crop model to the site specific characteristics of soil and crop using remote sensing data assimilation. *Agriculture, Ecosystems & Environment*. http://dx.doi.org/10.1016/S0167-8809(00)00168-7.

Jackson, T. J., Bindlish, R., Cosh, M. H., Zhao, T., Starks, P. J., Bosch, D. D., ... Leroux, D. J. (2012). Validation of soil moisture and Ocean Salinity (SMOS) soil moisture over watershed networks in the U.S. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(5), 1530–1543. http://dx.doi.org/10.1109/TGRS.2011.2168533.

Jackson, T. J., Cosh, M. H., Bindlish, R., Starks, P. J., Bosch, D. D., Seyfried, M., ... Du, J. (2010). Validation of advanced microwave scanning radiometer soil moisture products. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(12), 4256–4272. http://dx.doi.org/10.1109/TGRS.2010.2051035.

Jiang, L., & Islam, S. (2003). An intercomparison of regional latent heat flux estimation using remote sensing data. *International Journal of Remote Sensing*. http://dx.doi.org/10.1080/01431160210154821.

Kerr, Y. H., & Njoku, E. G. (1990). Semiempirical model for interpreting microwave emission from semiarid land surfaces as seen from space. *IEEE Transactions on Geoscience and Remote Sensing*, *28*(3), 384–393. http://dx.doi.org/10.1109/36.54364.

Kerr, Y. H., Berthon, L., Mialon, A., Cabot, F., Al Bitar, A., Richaume, P., ... Jacquette, E. (2014). *CATDS LEVEL 3 – data product description – soil moisture and brightness temperature*.

Kerr, Y. H., Jacquette, E., Al Bitar, A., Cabot, F., Mialon, A., & Richaume, P. (2013). CATDS SMOS L3 soil moisture retrieval processor. *Algorithm Theoretical Baseline Document (ATBD)*.

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. -P., Ferrazzoli, P., Mahmoodi, A., ... Delwart, S. (2012). The SMOS soil moisture retrieval algorithm. *Geoscience and Remote Sensing*, *50*(5), 1384–1403.

Kerr, Y. H., Waldteufel, P., Wigneron, J. -P., Delwart, S., Cabot, F., Boutin, J., ... Mecklenburg, S. (2010). The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proceedings of the IEEE*, *98*(5), 666–687. http://dx.doi.org/10.1109/JPROC.2010.2043032.

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*(8), 1729–1735. http://dx.doi.org/10.1109/36.942551.

Kim, J., & Hogue, T. S. (2012). Improving spatial soil moisture representation through integration of AMSR-E and MODIS products. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(2), 446–460. http://dx.doi.org/10.1109/TGRS.2011.2161318.

Laio, F., Porporato, A., Ridolfi, L., & Rodríguez-Fernández, N. J. (2002). On the seasonal dynamics of mean soil moisture. *Journal of Geophysical Research-Atmospheres*. http://dx.doi.org/10.1029/2001JD001252.

Leroux, D. J., Kerr, Y. H., Al Bitar, A., Bindlish, R., Jackson, T. J., Berthelot, B., & Portet, G. (2013). Comparison between SMOS, VUA, ASCAT, and ECMWF soil moisture products over four watersheds in U.S. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(3), 1–10.

Lievens, H., Tomer, S. K., Al Bitar, A., De Lannoy, G. J. M., Drusch, M., Dumedah, G., ... Pauwels, V. R. N. (2015). SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sensing of Environment*, *168*, 146–162. http://dx.doi.org/10.1016/j.rse.2015.06.025.

Malbéteau, Y., Merlin, O., Gascoin, S., Gastellu, J. P., Olivera, L., Mattar, C., & Khabba, S. (2016n). Correcting land surface temperature data for elevation and illumination effects in mountainous areas: A case study using ASTER data over the imlil valley, Morocco. (submitted for publication) *Remote Sensing of Environment*.

Malbéteau, Y., Merlin, O., Molero, B., Rüdiger, C., & Bacon, S. (2016). DisPATCh as a tool to evaluate coarse-scale remotely sensed soil moisture using localized in situ measurements: Application to SMOS and AMSR-E data in Southeastern Australia. *International Journal of Applied Earth Observation and Geoinformation*, *45*(Part B), 221–234. http://dx.doi.org/10.1016/j.jag.2015.10.002.

Manabe, S. (1969). Climate and the ocean circulation. I. The atmospheric circulation and the hydrology of the Earth's surface. *Monthly Weather Review*, *97*(11), 739–774.

Merlin, O., Al Bitar, A., Walker, J. P., & Kerr, Y. H. (2009). A sequential model for disaggregating near-surface soil moisture observations using multi-resolution thermal sensors. *Remote Sensing of Environment*, *113*(10), 2275–2284. http://dx.doi.org/10.1016/j.rse.2009.06.012.

Merlin, O., Al Bitar, A., Walker, J. P., & Kerr, Y. H. (2010a). An improved algorithm for disaggregating microwave-derived soil moisture based on red, near-infrared and thermal-infrared data. *Remote Sensing of Environment*, *114*(10), 2305–2316. http://dx.doi.org/10.1016/j.rse.2010.05.007.

Merlin, O., Chehbouni, A., Boulet, G., & Kerr, Y. H. (2006). Assimilation of disaggregated microwave soil moisture into a hydrologic model using coarse-scale meteorological data. *Journal of Hydrometeorology*. http://dx.doi.org/10.1175/JHM552.1.

Merlin, O., Duchemin, B., Hagolle, O., Jacob, F., Coudert, B., Chehbouni, G., ... Kerr, Y. H. (2010b). Disaggregation of MODIS surface temperature over an agricultural area using a time series of formosat-2 images. *Remote Sensing of Environment*, *114*(11), 2500–2512. http://dx.doi.org/10.1016/j.rse.2010.05.025.

Merlin, O., Escorihuela, M. -J., Mayoral, M. A., Hagolle, O., Al Bitar, A., & Kerr, Y. H. (2013). Self-calibrated evaporation-based disaggregation of SMOS soil moisture: An evaluation study at 3 km and 100 m resolution in Catalunya, Spain. *Remote Sensing of Environment*, *130*(2013), 25–38.

Merlin, O., Malbéteau, Y., Notfi, Y., Bacon, S., Er-raki, S., Khabba, S., & Jarlan, L. (2015). Performance metrics for soil moisture downscaling methods: Application to DISPATCH data in central Morocco. *Remote Sensing*, *7*(4), 3783–3807. http://dx.doi.org/10.3390/rs70403783.

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

Merlin, O., Rüdiger, C., Richaume, P., Al Bitar, A., Mialon, A., Walker, J. P., & Kerr, Y. H. (2010c). Disaggregation as a top-down approach for evaluating 40 km resolution SMOS data using point-scale measurements: Application to AACES-1. *Remote sensing for agriculture, ecosystems, and hydrology XII* (pp. *7824*), 78240i\rhttp://dx.doi.org/10.1117/12.865751 (\r666).

Merlin, O., Stefan, V. G., Amazirh, A., Chanzy, A., Ceschia, E., Tallec, T., ... Khabba, S. (2016). Modeling soil evaporation efficiency in a range of soil and atmospheric conditions: A downward approach based on multi-site data. (submitted for publication) *Water Resources Research*.

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

Mladenova, I., Lakshmi, V., Jackson, T. J., Walker, J. P., Merlin, O., & de Jeu, R. A. M. (2011). Validation of AMSR-E soil moisture using L-band airborne radiometer data from National Airborne Field Experiment 2006. *Remote Sensing of Environment*, *115*(8), 2096–2103. http://dx.doi.org/10.1016/j.rse.2011.04.011.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). Introduction to the theory of statistics. *McGrawHill series in probability and statistics. Vol. 3*, (Retrieved from http://www.librarything.com/work/1154157/book/32217714).

Moran, M. S., Clarke, T. R., Inoue, Y., & Vidal, A. (1994). Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index. *Remote Sensing of Environment*. http://dx.doi.org/10.1016/0034-4257(94)90020-5.

Narayan, U., Lakshmi, V., & Jackson, T. J. (2006). High-resolution change estimation of soil moisture using L-band radiometer and radar observations made during the SMEX02 experiments. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(6), 1545–1554. http://dx.doi.org/10.1109/TGRS.2006.871199.

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

Noilhan, J., & Planton, S. (1989). A simple parameterization of land surface processes for meteorological models. *Monthly Weather Review*, *117*(3), 536–549.

Oliva, R., Daganzo-Eusebio, E., Kerr, Y. H., Mecklenburg, S., Nieto, S., Richaume, P., & Gruhier, C. (2012). SMOS radio frequency interference scenario: Status and actions taken to improve the RFI environment in the 1400–1427-MHZ passive band. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(5 PART 1), 1427–1439. http://dx.doi.org/10.1109/TGRS.2012.2182775.

Panciera, R., Walker, J. P., Jackson, T. J., Gray, D. a., Tanase, M. a., Ryu, D., ... Hacker, J. M. (2014). The soil moisture active passive experiments (SMAPEx): Toward soil moisture retrieval from the SMAP mission. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(1), 490–507. http://dx.doi.org/10.1109/TGRS.2013.2241774.

Peischl, S., Walker, J. P., Rüdiger, C., Ye, N., Kerr, Y. H., Kim, E., ... Allahmoradi, M. (2012). The AACES field experiments: SMOS calibration and validation across the Murrumbidgee River catchment. *Hydrology and Earth System Sciences*, *16*(6), 1697–1708. http://dx.doi.org/10.5194/hess-16-1697-2012.

Piles, M., Camps, A., Vall-Llossera, M., Corbella, I., Panciera, R., Rüdiger, C., ... Walker, J. P. (2011). Downscaling SMOS-derived soil moisture using MODIS visible/infrared data. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(9), 3156–3166. http://dx.doi.org/10.1109/TGRS.2011.2120615.

Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, *42*(1), 59–66.

Salkind, N. J. (2010). Standard error of the estimate. In SAGE (Ed.), *Encyclopedia of research design. Vol. 3.* (pp. 1426–1430) (London).

Schmugge, T. J. (1998). Applications of passive microwave observations of surface soil moisture. *Journal of Hydrology*, *212-213*(1–4), 188–197. http://dx.doi.org/10.1016/S0022-1694(98)00209-1.

Smith, A. B., Walker, J. P., Western, A. W., Young, R. I., Ellett, K. M., Pipunic, R. C., ... Richter, H. (2012).). The Murrumbidgee soil moisture monitoring network data set. *Water Resources Research*, *48*(7), W07701. http://dx.doi.org/10.1029/2012WR011976.

Solano, R., Didan, K., Jacobson, A., & Huete, A. (2010). *MODIS Vegetation Index User's Guide. MOD13 Series, Vol. v2.0*, .

Stefan, V. G., Merlin, O., Er-Raki, S., Escorihuela, M. -J., & Khabba, S. (2015). Consistency between in situ, model-derived and image-based soil temperature endmembers: towards a robust data-based model for multi-resolution monitoring of crop evapotranspiration. *Remote Sensing, 7*(8), 10444–10479.

Walker, J. P., & Houser, P. R. (2004). Requirements of a global near-surface soil moisture satellite mission: accuracy, repeat time, and spatial resolution. *Advances in Water Resources*, *27*(8), 785–801. http://dx.doi.org/10.1016/j.advwatres.2004.05.006.

Wan, Z. (1999). *MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD).*

Wan, Z. (*2006*). *MODIS Land Surface Temperature Products Users' Guide - Collection 5.* South Dakota: Sioux Falls (Retrieved from http://www.icess.ucsb.edu/modis/LstUsrGuide/MODIS_LST_products_Users_guide_C5.pdf).

Wanders, N., Bierkens, M., de Jong, S. M., de Roo, A., & Karssenberg, D. (2014). The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models. *Water Resources Research*, *50*(8), 6874–6891. http://dx.doi.org/10.1002/2013WR014639.

Wigneron, J. -P., Kerr, Y. H., Waldteufel, P., Saleh, K., Escorihuela, M. -J., Richaume, P., ... Schwank, M. (2007). L-band microwave emission of the biosphere (L-MEB) model: Description and calibration against experimental data sets over crop fields. *Remote Sensing of Environment*, *107*, 639–655. http://dx.doi.org/10.1016/j.rse.2006.10.014.

Zhan, X., Houser, P. R., Walker, J. P., & Crow, W. T. (2006). A method for retrieving high-resolution surface soil moisture from hydros L-band radiometer and radar observations. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(6), 1534–1544. http://dx.doi.org/10.1109/TGRS.2005.863319.

Bibliography

Abarbanel, H. D. I. (1996), *Analysis of Observed Chaotic Data*, Springer Verlag.

Aguirre, L. A., and S. Billings (1995), Dynamical effects of overparametrization in nonlinear models, *Physica D*, *80(1,2)*, 26–40.

Al Bitar, A., D. Leroux, Y. H. Kerr, O. Merlin, P. Richaume, A. Sahoo, and E. F. Wood (2012), Evaluation of SMOS soil moisture products over continenal U.S. using the SCAN/SNOTEL network, *IEEE Trans. Geosci. Remote Sens.*, *50*(5), 1572–1586, doi:10.1109/TGRS.2012.2186581.

Al-Yaari, A., J.-P. Wigneron, A. Ducharne, Y. Kerr, P. de Rosnay, R. de Jeu, A. Govind, A. A. Bitar, C. Albergel, P. Richaume, and A. Mialon (2014), Global-scale evaluation of two satellite-based passive microwave soil moisture datasets (smos and amsr-e) with respect to land data assimilation system estimates, *Remote Sensing of Environmen*, *149*, 181–195 note =.

Allen, R. G. (2000), Using the FAO-56 dual crop coefficient method over an irrigated region as part of an evapotranspiration intercomparison study, *J. Hydrol.*, *229*(1-2), 27–41, doi:10.1016/S0022-1694(99)00194-8.

Allen, R. G., L. S. Pereira, D. Raes, and M. Smith (1998), *Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56-*, FAO, Rome.

Allen, R. G., M. Tasumi, A. Morse, and R. Trezza (2005), A landsat-based energy balance and evapotranspiration model in western us water rights regulation and planning, *Irrigation and Drainage Systems*, *19*(3-4), 251–268.

Allen, R. G., M. Tasumi, and R. Trezza (2007), Satellite-based energy balance for mapping evapotranspiration with internalized calibration (metric)—model, *Journal of irrigation and drainage engineering*, *133*(4), 380–394.

Anderson, M. C., J. M. Norman, G. R. Diak, W. P. Kustas, and J. R. Mecikalski (1997), A two-source time-integrated model for estimating surface fluxes using thermal infrared remote sensing, *Remote Sens. Environ.*, *60*, 195–216.

Balenzano, A., F. Mattia, G. Satalino, and M. Davidson (2011), Dense temporal series of c- and l-band sar data for soil moisture retrieval over agricultural crops, *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, *4*(2), 439–450, doi: 10.1109/JSTARS.2010.2052916.

Balsamo, G., S. Boussetta, E. Dutra, A. Beljaars, P. Viterbo, and B. Van den Hurk (2011), Evolution of land-surface processes in the IFS, *ECMWF Newsletter*, *127*, 17–22.

Bandara, R., J. P. Walker, C. Rüdiger, and O. Merlin (2015), Towards soil property retrieval from space: An application with disaggregated satellite observations, *Journal of Hydrology*, *522*, 582–593.

Bastiaanssen, W. G. M., M. Menenti, R. A. Feddes, and A. A. M. Holtslag (1998), A remote sensing surface energy balance algorithm for land (SEBAL) 1. Formulation, *J. Hydrology*, *212-213*, 198–212.

Bates, B., Z. W. Kundzewicz, S. Wu, and J. Palutikof (2008), *climate change and Water: technical Paper vi*, Intergovernmental Panel on Climate Change (IPCC).

Bausch, W., and C. Neale (1989), Spectral inputs improve corn crop coefficients and irrigation scheduling, *Transactions of the ASAE (USA)*.

Bergé, P., Y. Pommeau, and C. Vidal (1998), *L'ordre dans le chaos*, Hermann.

Béziat, P., E. Ceschia, and G. Dedieu (2009), Carbon balance of a three crop succession over two cropland sites in South West France, *Agr. For. Meteor.*, *149*(10), 1628–1645, doi:10.1016/j.agrformet.2009.05.004.

Bittelli, M., F. Ventura, G. S. Campbell, R. L. Snyder, F. Gallegati, and P. R. Pisa (2008), Coupling of heat, water vapor, and liquid water fluxes to compute evaporation in bare soils, *Journal of Hydrology*, *362*, 191–205.

Bouilloud, L., K. Chancibault, B. Vincendon, V. Ducrocq, F. Habets, G.-M. Saulnier, S. Anquetin, E. Martin, and J. Noilhan (2010), Coupling the ISBA land surface model and the TOPMODEL hydrological model for Mediterranean flash-flood forecasting: description, calibration, and validation, *J. Hydrometeor.*, *11*, 315–333.

Braud, I., A. Dantas-Antonino, M. Vauclin, J. Thony, and P. Ruelle (1995), A simple soil-plant-atmosphere transfer model (SiSPAT) development and field verification, *J. Hydrol.*, *166*, 213–250.

Brooks, R. H., and A. Corey (1964), *Hydraulic properties of porous media*, Paper No. 3, Colorado State Univ., Ft. Collins, CO.

Bsaibes, A., D. Courault, F. Baret, M. Weiss, A. Olioso, F. Jacob, O. Hagolle, O. Marloie, N. Bertrand, V. Desfond, and F. Kzemipour (2009), Albedo and LAI estimates from FORMOSAT-2 data for crop monitoring, *Remote Sens. Environ.*, *113*, 716–729, doi:10.1016/j.rse.2008.11.014.

Budyko, M. I. (1956), *Heat balance of the Earth's surface*, 255 pp., Gidrometeoizdat, Leningrad.

Cahill, A. T., M. B. Parlange, T. J. Jackson, P. O'Neill, and T. J. Schmugge (1999), Evaporation from nonvegetated surfaces: surface aridity methods and passive microwave remote sensing, *J. Appl. Meteor.*, *38*, 1346–1351.

Calvet, J.-C., J. Noilhan, J.-L. Roujean, P. Bessemoulin, M. Cabelguenne, A. Olioso, and J.-P. Wigneron (1998), An interactive vegetation svat model tested against data from six contrasting sites, *Agricultural and Forest Meteorology*, *92*(2), 73–95.

Camillo, P. J., and R. J. Gurney (1986), A resistance parameter for bare soil evaporation models, *Soil Sci.*, *141*, 95–105.

Campbell, G. S. (1974), A simple method for determining unsaturated conductivity from moisture retention data, *Soil Sci.*, *117*, 311–314.

Carlson, T. (2007), An overview of the 'triangle method' for estimating surface evapotranspiration and soil moisture from satellite imagery, *Sensors*, *7*, 1612–1629.

Carlson, T. N., R. R. Gillies, and E. M. Perry (1994), A method to make use of thermal infrared temperature and NDVI measurements to infer soil water content and fractional vegetation cover, *Remote Sens. Rev.*, *52*, 45–59.

Carlson, T. N., W. J. Capehart, and R. R. Gillies (1995), A new look at the simplified method for remote-sensing of daily evapotranspiration, *Remote Sens. Environ.*, *54*, 161–167.

Chang, S., D. Hahn, C.-H. Yang, and D. Norquist (1999), Validation study of the CAPS model land surface scheme using the 1987 Cabauw/PILPS dataset, *J. Appl. Meteor.*, *38*, 405–422.

Chanzy, A., and L. Bruckler (1993), Significance of soil surface moisture with respect to daily bare soil evaporation, *Water Resour. Res.*, *29*(4), 1113–1125.

Chanzy, A., L. Bruckler, and A. Perrier (1995), Soil evaporation monitoring: a possible synergism of microwave and infrared remote sensing, *J. Hydrol.*, *165*, 235–259.

Chanzy, A., M. Mumen, and G. Richard (2008), Accuracy of top soil moisture simulation using a mechanistic model with limited soil characterization, *Water Resour. Res.*, *44*(W03432), doi:10.1029/2006WR005765.

Chirouze, J., G. Boulet, L. Jarlan, R. Fieuzal, J. C. Rodriguez, J. Ezzahar, S. Er-raki, G. Bigeard, O. Merlin, J. Garatuza-Payan, C. Watts, and G. Chehbouni (2014), Inter-comparison of four remote sensing based surface energy balance methods to retrieve surface evapotranspiration and water stress of irrigated fields in semi-arid climate, *Hydrol. Earth Syst. Sci.*, *18*, 1165–1188, doi:10.5194/hess-18-1165-2014.

Chlouverakis, K., and J. Sprott (2004), A comparison of correlation and Lyapunov dimensions, *Physica D*, *200*, 156–164.

Choudhury, B., R. J. Reginato, and S. B. Idso (1986), An analysis of infrared temperature observations over wheat and calculation of latent heat flux, *Agr. For. Meteor.*, *37*, 75–88.

Collow, T. W., A. Robock, J. B. Basara, and B. G. Illston (2012), Evaluation of smos retrievals of soil moisture over the central united states with currently available in situ observations, *Journal of Geophysical Research*, *117*(D09113).

Comission, E. (2011), Roadmap to a resource efficient europe.

Cosby, B. J., G. M. Hornberger, R. B. Clapp, and T. R. Ginn (1984), A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils, *Water Resour. Res.*, *20*(6).

Coudert, B., C. Ottlé, B. Boudevillain, J. Demarty, and P. Guillevic (2006), Contribution of thermal infrared remote sensing data in multiobjective calibration of a dual-source SVAT model, *J. Hydrometeor.*, *7*, 404–420.

Crow, W. T., W. P. Kustas, and J. H. Prueger (2008), Monitoring root-zone soil moisture through the assimilation of a thermal remote sensing-based soil moisture proxy into a water balance model, *Remote Sens. Environ.*, *112*, 1268–1281.

Daly, E., and A. Porporato (2005), A Review of Soil Moisture Dynamics: From Rainfall Infiltration to Ecosystem Response, *Environmental Engineering Science*, doi:10.1089/ees.2005.22.9.

Das, N. N., D. Entekhabi, E. G. Njoku, J. 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. Remote Sens.*, *52*(4), 2018–2028, doi:10.1109/TGRS.2013.2257605.

de Tomás, A., H. Nieto, R. Guzinski, J. Salas, I. Sandholt, and P. Berliner (2014), Validation and scale dependencies of the triangle method for the evaporative fraction estimation over heterogeneous areas, *Remote Sens. Environ.*, *152*, 493–511.

Deardorff, J. W. (1978), Efficient prediction of ground temperature and moisture with inclusion of a layer of vegetation, *J. Geophys. Res.*, *83*, 1889–1903.

Dekic, L. J., D. T. Mihailovic, and B. Rajkovic (1995), A study of the sensitivity of bare soil evaporation schemes to soil surface wetness, using the coupled soil moisture and surface temperature prediction model, BARESOIL, *Meteorol. Atmos. Phys.*, *55*, 101–112.

Demaria, E. M., B. Nijssen, and T. Wagener (2007), Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model, *Journal of Geophysical Research*, *112*, D11,113.

Dente, L., Z. Su, and J. Wen (2012), Validation of SMOS Soil Moisture Products over the Maqu and Twente Regions, *Sensors*, *12*, 9965–9986.

Desborough, C. E., and A. J. Pitman (1998), The BASE land surface model, *Global and Planetary Change*, *19*, 3–18.

Desborough, C. E., A. J. Pitman, , and P. Irannejad (1996), Analysis of the relationship between bare soil evaporation and soil moisture simulated by 13 land surface schemes for a simple non-vegetated site, *Global and Planetary Change*, *13*, 47–56.

Dickinson, R., A. Henderson-Sellers, and P. Kennedy (1993), *Biosphere-Atmosphere Transfer Scheme (BATS) Version le as Coupled to the NCAR Community Climate Model*, vol. TN-387+STR, National Center for Atmospheric Research Technical Note, Boulder, Colorado.

Dickinson, R. E., A. Henderson-Sellers, P. J. Kennedy, and M. F. Wilson (1986), *Biosphere-Atmosphere Transfer Scheme (BATS) for the NCAR Community Climate model*, National Center for Atmospheric Research Technical Note, Boulder, Colorado.

Djamai, N., R. Magagi, K. Goïta, O. Merlin, Y. Kerr, and A. Roy (2016), A combination of DISPATCH downscaling algorithm with CLASS land surface scheme for soil moisture estimation at fine scale during cloudy days, *Remote Sensing of Environment*, *184*, 1–14.

Ducoudré, N. I., K. Laval, and A. Perrier (1993), SECHIBA, a new set of parameterizations of the hydrologic exchanges at the land-atmosphere interface within the LMD atmospheric general circulation model, *J. Climate*, *6*, 248–273.

Dumedah, G., J. P. Walker, and O. Merlin (2015), Root-zone soil moisture estimation from assimilation of downscaled soil moisture and ocean salinity data, *Advances in Water Resources*, *84*, 14–22.

Durand, Y., E. Brun, L. Merindol, G. Guyomarc'h, B. Lesaffre, and E. Martin (1993), A meteorological estimation of relevant parameters for snow models, *Annals of Glaciology*, *18*, 65–71.

Durand, Y., G. Giraud, E. Brun, L. Merindol, and E. Martin (1999), A computer-based system simulating snowpack structures as a tool for regional avalanche forecasting, *J. Glaciol.*, *45(151)*, 469–484.

ECMWF (2014), *IFS Documentation, Part IV: Physical Processes*, vol. Cy40r1, European Centre for Medium-Range Weather Forecasts, Reading, England.

Entekhabi, D., and P. S. Eagleson (1989), Land surface hydrology parameterization for atmospheric general circulation models including subgrid scale spatial variability, *J. Climate*, *2*, 816–831.

Entekhabi, D., E. G. Njoku, P. E. O'Neill, K. H. Kellogg, W. T. Crow, W. N. Edelstein, J. K. Entin, S. D. Goodman, T. J. Jackson, J. Johnson, J. Kimball, J. R. Piepmeier, R. Koster, N. Martin, K. C. McDonald, M. Moghaddam, S. Moran, R. Reichle, J. C. Shi, M. W. Spencer, S. W. Thurman, L. Tsang, and J. V. Zyl (2010), The Soil Moisture Active Passive (SMAP) mission, in *Proceedings of the IEEE*, vol. 98, pp. 704–716.

Er-Raki, S., A. Chehbouni, N. Guemouria, B. Duchemin, J. Ezzahar, and R. Hadria (2007), Combining FAO-56 model and ground-based remote sensing to estimate water consumptions of wheat crops in a semi-arid region, *Agr. Water Manag.*, *87*, 41–54, doi:10.1016/j.agwat.2006.02.004.

Er-Raki, S., A. Chehbouni, G. Boulet, and D. Williams (2010), Using the dual approach of fao-56 for partitioning et into soil and plant components for olive orchards in a semi-arid region, *Agricultural water management*, *97*(11), 1769–1778.

F. Grond, H. H. D. (2005), Local Lyapunov exponents for dissipative continuous systems, *Chaos, Solitons Fractals*, *23*, 1809–1817.

F. Grond, S. S. A. M. s. F. . O. E. R., H. H. Diebner (2003), A robust, locally interpretable algorithm for Lyapunov exponents, *Chaos, Solitons Fractals*, *16*, 841–852.

Famiglietti, J. S. (1994), Multiscale modeling of spatially variable water and energy balance processes, *Water Resour. Res.*, *30*(11), 3061–3078.

Fan, L., S. Liu, C. Bernhofer, H. Liu, and F. H. Berger (2007), Regional land surface energy fluxes by satellite remote sensing in the upper Xilin river watershed (inner Mongolia, China), *Theor. Appl. Climatol.*, *88*, 231–245, doi:10.1007/s00704-006-0241-9.

Fang, B., V. Lakshmi, R. Bindlish, T. J. Jackson, M. Cosh, and J. Basara (2013), Passive microwave soil moisture downscaling using vegetation index and skin surface temperature, *Vadose Zone J.*, *12*, n.a.–n.a., doi:10.2136/vzj2013.05.0089.

FAO (2009), Food and Agriculture Organization of the United Nations. Water and Food Security Statistics., `http://www.unwater.org/statistics_sec.html`.

FAO (2015), AQUASTAT Food and Agriculture Organization of the United Nations, `http://www.fao.org/nr/water/aquastat/countries_regions/africa/index.stm#a4`, [Online; accessed 11-November-2015].

Feddes, R. A., H. Hoff, M. Bruen, T. Dawson, P. de Rosnay, P. Dirmeyer, R. B. Jackson, P. Kabat, A. Kleidon, A. Lilly, and A. J. Pitman (2001), Modeling root water uptake in hydrological and climate models, *Bulletin of the American Meteorological Society*, *82*(12), 2797–2809.

Fieuzal, R., B. Duchemin, L. Jarlan, M. Zribi, F. Baup, O. Merlin, O. Hagolle, and J. Garatuza-Payan (2011), Combined use of optical and radar satellite data for the monitoring of irrigation and soil moisture of wheat crops, *Hydrol. Earth Syst. Sci.*, *15*, 1117–1129.

Findell, K. L., P. Gentine, B. R. Lintner, and C. Kerr (2011), Probability of afternoon precipitation in eastern United States and Mexico enhanced by high evaporation, *Nature Geosci.*, *4*, 434–439.

Freitas, U., C. Letellier, and L. Aguirre (2009), Failure in distinguishing colored noise from chaos using the "noise titration" technique, *Physical Review E*, *79*, 035201.

Galleguillos, M., F. Jacob, L. Prévot, and S. Liang (2011a), Mapping daily evapotranspiration over a Mediterranean vineyard watershed, *IEEE Geosci. Remote Sens. Lett.*, *8*(1), 168–172, doi:10.1109/LGRS.2010.2055230.

Galleguillos, M., F. Jacob, L. Prévot, A. French, and P. Lagacherie (2011b), Comparison of two temperature differencing methods to estimate daily evapotranspiration over a Mediterranean vineyard watershed from ASTER data, *Remote Sens. Environ.*, *115*, 1326–1340, doi:10.1016/j.rse.2011.01.013.

Gardner, W. R., and D. Hillel (1962), The relation of external evaporative conditions to the drying of soils, *J. Geophys. Res.*, *67*, 4319–4325.

Garrido, A., M. Llamas, C. Varela-Ortega, P. Novo, R. Rodgriguez-Casado, and M. Aldaya (2010), Water Footprint and virtual water trade in Spain. Policy implications, *Natural Resource Management and Policy*, *35*.

Gillespie, A., S. Rokugawa, T. Matsunaga, S. Cothern, S. Hook, and A. Khale (1998), A temperature and emissivity separation algorithm for Advanced Spaceborne Thermal Emission and Reflection (ASTER) images, *IEEE Trans. Geosci. Remote Sens.*, *36*, 1113–1126.

Gilmore, R. (1998), Topological analysis of chaotic dynamical systems, *Review of Modern Physics.*, *70*, 1455–1530.

Gómez, M., A. Olioso, J. A. Sobrino, and F. Jacob (2005), Retrieval of evapotranspiration over the Alpilles/ReSeDA experimental site using airborne POLDER sensor and a thermal camera, *Remote Sens. Environ.*, *96*, 399–408.

Good, S. P., D. Noone, and G. Bowen (2015), Hydrologic connectivity constrains partitioning of global terrestrial water fluxes, *Science*, *349*(6244), 175–177.

Gouesbet, G., and C. Letellier (1994), Global vector field reconstruction by using a multivariate polynomial l2-approximation on nets, *Physical Review E*, *49(6)*, 4955–4972.

Grassberger, P., and I. Procaccia (1983), Characterization of strange attractors, *Phys. Rev., Lett.*, *50*, 346–349.

Gusev, Y. M., and O. N. Nasonova (1998), The land surface parameterization scheme SWAP: description and partial validation, *Global Plan. Change*, *19*(1-4), 63–86.

Gutman, G., and A. Ignatov (1998), The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models, *Int. J. Remote Sens.*, *19*, 1533–1543.

Gutmann, E. D., and E. E. Small (2007), A comparison of land surface model soil hydraulic properties estimated by inverse modeling and pedotransfer functions, *Water Resources Research*, *43*(5), W05,418, doi:10.1029/2006WR005135.

Heitman, J. L., X. Xiao, R. Horton, and T. J. Sauer (2008), Sensible heat measurements indicating depth and magnitude of subsurface soil water evaporation, *Water Resour. Res.*, *44*, W00D05.

Huc, M., and S. Mangiarotti (2016), Un modèle chaotique pour l'épidémie de maladie à virus ebola d'afrique de l'ouest en 2014.

IPCC (2013), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Ed. Stocker, T.F., et al.. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535.

Jiang, L., and S. Islam (1999), A methodology for estimation of surface evapotranspiration over large areas using remote sensing observations, *Geophys. Res. Lett.*, *26*, 2773–2776.

Kalma, J. D., T. R. McVicar, and M. F. McCabe (2008), Estimating land surface evaporation: a review of methods using remotely sensed surface temperature data, *Surv. Geophys.*, *29*(4-5), 421–469, doi:10.1007/s10712-008-9037-z.

Kaplan, D., and L. Glasss (1992), Direct test for determinism, *Physical Review Letters*, *68(4)*, 427–430.

Kaplan, J. L., and J. A. Yorke (1979), *Chaotic Behavior of Multidimensional Difference Equations, In : Functional Differential Equations and Approximations of Fixed Points*, Springer, Berlin.

Kerr, Y., E. Jacquette, A. Al Bitar, F. Cabot, A. Mialon, P. Richaume, A. Quesney, and L. Berthon (2013), *CATDS SMOS L3 soil moisture retrieval processor: Algorithm Theoretical Baseline Document (ATBD)*, vol. SO-TN-CBSA-GS-0029, V2.0, CESBIO, Toulouse, France.

Kerr, Y. H., P. Waldteufel, J.-P. Wigneron, J.-M. Martinuzzi, J. Font, and M. Berger (2001), Soil moisture retrieval from space: the soil moisture and ocean salinity (SMOS) mission, *IEEE Trans. Geosci. Remote Sens.*, *39*, 1729–1735.

Kerr, Y. H., P. Waldteufel, J.-P. Wigneron, S. Delwart, F. Cabot, J. Boutin, M. J. Escorihuela, J. Font, N. Reul, C. Gruhier, S. E. Juglea, M. R. Drinkwater, A. Hahne, M. Martin-Neira, and S. Mecklenburg (2010), The SMOS mission: new tool for monitoring key elements of the global water cycle, in *IEEE*, vol. 98, pp. 666–687, doi:10.1109/JPROC.2010.2043032.

Kerr, Y. H., P. Waldteufel, P. Richaume, J. P. Wigneron, P. Ferrazzoli, A. Mahmoodi, A. Al Bitar, F. Cabot, C. Gruhier, D. Leroux, A. Mialon, and S. Delwart (2012), The SMOS soil moisture retrieval algorithm, *IEEE Trans. Geosci. Remote Sens.*, doi:10.1109/TGRS.2012.2184548.

Kerr, Y. H., L. Berthon, A. Mialon, F. Cabot, A. Al Bitar, P. Richaume, and E. Jacquette (2014), *CATDS LEVEL 3 - Data product description - Soil Moisture and Brightness Temperature*, CESBIO, Toulouse, France.

Kharrou, M., M. L. Page, A. Chehbouni, V. Simonneaux, S. Er-Raki, L. Jarlan, L. Ouzine, S. Khabba, and G. Chehbouni (2013), Assessment of Equity and Adequacy of Water Delivery in Irrigation Systems Using Remote Sensing-Based Indicators in Semi-Arid Region, Morocco, *Water Resources Management*, *27(13)*, 4697–4714.

Komatsu, T. S. (2003), Towards a robust phenomenological expression of evaporation efficiency for unsaturated soil surfaces, *J. Appl. Meteorol.*, *42*, 1330–1334.

Kool, D., N. Agam, N. Lazarovitch, J. L. Heitman, T. J. Sauer, and A. Ben-Gal (2014), A review of approaches for evapotranspiration partitioning, *Agricultural and Forest Meteorology*, *184*, 56–70.

Koster, R. D., and M. J. Suarez (1996), *Energy and water balance calculations in the Mosaic LSM*, vol. 9, 10606, NASA Tech. Memo.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model 1. Model structure, *J. Geophys. Res.*, *105*(D20), 24,809–24,822.

Kowalczyk, E. A., Y. P. Wang, R. M. Law, H. L. Davies, J. L. McGregor, and G. Abramowitz (November 2006), *The CSIRO Atmosphere Biosphere Land Exchange (CABLE) model for use in climate models and as an offline model*, vol. paper 013, CSIRO Marine and Atmospheric Research, Aspendale, Australia.

Kustas, W. P., T. J. Schmugge, K. S. Humes, T. J. Jackson, R. Parry, M. A. Weltz, and M. S. Moran (1993), Relationships between evaporative fraction and remotely sensed vegetation index and microwave brightness temperature for semiarid rangelands, *J. Appl. Meteor.*, *32*, 1781–1790.

Lawrence, D. M., P. E. Thornton, K. W. Oleson, and G. B. Bonan (2007), The Partitioning of Evapotranspiration into Transpiration, Soil Evaporation, and Canopy Evaporation in a GCM: Impacts on Land–Atmosphere Interaction, *J. Hydrometeor.*, *8*, 862–880.

Lee, H. (1997), Notes and correspondence: improvement of surface flux calculation in the atmospheric surface layer, *J. Appl. Meteor.*, *36*, 1416–1423.

Lee, T. J., and R. A. Pielke (1992), Estimating the soil surface specific humidity, *J. Appl. Meteor.*, *31*, 480–484.

Letellier, C. (2006), *Le chaos dans la nature*.

Letellier, C., L. L. Sceller, P. Dutertre, G. Gouesbet, Z. Fei, and J. Hudson (1995), Topological characterization and global vector field reconstruction of an experimental electrochemical system, *The Journal of Physical Chemistry*, *99 (18)*, 7016–7027.

Letellier, C., L. A. Aguirre, and J. Maquet (2005), Relation between observability and differential embeddings for nonlinear dynamics, *Physical Review E*, *71, 066213*, 7016–7027.

Letellier, C., L. A. Aguirre, and U. S. Freitas (2009), Frequently asked questions about global modeling, *Chaos*, *023103*.

Li, F., W. P. Kustas, M. C. Anderson, T. J. Jackson, R. Bindlish, and J. H. Prueger (2006), Comparing the utility of microwave and thermal remote-sensing constraints in two-source energy balance modeling over an agricultural landscape, *Remote sensing of environment*, *101*(3), 315–328.

Li, Z.-L., R. Tang, Z. Wan, Y. Bi, C. Zhou, B. Tang, G. Yan, and X. Zhang (2009), A review of current methodologies for regional evapotranspiration estimation from remotely sensed data, *Sensors*, *9*, 3801–3853, doi:10.3390/s90503801.

Liu, S., D. Mao, and L. Jia (2007), Evaluating parameterizations of aerodynamic resistance to heat transfer using field measurements, *Hydrol. Earth Syst. Sci.*, *11*, 769–783.

Liu, Y., T. Hiyama, and Y. Yamaguchi (2006), Scaling of land surface temperature using satellite data: a case examination on ASTER and MODIS products over a heterogeneous terrain area, *Remote Sens. Environ.*, *105*, 115–128.

Long, D., and V. P. Singh (2010), Integration of the gg model with sebal to produce time series of evapotranspiration of high spatial resolution at watershed scales, *Journal of Geophysical Research: Atmospheres (1984–2012)*, *115*(D21).

Long, D., and V. P. Singh (2012), A two-source trapezoid model for evapotranspiration (TTME) from satellite imagery, *Remote Sens. Environ.*, *121*, 370–388, doi:10.1016/j.rse.2012.02.015.

Long, D., and V. P. Singh (2013), Assessing the impact of end-member selection on the accuracy of satellite-based spatial variability models for actual evapotranspiration estimation, *Water Resour. Res.*, *49*, 2601–2618, doi:10.1002/wrcr.20208.

Long, D., V. P. Singh, and Z.-L. Li (2011), How sensitive is SEBAL to changes in input variables, domain sizes and satellite sensor?, *J. Geophys. Res.*, *116*(D21107), doi:10.1029/2011JD016542.

Long, D., V. P. Singh, and B. R. Scalon (2012), Deriving theoretical boundaries to address scale dependencies of triangle models for evapotranspiration estimation, *J. Geophys. Res.*, *117*(D05113), doi:10.1029/2011JD017079.

Lorenz, E. (1963), Deterministic Nonperiodic Flow, *Journal of the Atmospheric Sciences*, *20*, 130–141.

Mahfouf, J. F., and J. Noilhan (1991), Comparative study of various formulations of evaporation from bare soil using in situ data, *J. Appl. Meteor.*, *30*, 1354–1365.

Mahrt, L., and H. Pan (1984), A two-layer model of soil hydrology, *Bound.-Layer Meteor.*, *29*, 1–20.

Manabe, S. (1969), Climate and the Ocean Circulation: I. The Atmospheric Circulation and the Hydrology of the Earth's Surface, *Monthly Weather Review*, *97*(11), 739–774.

Mangiarotti, S. (2015), Low dimensional chaotic models for the plague epidemic in Bombay (1896 - 1911), *Chaos, Solitons and Fractals*, *81A*, 184–196.

Mangiarotti, S., C. Coudret, L. Drapeau, and L. Jarlan (2012a), Polynomial search and Global modelling: two algorithms for modeling chaos, *Physical Review E*, *86(4)*(046205).

Mangiarotti, S., P. Mazega, E. Mougin, and P. Hiernaux (2012b), Predictability of vegetation cycles over the semi-arid region of Gourma (Mali) from forecasts of AVHRR-NDVI signals, *Remote Sensing of Environment*, *123*, 246–257.

Mangiarotti, S., L. Drapeau, and C. Letellier (2014), Two chaotic global models for cereal crops cycles observed from satellite in Northern Morocco, *Chaos*, *24*(023130).

Mangiarotti, S., F. L. Jean, M. Huc, and C. Letellier (2016), Global Modeling of aggregated and associated chaotic dynamics, *Chaos, Solitons and Fractals*, *83*, 82–96.

Maquet, J., C. Letellier, and L. A. Aguirre (2007), Global models from the Canadian lynx cycles as a direct evidence for chaos in real ecosystems, *Journal of mathematical biology*, *55(1)*, 21–39.

Menenti, M., W. Bastiaanssen, D. van Eick, and M. A. Abd el Karim (1989), Linear relationships between surface reflectance and temperature and their application to map actual evaporation of groundwater, *Adv. Space Res.*, *9*(1), 165–176.

Mengelkamp, H.-T., K. Warrach, and E. Raschke (1999), SEWAB - a parameterization of the Surface Energy and Water Balance for atmospheric and hydrologic models, *Advances in Water Resources*, *23*, 165–175.

Merlin, O. (2013), An original interpretation of the wet edge of the surface temperature–albedo space to estimate crop evapotranspiration (SEB-1S), and its validation over an irrigated area in northwestern Mexico, *Hydrol. Earth Syst. Sci.*, *17*, 3623–3637.

Merlin, O., G. Chehbouni, Y. Kerr, and D. Goodrich (2006a), A downscaling method for distributing surface soil moisture within a microwave pixel: Application to the Monsoon'90 data, *Remote Sens. Environ.*, *101*, 379–389.

Merlin, O., G. Chehbouni, G. Boulet, and Y. Kerr (2006b), Assimilation of disaggregated microwave soil moisture into a hydrologic model using coarse-scale meterological data, *J. Hydrometeor.*, *7*, 1308–1322.

Merlin, O., G. Chehbouni, J. P. Walker, R. Panciera, and Y. Kerr (2008a), A simple method to disaggregate passive microwave based soil moisture, *IEEE Trans. Geosc. Remote Sens., SMOS Special Issue*, *46*(3), 786–796, doi:10.1109/TGRS.2007.914807.

Merlin, O., J. P. Walker, A. Chehbouni, and Y. Kerr (2008b), Towards deterministic downscaling of SMOS soil moisture using MODIS derived soil evaporative efficiency, *Remote Sens. Environ.*, *112*, 3935–3946, doi:10.1016/j.rse.2008.06.012.

Merlin, O., B. Duchemin, O. Hagolle, F. Jacob, B. Coudert, G. Chehbouni, G. Dedieu, J. Garatuza, and Y. Kerr (2010a), Disaggregation of MODIS Surface Temperature over an Agricultural Area Using a Time Series of Formosat-2 Images, *Remote Sens. Environ.*, *114*(11), 2500–2512, doi:10.1016/j.rse.2010.05.025.

Merlin, O., C. Rüdiger, P. Richaume, A. Al Bitar, A. Mialon, J. P. Walker, and Y. Kerr (2010b), Disaggregation as a top-down approach for evaluating 40 km resolution SMOS data using point-scale measurements: application to AACES-1, in *SPIE, Remote Sensing of Agriculture, Ecosystems, and Hydrology XII*, pp. 78,240I–1–8, Toulouse, France, doi:10.1117/12.865751.

Merlin, O., A. Al Bitar, V. Rivalland, P. Béziat, E. Ceschia, and G. Dedieu (2011), An analytical model of evaporation efficiency for unsaturated soil surfaces with an arbitrary thickness, *J. Appl. Meteor. Clim.*, *50*(2), 457–471, doi:10.1175/2010JAMC2418.1.

Merlin, O., C. Rüdiger, A. Al Bitar, P. Richaume, J. P. Walker, and Y. H. Kerr (2012a), Disaggregation of SMOS soil moisture in Southeastern Australia, *IEEE Trans. Geosci. Remote Sens.*, *50*(5), 1557–1571, doi:10.1109/TGRS.2011.2175000.

Merlin, O., F. Jacob, J.-P. Wigneron, J. Walker, and G. Chehbouni (2012b), Multidimensional disaggregation of land surface temperature using high-resolution red, near-infrared, shortwave-infrared, and microwave-L bands, *IEEE Trans. Geosci. Remote Sens.*, *50*(5), 1864–1880, doi:10.1109/TGRS.2011.2169802.

Merlin, O., V. Stefan, Y. Malbéteau, M.-J. Escorihuela, C. Rüdiger, and L. Jarlan (2013a), Data disaggregation and evapotranspiration modeling: a synergism between multi-spectral/multi-resolution remote sensing data, in *American Geophysical Union (AGU)*.

Merlin, O., M. J. Escorihuela, M. A. Mayoral, O. Hagolle, A. A. Bitar, and Y. Kerr (2013b), Self-calibrated evaporation-based disaggregation of SMOS soil moisture: an evaluation study at 3 km and 100 m resolution in Catalunya, Spain, *Remote Sensing of Environment*, *130*, 25–38.

Merlin, O., J. Chirouze, A. Olioso, L. Jarlan, G. Chehbouni, and G. Boulet (2014), An image-based four-source surface energy balance model to estimate crop evapotranspiration from solar reflectance/thermal emission data (SEB-4S), *Agricultural and Forest Meteorology*, *184*, 188–203.

Merlin, O., V. G. Stefan, A. Amazirh, A. Chanzy, E. Ceschia, S. Er-Raki, P. Gentine, T. Tallec, J. Ezzahar, S. Bircher, J. Beringer, and S. Khabba (2016), Modeling soil evaporation efficiency in a range of soil and atmospheric conditions using a meta-analysis approach, *Water Resources Research*, submitted.

Mihailović, D. T. (1996), Description of land-air parameterization scheme (LAPS), *Global and Planetary Change*, *13*, 207–215.

Mihailovic, D. T., B. Rajkovic, L. Dekic, R. A. Pielke, T. J. Lee, and Z. Ye (1995), The validation of various schemes for parameterizing evaporation from bare soil for use in meteorological models: a numerical study using in situ data, *Boundary-Layer Meteorology*, *76*, 259–289.

Milly, P. C. D. (1986), An event-based simulation model of moisture and energy fluxes at a bare soil surface, *Water Resour. Res.*, *22*(12), 1680–1692.

Milly, P. C. D., and A. B. Shmakin (2002), Global modeling of land water and energy balances. Part I: the Land Dynamics (LaD) model, *J. Hydrometeor.*, *3*, 283–299.

Molero, B., O. Merlin, Y. Malbéteau, A. A. Bitar, F. Cabot, V. G. Stefan, S. Bacon, Y. Kerr, R. Bindlish, and T. J. Jackson (2016), New smos disaggregated soil moisture product at 1km resolution: processor overview and first validation results, *Remote Sensing of Environment*.

Monteith, J. L. (1973), *Principles of environmental physics*, 241 pp., Edward Arnold Press.

Moran, M. S., T. R. Clarke, Y. Inoue, and A. Vidal (1994), Estimating crop water deficit using the relation between surface-air temperature and spectral vegetation index, *Remote Sens. Environ.*, *49*, 246–263.

Mueller, N. D., J. S. Gerber, M. Johnston, D. Ray, N. Ramankutty, and J. A. Foley (2012), Closing yield gaps through nutrient and water management, *Nature*, *490(7419)*, 254–257.

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, doi:10.1109/TGRS.2008.2011617.

Nichols, W. E., and R. H. Cuenca (1993), Evaluation of the evaporative fraction for parameterization of the surface, *Water Resour. Res.*, *29*(11), 3681–90.

Niu, G., Z. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari, and Y. Xia (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, *J. Geophys. Res.*, *116*(D12109), doi:10.1029/2010JD015139.

Njoku, E., T. Jackson, V. Lakshmi, T. Chan, and S. Nghiem (2003), Soil moisture retrieval from AMSR-E, *IEEE Trans. Geosci. Remote Sens.*, *41*, 215–229.

Njoku, E. G., and D. Entekhabi (1996), Passive microwave remote sensing of soil moisture, *J. Hydrol.*, *184*, 101–129.

Noilhan, J., and J.-F. Mahfouf (1996), The ISBA land surface parameterisation scheme, *Glob. Planet. Change*, *13*, 145–159.

Noilhan, J., and S. Planton (1989), A simple parameterization of land surface processes for meteorological models, *Mon. Wea. Rev.*, *117*, 536–549.

Norman, J. M., W. P. Kustas, and K. S. Humes (1995), Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature, *Agr. For. Meteorol.*, *77*, 263–293.

Oki, T., and S. Kanae (2006), Global hydrological cycles and world water resources, *Science*, *313*, 1068–1072.

Oleson, K. W., D. M. Lawrence, G. B. Bonan, M. G. Flanner, E. Kluzek, P. J. Lawrence, S. Levis, S. C. Swenson, P. E. Thornton, A. Dai, M. Decker, R. Dickinson, J. Feddema, C. L. Heald, F. Hoffman, J.-F. Lamarque, N. Mahowald, G.-Y. Niu, T. Qian, J. Randerson, S. Running, K. Sakaguchi, A. Slater, R. Stöckli, A. Wang, Z.-L. Yang, X. Zeng, and X. Zeng (2010), *Technical Description of version 4.0 of the Community Land Model (CLM)*, vol. TN-478+STR, National Center for Atmospheric Research, Boulder, Colorado.

Oleson, K. W., D. M. Lawrence, G. B. Bonan, B. Drewniak, M. Huang, C. D. Koven, S. Levis, F. Li, W. J. Riley, Z. M. Subin, S. C. Swenson, P. E. Thornton, A. Bozbiyik, R. Fisher, C. L. Heald, E. Kluzek, J.-F. Lamarque, P. J. Lawrence, L. R. Leung, W. Lipscomb, S. Muszala, D. M. Ricciuto, W. Sacks, Y. Sun, J. Tang, and Z.-L. Yang (2013), *Technical Description of version 4.5 of the Community Land Model (CLM)*, vol. TN-503+STR, National Center for Atmospheric Research, Boulder, Colorado.

Petropoulos, G., T. Carlson, M. Wooster, and S. Islam (2009), A review of Ts/VI remote sensing based methods for the retrieval of land surface energy fluxes and soil surface moisture, *Prog. Phys. Geography*, *33*(2), 224–250.

Philip, J. R., and D. A. de Vries (1957), Moisture movement in porous materials under temperature gradients, *Trans. Am. Geophys. Union*, *38*, 222–232.

Piles, M., A. Camps, M. Vall-llossera, I. Corbella, R. Panciera, C. Rüdiger, Y. H. Kerr, and J. P. Walker (2011), Downscaling SMOS-derived soil moisture using MODIS visible/infrared data, *IEEE Trans. Geosci. and Remote Sens.*, in press.

Pitman, A. J., Z. L. Yang, J. Cogley, and A. Henderson-Sellers (1991), *Description of bare essentials of surface transfer for the Bureau of Meteorology Research Center AGCM*, Macquarie University, North Ryde, Australia.

Porporato, A., F. Laio, L. Ridolfi, and I. Rodriguez-Iturbe (2001), Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress: III. Vegetation water stress, *Advances in Water Resources*, *24*(7), 725–744.

Prihodko, L., and S. N. Goward (1997), Estimation of air temperature from remotely sensed surface observation, *Remote Sens. Environ.*, *60*, 335–346.

Quintana-Seguí, P., M. C. Peral, M. Turco, M.-C. Llasat, and E. Martin (), Meteorological analysis systems in north-east spain. validation of safran and span, *Journal of Environmental Informatics*, in press.

Quintana-Seguí, P., P. L. Moigne, Y. Durand, E. Martin, and F. Habets (2008), Analysis of near-surface atmospheric variables: Validation of the safran analysis over france, *Journal of Applied Meteorology and Climatology*, *47*, 92–107.

Rawls, W. J., and D. L. Brackensiek (1985), *Prediction of soil water properties for hydrologic modelling, in Proceedings of the Symposium on Watershed Management in the Eighties, New York*, 293–299 pp., Am. Soc. of Civ. Eng., edited by E. B. Jones and T. J. Ward.

Ritter, B., and J. F. Geleyn (1992), A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations., *Mon. Wea. Rev.*, *120*, 303–325.

Robock, A., K. Y. Vinnikov, C. A. Schlosser, N. A. Speranskaya, and Y. Xue (1995), Use of midlatitude soil moisture and meteorological observations to validate soil moisture simulations with Biosphere and Bucket models, *Journal of Climate*, *8*, 15–35.

Roerink, G. J., Z. Su, and M. Menenti (2000), S-SEBI: a simple remote sensing algorithm to estimate the surface energy balance, *Phys. Chem. Earth*, *25*(2), 147–157.

Rosenzweig, C., and F. Abramopoulos (1996), Land-Surface Model Development for the GISS GCM, *J. Climate*, *10*, 2040–.

Sakaguchi, K., and X. Zeng (2009), Effects of soil wetness, plant litter, and under-canopy atmospheric stability on ground evaporation in the Community Land Model (CLM3.5), *J. Geophys. Res.*, *114*, D01,107, doi:10.1029/2008JD010834.

Salvador, R., A. Martínez-Cob, J. Cavero, and E. Playán (2011), Seasonal on-farm irrigation performance in the Ebro basin (Spain): Crops and irrigation systems, *Agricultural Water Management*, *98*, 577–587.

Satalino, G., A. Balenzano, F. Mattia, and M. W. Davidson (2014), C-band sar data for mapping crops dominated by surface or volume scattering, *Geoscience and Remote Sensing Letters, IEEE*, *11*(2), 384–388.

Saxton, K. E., W. J. Rawls, J. S. Romberger, and R. I. Papendick (1986), Estimating Generalized Soil-water Characteristics from Texture, *Soil Sci. Soc. Amer. J.*, *50*(4), 1031–1036.

Schaake, J. C., V. I. Koren, Q.-Y. Duan, K. Mitchell, and F. Chen (1996), Simple water balance model for estimating runoff at different spatial and temporal scales, *J. Geophys. Res.*, *101*(D3), 7461–7475.

Schlesinger, W. H., and S. Jasechko (2014), Transpiration in the global water cycle, *Agricultural and Forest Meteorology*, *189*, 115–117.

Schmugge, T. J. (1998), Applications of passive microwave observations of surface soil moisture, *J. Hydrol.*, *212/213*, 188–197.

Schmugge, T. J., S. J. Hook, and C. Coll (1998), Recovering surface temperature and emissivity from thermal infrared multispectral data, *Remote Sens. Environ.*, *65*, 121–131.

Schmugge, T. J., W. P. Kustas, J. C. Ritchie, T. J. Jackson, and A. Rango (2002), Remote sensing in hydrology, *Advances in water resources*, *25*(8), 1367–1385.

Sellers, P., Y. Mintz, Y. e. a. Sud, and A. Dalcher (1986), A simple biosphere model (sib) for use within general circulation models, *Journal of the Atmospheric Sciences*, *43*(6), 505–531.

Sellers, P. J., M. D. Heiser, and F. G. Hall (1992), Relations between surface conductance and spectral vegetation indices at intermediate (100 m^2 to 15 km^2) length scales, *J. Geophys. Res.*, *97*(D17), 19,033–19,059.

Sellers, P. J., D. A. Randall, G. J. Collatz, J. A. Berry, C. B. Field, D. A. Dazlich, C. Zhang, G. D. Collelo, and L. Bounoua (1996), A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation, *J. Climate*, *9*, 676–705.

Senay, G., S. Bohms, R. Singh, P. Gowda, and N. Velpuri (2013), Operational Evapotranspiration Mapping Using Remote Sensing and Weather Datasets: A New Parametrization for the SSEB Approach, *J. of the American Water Assoc.*, *49*, 577–591.

Seneviratne, S. I., D. Luthi, M. Litschi, and C. Schar (2006), Land-atmosphere coupling and climate change in Europe, *Nature*, *443*, 205–209.

Shahraeeni, E., P. Lehmann, and D. Or (2012), Coupling of evaporative fluxes from drying porous surfaces with air boundary layer: Characteristics of evaporation from discrete pores, *Water Resour. Res.*, *48*, W09,525.

Sheffield, J., and E. F. Wood (2008), Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations, *Climate Dyn.*, *31*, 79–105.

Shuttleworth, W. J., and J. S. Wallace (1985), Evaporation from sparse canopies-an energy combination theory, *Q. J. R. Meteorol. Soc.*, *111*, 839–855.

Shuttleworth, W. J., R. J. Gurney, A. Y. Hsu, and J. P. Ormsby (1989), FIFE: the variation in energy partition at surface flux sites, *IAHS Publ.*, *186*, 67–74.

Simonneaux, V., B. Duchemin, D. Helson, S. Er-Raki, A. Olioso, and A. Chehbouni (2008), he use of high-resolution image time series for crop classification and evapotranspiration estimate over an irrigated area in central Morocco, *Int. J. Remote Sens.*, *29 (1)*, 95–116.

Sivapalan, M., G. Bloschl, L. Zhang, and R. Vertessy (2003), Downward approach to hydrological prediction, *Hydrol. Proc.*, *17*, 2101–2111.

Soarès, J., R. Bernard, O. Taconet, D. Vidal-Madjar, and A. Weill. (1988), Estimation of bare soil evaporation from airborne measurements, *J. Hydrol.*, *99*, 281–296.

Sobrino, J. A., M. Gómez, J. C. Jiménez-Munoz, A. Olioso, and G. Chehbouni (2005), A simple algorithm to estimate evapotranspiration from DAIS data: application to the DAISEX campaigns, *J. Hydrol.*, *315*, 117–125, doi:10.1016/j.jhydrol.2005.03.027.

Sobrino, J. A., M. Gómez, J. C. Jiménez-Munoz, and A. Olioso (2007), Application of a simple algorithm to estimate daily evapotranspiration from NOAA-AVHRR images for the Iberian Peninsula, *Remote Sens. Environ.*, *110*, 139–148, doi:10.1016/j.rse.2007.02.017.

Soet, M., and J. Stricker (2003), Functional behaviour of pedotransfer functions in soil water flow simulation, *Hydrological processes*, *17*(8), 1659–1670.

Solano, R., K. Didan, A. Jacobson, and A. Huete (2010), *MODIS Vegetation Index User's Guide (MOD13 Series) (Vol. v2.0)*.

Solomon, S. e. a. (2007), *Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*, Cambridge Univ. Press.

Stefan, V. G., O. Merlin, S. Er-Raki, M.-J. Escorihuela, and S. Khabba (2015), Consistency between in situ, model-derived and high-resolution-image-based soil temperature endmembers: towards a robust data-based model for multi-resolution monitoring of crop evapotranspiration, *Remote Sensing*, *7*, 10,444–10,479, doi:10.3390/rs70810444.

Stisen, S., I. Sandholt, A. Nørgaard, R. Fensholt, and K. H. Jensen (2008), Combining the triangle method with thermal inertia to estimate regional evapotranspiration - Applied to MSG-SEVERI data in the Senegal River basin, *Remote Sens. Environ.*, *112*, 1242–1255.

Stocker, T., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley (2014), *Climate change 2013: The physical science basis*, Cambridge University Press Cambridge, UK, and New York.

Su, Z. (2002), The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, *Hydrol. Earth Syst. Sci.*, *6*(1), 85–99.

Sun, S. F. (1982), *Moisture and heat transport in a soil layer forced by atmospheric conditions*, vol. M. S. Thesis, 72 pp., Univ. Conn., Conn., USA.

Takens, F. (1981), Detecting strange attractors in turbulence, *Lecture Notes in Mathematics*, *898*, 366–381.

Tang, R., Z.-L. Li, and B. Tang (2010), An application of the Ts-VI method with enhanced edges determination for evapotranspiration estimation from MODIS data in arid and semi-arid regions: Implementation and validation, *Remote Sens. Environ.*, *114*, 540–551, doi:10.1016/j.rse.2009.10.012.

Timmermans, W. J., W. P. Kustas, M. C. Anderson, and A. N. French (2007), An intercomparison of the Surface Energy Balance Algorithm for Land (SEBAL) and the Two-Source Energy Balance (TSEB) modeling schemes, *Remote Sensing of Environment*, *108*, 369–384.

Torres, R., P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Floury, M. Brown, et al. (2012), Gmes sentinel-1 mission, *Remote Sensing of Environment*, *120*, 9–24.

Twarakavi, N. K. C., J. Simunek, and M. G. Schaap (2009), Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines, *Soil Sci. Soc. Am.*, *73*(5), 1443–1452.

Van de Griend, A. A., and M. Owe (1994), Bare soil surface resistance to evaporation by vapor diffusion under semiarid conditions, *Water Resour. Res.*, *30*(2), 181–188.

Van den Hurk, B. J. J. M., P. Viterbo, A. C. M. Beljaars, and A. K. Betts (2000), *Offline validation of the ERA40 surface scheme*, 295, ECMWF Tech. Memo, Reading, UK.

Van Genuchten, M. T. (1980), A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. Am.*, *44*, 892–898.

Vereecken, H., J. Maes, and P. Darius (1989), Estimating the soil moisture retention characteristic from texture, bulk density, and carbon content, *Soil Sci.*, *148*, 389–403.

Verseghy, D. L. (1996), Local climates simulated by two generations of Canadian GCM land surface schemes, *Atmosphere-Ocean*, *34*(3), 435–456.

Viterbo, P., and A. C. M. Beljaars (1995), *An improved land surface parametrization scheme in the ECMWF model and its validation*, 75, ECMWF Tech. Report, Research Department, ECMWF, Reading, UK.

Wallace, J. S. (2000), Increasing agricultural water use efficiency to meet future food production, *Agriculture, Ecosystems and Environment*, *82*, 105–119.

Wan, Z. (1999), *MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD)*.

Wan, Z. (2006), *MODIS Land Surface Temperature Products Users' Guide - Collection 5. Sioux Falls, South Dakota*.

Wang, Y. P., E. Kowalczyk, R. Leuning, G. Abramowitz, M. R. Raupach, B. Pak, E. van Gorsel, and A. Luhar (2011), Diagnosing errors in a land surface model (CABLE) in the time and frequency domains, *J. Geosphys. Res. Biosgeosci.*, *116*(G1), doi:10.1029/2010JG001385.

Weiss, M., F. Baret, M. Leroy, A. Begué, O. Hautecoeur, and R. Santer (1999), Hemispherical reflectance and albedo estimates from the accumulation of across track sun synchroneous satellite data, *J. Geophys. Res.*, *104*, 221–232.

Wetzel, P. J., and A. Boone (1995), A parameterization for land-atmosphere-cloud exchange (place): documentation and testing of a detailed process model of the partly cloudy boundary layer over heterogeneous land, *J. Climate*, *8*, 1810–1837.

Wetzel, P. J., and J.-T. Chang (1988), Evapotranspiration from nonuniform surfaces: a first approach for short-term numerical weather prediction, *Montly Weather Review*, *116*, 600–621.

Wigneron, J.-P., Y. Kerr, P. Waldteufel, K. Saleh, M.-J. Escorihuela, P. Richaume, P. Ferrazzoli, P. de Rosnay, R. Gurneye, J.-C. Calvet, J. P. Grant, M. Guglielmettih, B. Hornbuckle, C. Matzler, T. Pellarin, and M. Schwank (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*(4), 639–655, doi:10.1016/j.rse.2008.10.014.

Williams, D., W. Cable, K. Hultine, J. Hoedjes, E. Yepez, V. Simonneaux, S. Er-Raki, G. Boulet, H. De Bruin, A. Chehbouni, et al. (2004), Evapotranspiration components determined by stable isotope, sap flow and eddy covariance techniques, *Agricultural and Forest Meteorology*, *125*(3), 241–258.

Wolf, A., J. B. Swift, H. L. Swinney, and J. A. Vastano (1985), Determining Lyapunov exponents from a time series, *Physica D*, *16*, 285–317.

Wood, E. F., D. P. Lettenmaier, and V. G. Zartarian (1992), A land-surface hydrology parameterization with subgrid variability for general circulation models, *J. Geophys. Res.*, *97*(D3), 2717–2728.

Wösten, J. H. M. (1997), *Pedotransfer functions to evaluate soil quality, in Developments in Soils Sciences*, 221– 245 pp., Elsevier, edited by E. G. Gregorich and M. R. Carter.

Xue, Y., P. J. Sellers, J. L. Kinter, and J. Shukla (1991), A simplified biosphere model for global climate studies, *J. Climate*, *4*, 345–364.

Yang, K., N. Tamai, and T. Koike (2001), Analytical solution of surface layer similarity equations, *J. Appl. Meteor.*, *40*, 1647–1653.

Yang, K., T. Koike, and H. Ishikawa (2008), Turbulent flux transfer over bare soil surfaces: Characteristics and parameterization, *J. Appl. Meteor.*, *40(1)*, 276–290.

Yang, Y., and S. Shang (2013), A hybrid dual-source scheme and trapezoid framework-based evapotranspiration model (HTEM) using satellite images: algorithm and model test, *J. Geophys. Res. Atm.*, *118*, 2284–2300, doi:10.1002/jgrd.50259.

Yang, Z.-L., R. E. Dickinson, W. J. Shuttleworth, and M. Shaikh (1998), Treatment of soil, vegetation and snow in land surface models: a test of the biosphere-atmosphere transfer scheme with the hapex-mobilhy, abracos and russian data, *Journal of Hydrology*, *212-213*, 109–127.

Ye, Z., and R. A. Pielke (1993), Atmospheric parameterization of evaporation from non-plant-covered surfaces, *J. Appl. Meteor.*, *32*, 1248–1258.

Yongjiu, D., and Z. Qingcun (1997), A land surface model (IAP94) for climate studies Part I: formulation and validation in off-line experiments, *Advances in Atmospheric Sciences*, *14*(4), 434–460.

Zeng, J., L. Zhen, Q. Chen, H. Bi, J. Qiu, and P. Zou (2015), Evaluation of remotely sensed and reanalysis soilmoisture products over the Tibetan Plateau using in-situ observations, *Remote Sensing of environment*, *163*, 91–110.

Zhan, X., S. Miller, N. Chauhan, L. Di, and P. Ardanuy (2002), *Soil moisture visible/infrared radiometer suite algorithm theoretical basis document*, vol. Tech. Rep., Raytheon Syst. Company, Lanham, MD.

Zhan, X., P. R. Houser, J. P. Walker, and W. Crow (2006), A method for retrieving high resolution surface soil moisture from Hydros L-band radiometer and radar observations, *IEEE Trans. Geosci. Remote Sens.*, *44*(6), 1534–1544.

Zhang, Y., S. Mangiarotti, and M. Leblanc (2016), Modélisation par technique globale du débit de trois sources karstiques (le Doubs, la Touvre et le Lez), 19e Rencontre du Non Linéaire, 19, Université Paris Diderot, Paris, France, 127-132, ISBN 978-2-9538596-5-2.

Zribi, M., A. Chahbi, M. Shabou, Z. Lili-Chabaane, B. Duchemin, N. Baghdadi, R. Amri, and A. Chehbouni (2011), Soil surface moisture estimation over a semi-arid region using ENVISAT ASAR radar data for soil evaporation evaluation, *Hydrol. Earth Syst. Sci.*, *15*, 345–358.